Algebra & Number Theory

Volume 10

2016

No. 6

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2016 is US $290/year for the electronic version, and $485/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1147

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

$$\prod_{\mathfrak{q} \in S} \mathrm{Norm}_{\mathbb{F}_\mathfrak{q}/\mathbb{F}_5}(u \bmod \mathfrak{q}) \neq \bar{1}.$$

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 ℓ, 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 ℓ 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

$$x^p + y^q = z^r \tag{1}$$

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

The authors are supported by EPSRC Programme Grant "LMF: L-Functions and Modular Forms" EP/K034383/1.
MSC2010: primary 11D41, 11F80; secondary 11G05, 11F41.
Keywords: elliptic curves, modularity, Galois representation, level lowering, irreducibility, generalized Fermat, Fermat–Catalan, Hilbert modular forms.

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

$$\begin{aligned} 1+2^3 &= 3^2, & 2^5+7^2 &= 3^4, \\ 7^3+13^2 &= 2^9, & 2^7+17^3 &= 71^2, \\ 3^5+11^4 &= 122^2, & 1414^3+2213459^2 &= 65^7, \\ 17^7+76271^3 &= 21063928^2, & 9262^3+15312283^2 &= 113^7, \\ 43^8+96222^3 &= 30042907^2, & 33^8+1549034^2 &= 15613^3. \end{aligned}$$

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*

$$x^{2\ell} + y^{2m} = z^p \tag{2}$$

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 ζ 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 ℓ representations of these Frey curves using modularity and level lowering. Several recent papers [Dieulefait and

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 ℓ). 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 $\mathfrak{q}$ 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*

$$\prod_{\mathfrak{q}\in S} \mathrm{Norm}_{\mathbb{F}_\mathfrak{q}/\mathbb{F}_5}(u \bmod \mathfrak{q}) \neq \bar{1}. \tag{3}$$

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 ℓ representation, one must first show that this representation is irreducible. Let $G_K = \mathrm{Gal}(\overline{K}/K)$. The mod ℓ representation that concerns us, denoted $\bar{\rho}_{E,\ell} : G_K \to \mathrm{GL}_2(\mathbb{F}_\ell)$, is the one attached to the ℓ-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 $\bar{\rho}_{E,\ell}$ is reducible then E or some ℓ-isogenous curve possesses nontrivial K-rational ℓ-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. [$\geq$ 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 $2\mathcal{O}_K$, $2\mathfrak{p}$. 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 $\mathfrak{q}$ 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) *φ is unramified outside S and the infinite places of K*;

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

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

$$\prod_{\mathfrak{q} \in S} \mathrm{Norm}_{\mathbb{F}_\mathfrak{q}/\mathbb{F}_q}(u \bmod \mathfrak{q}) = \bar{1}.$$

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 φ. Then we may view φ as a character $\mathrm{Gal}(L/K) \to \mathbb{F}_q^\times$. Write M_K for the places of K. For $\upsilon \in M_K$, let $\Theta_\upsilon : K_\upsilon^* \to \mathrm{Gal}(L/K)$ be the local Artin map. Let $u \in \mathcal{O}_K$ be a unit that is positive in each real embedding. We consider the values $\varphi(\Theta_\upsilon(u)) \in \mathbb{F}_q^\times$ as υ ranges over M_K.

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

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

Finally, by global reciprocity, $\prod_{\upsilon \in M_K} \Theta_\upsilon(u) = 1$. Applying φ 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*

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

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

(i) *the splitting field $K(\psi_1/\psi_2)$ of ψ_1/ψ_2 is abelian over $\mathbb{Q}$;*

(ii) $(\psi_1/\psi_2)(\tau) = -1$ *for each complex conjugation τ;*

(iii) $(\psi_1/\psi_2)|_{D_\mathfrak{q}} \neq 1$ *for each $\mathfrak{q} \mid q$;*

(iv) *for all* $\mathfrak{q} \mid q$,

$$\rho|_{D_{\mathfrak{q}}} \sim \begin{pmatrix} \phi_1^{(\mathfrak{q})} \cdot \tilde{\psi}_1 & * \\ 0 & \phi_2^{(\mathfrak{q})} \cdot \tilde{\psi}_2 \end{pmatrix}$$

with $\phi_2^{(\mathfrak{q})}$ *factoring through a pro-*q *extension of* $K_{\mathfrak{q}}$ *and* $\phi_2^{(\mathfrak{q})}|_{I_{\mathfrak{q}}}$ *having finite order, and where* $\tilde{\psi}_i$ *is a Teichmüller lift of* ψ_i;

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

Then the representation ρ *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 $\bar{\rho}^{\mathrm{ss}} = \psi_1 \oplus \psi_2$. We verify hypotheses (i)–(v) in the theorem of Skinner and Wiles (with $q = 5$) to deduce the modularity of $\rho : G_K \to \mathrm{Aut}(T_5(E)) \cong \mathrm{GL}_2(\mathbb{Z}_5)$, where $T_5(E)$ is the 5-adic Tate module of E. If E has good supersingular reduction at some $\mathfrak{q} \mid 5$ then (as $\mathfrak{q}$ is unramified) $\bar{\rho}|_{I_{\mathfrak{q}}}$ is irreducible [Serre 1972, Proposition 12], contradicting the reducibility of $\bar{\rho}$. It follows that E has good ordinary or multiplicative reduction at all $\mathfrak{q} \mid 5$. In particular, hypothesis (iv) holds with $\phi_i^{(\mathfrak{q})} = 1$.

Now $\psi_1\psi_2 = \det(\rho) = \chi_5$ so hypothesis (v) holds with $\psi = 1$ and $k = 2$. Moreover, for each complex conjugation τ, 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 $\mathfrak{q} \mid 5$, that $(\bar{\rho}|_{I_{\mathfrak{q}}})^{\mathrm{ss}} = \chi_5|_{I_{\mathfrak{q}}} \oplus 1$ and so ψ_1/ψ_2 is nontrivial when restricted to $I_{\mathfrak{q}}$ (again as $\mathfrak{q}$ 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 ψ_1 and ψ_2 are unramified at all finite primes $\mathfrak{p} \nmid 5$. Let S be the subset of $\mathfrak{q} \in S_5$ such that ψ_1 is unramified at $\mathfrak{q}$. By the above, we know that these are precisely the $\mathfrak{q} \in S_5$ such that $\psi_2|_{I_{\mathfrak{q}}} = \chi_5|_{I_{\mathfrak{q}}}$. By assumption (c) and Proposition 2.1, we have that either $S = \varnothing$ or $S = S_5$. If $S = \varnothing$ then ψ_2 is unramified at all $\mathfrak{q} \mid 5$, and if $S = S_5$ then ψ_1 is unramified at all $\mathfrak{q} \mid 5$. 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 ζ 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

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

Let $\mathcal{O}_K$ be the ring of integers of K. Let $\mathfrak{p}$ be the unique prime ideal of K above p. Then $p\mathcal{O}_K = \mathfrak{p}^{(p-1)/2}$.

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

$$\theta_j \in \mathcal{O}_K^\times, \qquad \theta_j + 2 \in \mathcal{O}_K^\times, \qquad (\theta_j - 2)\mathcal{O}_K = \mathfrak{p}.$$

Moreover, $(\theta_j - \theta_k)\mathcal{O}_K = \mathfrak{p}$ for $1 \le j < k \le \frac{1}{2}(p-1)$.

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 $\theta_j + 2 = \theta_r^2 \in \mathcal{O}_K^\times$.

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

For the final part, $j \not\equiv \pm k \pmod p$. Thus there exist $u, v \not\equiv 0 \pmod p$ such that

$$u + v \equiv j, \qquad u - v \equiv k \pmod p.$$

Then

$$(\zeta^u - \zeta^{-u})(\zeta^v - \zeta^{-v}) = \theta_j - \theta_k,$$

and so $(\theta_j - \theta_k)\mathcal{O}_L = \mathfrak{P}^2 = \mathfrak{p}\mathcal{O}_L$. This completes the proof. □

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^ℓ 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 $x^\ell + y^m i$ and $x^\ell - y^m i$ are coprime. There exist coprime rational integers a, b such that

$$x^\ell + y^m i = (a + bi)^p, \qquad z = a^2 + b^2.$$

Then

$$\begin{aligned} x^\ell &= \tfrac{1}{2}((a+bi)^p + (a-bi)^p) \\ &= a \cdot \prod_{j=1}^{p-1}((a+bi) + (a-bi)\zeta^j) \\ &= a \cdot \prod_{j=1}^{(p-1)/2} ((a+bi) + (a-bi)\zeta^j) \cdot ((a+bi) + (a-bi)\zeta^{-j}). \end{aligned}$$

In the last step we have paired up the complex conjugate factors. Multiplying out these pairs we obtain a factorization of x^ℓ over $\mathcal{O}_K$:

$$x^\ell = a \cdot \prod_{j=1}^{(p-1)/2} ((\theta_j + 2)a^2 + (\theta_j - 2)b^2). \tag{4}$$

To ease notation, write

$$\beta_j = (\theta_j + 2)a^2 + (\theta_j - 2)b^2, \quad j = 1, \ldots, \tfrac{1}{2}(p-1). \tag{5}$$

Lemma 4.1. *Write $n = \upsilon_2(x) \geq 1$.*

(i) *If $p \nmid x$ then*

$$a = 2^{\ell n}\alpha^\ell, \qquad \beta_j \mathcal{O}_K = \mathfrak{b}_j^\ell,$$

where α is a rational integer and $\alpha\mathcal{O}_K, \mathfrak{b}_1, \ldots, \mathfrak{b}_{(p-1)/2}$ are pairwise coprime ideals of $\mathcal{O}_K$, all of which are coprime to $2p$.

(ii) *If $p \mid x$ then*

$$a = 2^{\ell n} p^{\kappa\ell-1}\alpha^{\ell}, \qquad \beta_j \mathcal{O}_K = \mathfrak{p}\cdot \mathfrak{b}_j^{\ell},$$

where $\kappa = \upsilon_p(x) \geq 1$, α is a rational integer and $\alpha\cdot\mathcal{O}_K, \mathfrak{b}_1, \ldots, \mathfrak{b}_{(p-1)/2}$ 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 β_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, β_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 $\mathfrak{p} \parallel \beta_j$. However, since $(\theta_j - 2)\mathcal{O}_K = \mathfrak{p}$ and $\theta_j + 2 \in \mathcal{O}_K^{\times}$, reducing (4) modulo $\mathfrak{p}$ shows that $a^p \equiv 0 \pmod{\mathfrak{p}}$, and hence that $p \mid a$. Since a, b are coprime, it follows that $\upsilon_{\mathfrak{p}}(\beta_j) = 1$. Now, from (4),

$$\tfrac{1}{2}(p-1)\,\upsilon_p(a) = \upsilon_{\mathfrak{p}}(a) = \ell\,\upsilon_{\mathfrak{p}}(x) - \sum_{j=1}^{(p-1)/2} \upsilon_{\mathfrak{p}}(\beta_j) = \tfrac{1}{2}(p-1)(\kappa\ell-1),$$

giving the desired exponent of p in the factorization of a. □

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,

$$a = 2^{\ell n}\alpha^{\ell}, \qquad a^2 - 3b^2 = \gamma^{\ell}$$

for some coprime odd rational integers α and γ. 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,

$$a = 2^{\ell n}3^{\kappa\ell-1}\alpha^{\ell}, \qquad a^2 - 3b^2 = 3\gamma^{\ell}$$

for coprime rational integers α, γ 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 (α, γ, b) whose mod ℓ 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 β_j, β_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, β_j, β_k are roughly ℓ-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 (ℓ, ℓ, ℓ). 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 ℓ representation has very little ramification. In what follows we take care to scale the expressions a^2, β_j, β_k appropriately so that the Frey curve is semistable.

Case I: $p \nmid x$. Let

$$u = \beta_j, \qquad v = -\frac{(\theta_j - 2)}{(\theta_k - 2)}\beta_k, \qquad w = \frac{4(\theta_j - \theta_k)}{(\theta_k - 2)} \cdot a^2. \tag{6}$$

Then $u + v + w = 0$. Moreover, by Lemmas 3.1 and 4.1,

$$u\mathcal{O}_K = \mathfrak{b}_j^{\ell}, \qquad v\mathcal{O}_K = \mathfrak{b}_k^{\ell}, \qquad w\mathcal{O}_K = 2^{2\ell n+2} \cdot \alpha^{2\ell}\mathcal{O}_K.$$

We let the Frey curve be

$$E = E_{j,k} : Y^2 = X(X-u)(X+v). \tag{7}$$

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

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 $\mathfrak{p}$. It has minimal discriminant and conductor*

$$\mathcal{D}_{E/K} = 2^{4\ell n - 4}\alpha^{4\ell}\mathfrak{b}_j^{2\ell}\mathfrak{b}_k^{2\ell}, \qquad \mathcal{N}_{E/K} = 2 \cdot \mathrm{Rad}(\alpha \mathfrak{b}_j \mathfrak{b}_k).$$

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

$$\begin{gathered} c_4 = 16(u^2 - vw) = 16(v^2 - wu) = 16(w^2 - uv), \\ c_6 = -32(u-v)(v-w)(w-u), \quad \Delta = 16u^2v^2w^2, \quad j(E) = c_4^3/\Delta. \end{gathered} \tag{8}$$

By Lemma 4.1, we have that $\alpha\mathcal{O}_K$, $\mathfrak{b}_j$ and $\mathfrak{b}_k$ are pairwise coprime, and all coprime to $2p$. In particular $\mathfrak{p} \nmid \Delta$ and so E has good reduction at $\mathfrak{p}$. Moreover, c_4 and Δ 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

$$u \equiv (\theta_j - 2)b^2 \pmod{2^{2\ell}}, \qquad v \equiv -(\theta_j - 2)b^2 \pmod{2^{2\ell}}, \qquad w \equiv 0 \pmod{2^{2\ell+2}}.$$

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

$$\tfrac{1}{16} c_4 = (u^2 - vw) \equiv (\theta_j - 2)^2 \cdot b^4 \pmod{2^{2\ell}},$$

which shows that c_4 is a square in $K_{\mathfrak{q}}$. Moreover,

$$-\tfrac{1}{16} c_6 = 2(u-v)(v-w)(w-u) \equiv 4 \cdot (\theta_j - 2)^3 \cdot b^6 \pmod{2^{2\ell+1}}.$$

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

Finally 2 is unramified in K, and so $\upsilon_{\mathfrak{q}}(c_4) = \upsilon_2(16) = 4$. It follows that $\mathcal{D}_{E/K} = (\Delta/2^{12}) \cdot \mathcal{O}_K$, as required. □

Case II: $p \mid x$. Let

$$u = \frac{\beta_j}{(\theta_j - 2)}, \qquad v = -\frac{\beta_k}{(\theta_k - 2)}, \qquad w = \frac{4(\theta_j - \theta_k)}{(\theta_j - 2)(\theta_k - 2)} \cdot a^2. \tag{9}$$

Then, from Lemmas 3.1 and 4.1,

$$u\mathcal{O}_K = \mathfrak{b}_j^\ell, \qquad v\mathcal{O}_K = \mathfrak{b}_k^\ell, \qquad w\mathcal{O}_K = 2^{2\ell n+2} \cdot \mathfrak{p}^\delta \cdot \alpha^{2\ell} \mathcal{O}_K,$$

where

$$\delta = (\kappa\ell - 1)(p-1) - 1. \tag{10}$$

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 $\mathfrak{p}$ and at all primes above 2. It has minimal discriminant and conductor*

$$\mathcal{D}_{E/K} = 2^{4\ell n - 4} \mathfrak{p}^{2\delta} \alpha^{4\ell} \mathfrak{b}_j^{2\ell} \mathfrak{b}_k^{2\ell}, \qquad \mathcal{N}_{E/K} = 2\mathfrak{p} \cdot \mathrm{Rad}(\alpha \mathfrak{b}_j \mathfrak{b}_k).$$

Proof. The proof is an easy modification of the proof of Lemma 6.1. □

7. A closer look at the Frey curve for $p \equiv 1 \pmod 4$

In this section we suppose that $p \equiv 1 \pmod 4$. 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 $\tau \in \mathrm{Gal}(K/\mathbb{Q})$, 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 $1 \le j, k \le \frac{1}{2}(p-1)$ with $j \ne k$. 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

$$u' = (\theta_k - 2)\beta_j, \qquad v' = -(\theta_j - 2)\beta_k, \qquad w' = 4(\theta_j - \theta_k) \cdot a^2, \tag{11}$$

and let

$$E' : Y^2 = X(X - u')(X + v').$$

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

$$\mathfrak{b}_{j,k} = \mathrm{Norm}_{K/K'}(\mathfrak{b}_j) = \mathrm{Norm}_{K/K'}(\mathfrak{b}_k).$$

It follows from Lemma 4.1 that the $\mathcal{O}_{K'}$-ideal $\mathfrak{b}_{j,k}$ is coprime to α 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 $\mathfrak{B}$, with minimal discriminant and conductor*

$$\mathcal{D}_{E'/K'} = 2^{4\ell n - 4} \mathfrak{B}^3 \alpha^{4\ell} \mathfrak{b}_{j,k}^{2\ell}, \qquad \mathcal{N}_{E'/K'} = 2 \cdot \mathfrak{B}^2 \cdot \mathrm{Rad}(\alpha \mathfrak{b}_{j,k}).$$

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*

$$\mathcal{D}_{E'/K'} = 2^{4\ell n - 4} \mathfrak{B}^\delta \alpha^{4\ell} \mathfrak{b}_{j,k}^{2\ell}, \qquad \mathcal{N}_{E'/K'} = 2 \cdot \mathfrak{B} \cdot \mathrm{Rad}(\alpha \mathfrak{b}_{j,k}),$$

where δ 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 $\ell, m \ge C_p^{(1)}$ 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 $j_1, \ldots, j_r$ be the values of these j-invariants. Let $\mathfrak{q}$ be a prime of K above 2. By Lemmas 6.1 and 6.2, we have $\upsilon_{\mathfrak{q}}(j(E)) = -(4\ell n - 4)$ with $n \geq 1$. Thus for ℓ, m sufficiently large we have $\upsilon_{\mathfrak{q}}(j(E)) < \upsilon_{\mathfrak{q}}(j_i)$ 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 ≥ 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 ℓ 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 ℓ 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 $\lambda \mid \ell$, the mod ℓ representation $\bar{\rho}_{E,\ell} : G_K \to \mathrm{GL}_2(\mathbb{F}_\ell)$ 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 $\bar{\rho}_{E,\ell} \sim \bar{\rho}_{\mathfrak{f},\lambda}$ for some Hilbert cuspidal eigenform $\mathfrak{f}$ over K of parallel weight 2 that is new at level $\mathcal{N}_\ell$, where*

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

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

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 $\bar{\rho}_{E',\ell} \sim \bar{\rho}_{\mathfrak{f},\lambda}$ for some Hilbert cuspidal eigenform $\mathfrak{f}$ over K of parallel weight 2 that is new at level $\mathcal{N}'_\ell$, where

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

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 $\bar{\rho}_{E,\ell}$ is irreducible and modular, and it follows from Lemma 8.3 that $\bar{\rho}_{E,\ell} \sim \bar{\rho}_{\mathfrak{f},\lambda}$ for some Hilbert eigenform over K of parallel weight 2 that is new at level $\mathcal{N}_\ell$, where $\mathcal{N}_\ell = 2\mathcal{O}_K$ or $2\mathfrak{p}$. 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 $\mathfrak{f}$ is $\mathbb{Q}$. Observe that the level $\mathcal{N}_\ell$ is nonsquarefull (meaning there is a prime $\mathfrak{q}$ at which the level has valuation 1). As the level is nonsquarefull and the field of eigenvalues is $\mathbb{Q}$, the eigenform $\mathfrak{f}$ is known to correspond to some elliptic curve E_1/K of conductor $\mathcal{N}_\ell$ [Blasius 2004], and $\bar{\rho}_{E,\ell} \sim \bar{\rho}_{E_1,\ell}$. 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 $\bar{\rho}_{E,\ell} \sim \bar{\rho}_{E_1,\ell}$ 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 $2\mathcal{O}_K$, $2\mathfrak{p}$, 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 ζ 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 $\bar{\rho}_{E,\ell}$ 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 $\bar{\rho}_{E,\ell}$ for $\ell \geq 5$; equivalently, we need to show that E does not have an ℓ-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 τ 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 $\mathfrak{q} \mid 2$ be a prime of K'. Then $\upsilon_{\mathfrak{q}}(j(E')) = 4 - 4\ell n$, which is negative but not divisible by ℓ. Thus $\bar{\rho}_{E',\ell}(G_{K'})$ contains an element of order ℓ [Silverman 1994, Proposition V.6.1]. By Dickson's classification [Swinnerton-Dyer 1973] of subgroups of $\mathrm{GL}_2(\mathbb{F}_\ell)$ we see that $\bar{\rho}_{E',\ell}(G_{K'})$ must contain $\mathrm{SL}_2(\mathbb{F}_\ell)$. 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 ℓ-torsion, or is ℓ-isogenous to an elliptic curve defined over K that has nontrivial ℓ-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 ψ_1 or ψ_2 is trivial. It follows in the former case that E has nontrivial ℓ-torsion, and in the latter case that the K-isogenous curve $E/\operatorname{Ker}(\psi_1)$ has nontrivial ℓ-torsion.

As K has narrow class number 1 for $p = 5, 7, 11, 13$, it is sufficient to show that one of ψ_1, ψ_2 is unramified at all finite places. As E is semistable, the characters ψ_1 and ψ_2 are unramified away from ℓ and the infinite places. Let S_ℓ 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 ψ_1 is ramified at λ. Then ψ_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_ℓ. Fix $\lambda \in S$ and let

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

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

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

As S is a nonempty proper subset of S_ℓ, we have that T' is a nonempty proper subset of T. Now, by Proposition 2.1, for any totally positive unit u of $\mathcal{O}_K$,

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

But

$$\begin{aligned}
\mathrm{Norm}_{\mathbb{F}_\lambda/\mathbb{F}_\ell}(u + \lambda^{\tau^{-1}}) &= \prod_{\sigma \in D} (u + \lambda^{\tau^{-1}})^\sigma \\
&= \prod_{\sigma \in D} (u^\sigma + \lambda^{\tau^{-1}}) \\
&= \Big(\prod_{\sigma \in D} (u^{\sigma\tau} + \lambda) \Big)^{\tau^{-1}} \\
&= \prod_{\sigma \in D} (u^{\sigma\tau} + \lambda) \qquad \text{(as this expression belongs to } \mathbb{F}_\ell\text{)}.
\end{aligned}$$

Let

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

It follows that $\ell \mid B_{T',D}(u)$. Now let $u_1, \ldots, u_d$ be a system of totally positive units. Then ℓ 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 ℓ, then there is some subgroup D of G and some nonempty proper subset T' of G/D such that ℓ 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 $\mathcal{O}_K$. We run through the subgroups D of $G = \mathrm{Gal}(K/\mathbb{Q})$. 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 $B_{T',D}(u_1, \dots, u_d)$ 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 $S \subset S_\ell$ and check that there is some totally positive unit u such that $\prod_{\lambda \in S} \operatorname{Norm}(u + \lambda) \neq \bar{1}$. This completes the proof. $\square$

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 ℓ-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 ℓ, 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:

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

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

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

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

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 $\ell \le 5$. However we are assuming throughout that $\ell \ge 5$ 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. $\square$

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 $\operatorname{Ker}(\phi) \cap \langle Q \rangle = 0$, 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 ℓ on E_1 survives the isogeny, and so E_2 has a K-rational cyclic subgroup of order 4ℓ.

Proof. In this case K is a cubic field. By the classification of cyclic ℓ-torsion on elliptic curves over cubic fields [Parent 2000; 2003], we know $\ell \leq 13$. Since $\ell \neq p$, we need only deal with the case $\ell = 5, 11, 13$. To eliminate $\ell = 5$ and $\ell = 11$ we computed the K-points on the modular curves $X_0(20)$ and $X_0(11)$. These both have rank 0 and their K-points are in fact defined over $\mathbb{Q}$. The $\mathbb{Q}$-points of $X_0(20)$ are cuspidal thus $\ell \neq 5$. The three noncuspidal $\mathbb{Q}$-points on $X_0(11)$ all have integral j-invariants. As our curve E has multiplicative reduction at $2\mathcal{O}_K$, it follows that $\ell \neq 11$.

We suppose $\ell = 13$. We now apply [Bruin and Najman 2016, Theorem 1]. That theorem gives a useful and practical criterion for ruling out the existence of torsion subgroups $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ on elliptic curves over a given number field K (the remarks at the end of Section 2 of [Bruin and Najman 2016] are useful when applying that theorem). The theorem involves making certain choices and we indicate them briefly; in the notation of the theorem, we take $A = \mathbb{Z}/26\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 26$, $X = X' = X_1(26)$, $p = \mathfrak{p}_0 = 7$. 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. $\square$

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

Proof. Now K has degree 5. By the classification of cyclic ℓ-torsion on elliptic curves over quintic fields [Derickx et al. $\geq$ 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 $A = \mathbb{Z}/26\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 26$, $X = X' = X_1(26)$, $p = \mathfrak{p}_0 = 11$ (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 $A = \mathbb{Z}/34\mathbb{Z}$, $L = \mathbb{Q}$, $m = 1$, $n = 34$, $X = X' = X_1(34)$, $p = \mathfrak{p}_0 = 11$. 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. $\square$

It remains to deal with $p = 13$. Unfortunately the field K in this case is sextic, and the known bound [Derickx et al. $\geq$ 2016] for cyclic ℓ-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 $\bar{\rho}_{E',\ell}$, where E' is the Frey curve defined over the degree 3 subfield K'.

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

Proof. Suppose $\bar{\rho}_{E',\ell}$ 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 ℓ-torsion, or is ℓ-isogenous to an elliptic curve with nontrivial ℓ-torsion. Thus there is an elliptic curve over K' with a point of exact order 2ℓ. 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 $\bar{\rho}_{E',\ell}$ is reducible, we have that $\bar{\rho}_{E,\ell}$ is reducible (Lemma 10.1). Now we may apply Lemma 10.2 to deduce the existence of E_1/K (which is E or ℓ-isogenous to it) that has a subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\ell\mathbb{Z}$. As before, let $\mathfrak{p}$ be the unique prime of K above 13. By Lemma 6.1 the Frey curve E has good reduction at $\mathfrak{p}$. As $\mathfrak{p} \nmid 2\ell$, we know from injectivity of torsion that $4\ell \mid \#E(\mathbb{F}_\mathfrak{p})$. But $\mathbb{F}_\mathfrak{p} = \mathbb{F}_{13}$. 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 $\upsilon_2(a) = 5n$, where $n > 0$. Write $a = 2^{5n}a'$, where a' is odd. We know that $\upsilon_2(j(E)) = -(20n - 4)$. 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 τ 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 $\bar{\rho}_{\mathcal{E},\ell} \sim \bar{\rho}_{\mathfrak{f},\lambda}$, 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 $\mathfrak{q}$ a prime of $\mathcal{K}$ such that $\mathfrak{q} \nmid 2p\ell$. In particular, $\mathfrak{q}$ does not divide the level of $\mathfrak{f}$, and $\mathcal{E}$ has good or multiplicative reduction at $\mathfrak{q}$. Write $\sigma_{\mathfrak{q}}$ for a Frobenius element of $G_{\mathcal{K}}$ at $\mathfrak{q}$. Comparing the traces of $\bar{\rho}_{\mathcal{E},\ell}(\sigma_{\mathfrak{q}})$ and $\bar{\rho}_{\mathfrak{f},\lambda}(\sigma_{\mathfrak{q}})$ we obtain

(i) if $\mathcal{E}$ has good reduction at $\mathfrak{q}$ then $a_{\mathfrak{q}}(\mathcal{E}) \equiv a_{\mathfrak{q}}(\mathfrak{f}) \pmod{\lambda}$;

(ii) if $\mathcal{E}$ has split multiplicative reduction at $\mathfrak{q}$ then $\mathrm{Norm}(\mathfrak{q}) + 1 \equiv a_{\mathfrak{q}}(\mathfrak{f}) \pmod{\lambda}$;

(iii) if $\mathcal{E}$ has nonsplit multiplicative reduction at $\mathfrak{q}$ then $-(\mathrm{Norm}(\mathfrak{q}) + 1) \equiv a_{\mathfrak{q}}(\mathfrak{f}) \pmod{\lambda}$.

[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.

<table>
<tr><th>p</th><th>case</th><th>field $\mathcal{K}$</th><th>Frey curve $\mathcal{E}$</th><th>level $\mathcal{N}$</th><th>eigenforms $\mathfrak{f}$</th><th>$[\mathbb{Q}_\mathfrak{f} : \mathbb{Q}]$</th></tr>
<tr><td rowspan="2">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>$5 \mid x$</td><td>K</td><td>E</td><td>$2\mathfrak{p}$</td><td>–</td><td>–</td></tr>
<tr><td rowspan="2">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>$7 \mid x$</td><td>K</td><td>E</td><td>$2\mathfrak{p}$</td><td>$\mathfrak{f}_1$</td><td>1</td></tr>
<tr><td rowspan="2">11</td><td>$11 \nmid x$</td><td>K</td><td>E</td><td>$2\mathcal{O}_K$</td><td>$\mathfrak{f}_2$</td><td>2</td></tr>
<tr><td>$11 \mid x$</td><td>K</td><td>E</td><td>$2\mathfrak{p}$</td><td>$\mathfrak{f}_3, \mathfrak{f}_4$</td><td>5</td></tr>
<tr><td rowspan="2">13</td><td>$13 \nmid x$</td><td>K</td><td>E</td><td>$2\mathcal{O}_K$</td><td>$\mathfrak{f}_5, \mathfrak{f}_6$
$\mathfrak{f}_7$
$\mathfrak{f}_8$</td><td>1
2
3</td></tr>
<tr><td>$13 \mid x$</td><td>K'</td><td>E'</td><td>$2\mathfrak{B}$</td><td>$\mathfrak{f}_9, \mathfrak{f}_{10}$
$\mathfrak{f}_{11}, \mathfrak{f}_{12}$</td><td>1
3</td></tr>
</table>

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

Let $q \nmid 2p\ell$ be a rational prime and let

$$\mathcal{A}_q = \{(\eta, \mu) : 0 \le \eta, \mu \le q-1,\ (\eta, \mu) \neq (0,0)\}.$$

For $(\eta, \mu) \in \mathcal{A}_q$ let

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

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

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 $(\eta, \mu) \in \mathcal{A}_q$. In particular, $(a, b) \equiv (\eta, \mu) \pmod{\mathfrak{q}}$ for all primes $\mathfrak{q} \mid q$ of $\mathcal{K}$. From the definitions of the Frey curves E and E' in Sections 6 and 7 we see that $\mathcal{E}$ has good reduction at $\mathfrak{q}$ if and only if

$\mathfrak{q} \nmid \Delta(\eta, \mu)$, and in this case $a_{\mathfrak{q}}(\mathcal{E}) = a_{\mathfrak{q}}(E_{(\eta,\mu)})$. Let

$$B_{\mathfrak{q}}(\mathfrak{f}, \eta, \mu) = \begin{cases} a_{\mathfrak{q}}(E_{(\eta,\mu)}) - a_{\mathfrak{q}}(\mathfrak{f}) & \text{if } \mathfrak{q} \nmid \Delta(\eta, \mu), \\ \mathrm{Norm}(\mathfrak{q}) + 1 - a_{\mathfrak{q}}(\mathfrak{f}) & \text{if } \mathfrak{q} \mid \Delta(\eta, \mu) \text{ and } \overline{\gamma(\eta, \mu)} \in (\mathbb{F}_{\mathfrak{q}}^*)^2, \\ \mathrm{Norm}(\mathfrak{q}) + 1 + a_{\mathfrak{q}}(\mathfrak{f}) & \text{if } \mathfrak{q} \mid \Delta(\eta, \mu) \text{ and } \overline{\gamma(\eta, \mu)} \notin (\mathbb{F}_{\mathfrak{q}}^*)^2. \end{cases}$$

From (i)–(iii) above we see that $\lambda \mid B_{\mathfrak{q}}(\mathfrak{f}, \eta, \mu)$. Now let

$$B_q(\mathfrak{f}, \eta, \mu) = \sum_{\mathfrak{q} \mid q} B_{\mathfrak{q}}(\mathfrak{f}, \eta, \mu) \cdot \mathcal{O}_{\mathfrak{f}},$$

where $\mathcal{O}_{\mathfrak{f}}$ is the ring of integers of $\mathbb{Q}_{\mathfrak{f}}$. Since $(a, b) \equiv (\eta, \mu) \pmod{\mathfrak{q}}$ for all $\mathfrak{q} \mid q$, we have that $\lambda \mid B_q(\mathfrak{f}, \eta, \mu)$. Now (η, μ) is some unknown element of $\mathcal{A}_q$. Let

$$B'_q(\mathfrak{f}) = \prod_{(\eta,\mu) \in \mathcal{A}_q} B_q(\mathfrak{f}, \eta, \mu).$$

Then $\lambda \mid B'_q(\mathfrak{f})$. Previously, we have supposed that $q \nmid 2p\ell$. This is inconvenient as ℓ is unknown. Now we simply suppose $q \nmid 2p$, and let $B_q(\mathfrak{f}) = q B'_q(\mathfrak{f})$. Then, since $\lambda \mid \ell$, we certainly have that $\lambda \mid B_q(\mathfrak{f})$ regardless of whether $q = \ell$ or not.

Finally, if $S = \{q_1, q_2, \dots, q_r\}$ is a set of rational primes with $q_i \nmid 2p$, then λ divides the $\mathcal{O}_{\mathfrak{f}}$-ideal $\sum_{i=1}^r B_{q_i}(\mathfrak{f})$, and thus ℓ divides $B_S(\mathfrak{f}) = \mathrm{Norm}\left(\sum_{i=1}^r B_{q_i}(\mathfrak{f})\right)$. Table 2 gives our choices for the set S and the corresponding value of $B_S(\mathfrak{f})$ for each of the eigenforms $\mathfrak{f}_1, \dots, \mathfrak{f}_{12}$ appearing in Table 1. Recalling that $\ell \geq 5$ and $\ell \neq p$ gives a contradiction unless $p = 13$ and $\ell = 7$. 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 $\mathfrak{f}_9$ and $\mathfrak{f}_{11}$. Consider $(\eta, \mu) = (0, 1)$, which belongs to $\mathcal{A}_q$ for any q. The corresponding Weierstrass model $E_{(0,1)}$ is singular with a split note. It follows that

$$B_{\mathfrak{q}}(\mathfrak{f}, 0, 1) = \mathrm{Norm}(\mathfrak{q}) + 1 - a_{\mathfrak{q}}(\mathfrak{f}).$$

Note that if λ is a prime of $\mathbb{Q}_{\mathfrak{f}}$ that divides $\mathrm{Norm}(\mathfrak{q}) + 1 - a_{\mathfrak{q}}(\mathfrak{f})$ for all $\mathfrak{q} \nmid 26$, then ℓ will divide $B_S(\mathfrak{f})$ for any set S where $\lambda \mid \ell$. This appears to be the case with $\ell = 7$ for $\mathfrak{f}_{11}$, and we now show that it is indeed the case for $\mathfrak{f}_9$. Let F be the elliptic curve with Cremona label `26b1`:

$$F : y^2 + xy + y = x^3 - x^2 - 3x + 3,$$

which has conductor $2\mathfrak{B}$ as an elliptic curve over $\mathcal{K}$. As $\mathcal{K}/\mathbb{Q}$ is cyclic, we know that F is modular over $\mathcal{K}$ and hence corresponds to a Hilbert modular form of parallel weight 2 and level $2\mathfrak{B}$, and by comparing eigenvalues we can show that it in fact corresponds to $\mathfrak{f}_9$. Now the point $(1, 0)$ on F has order 7. It follows that $7 \mid \#E(\mathbb{F}_{\mathfrak{q}}) = \mathrm{Norm}(\mathfrak{q}) + 1 - a_{\mathfrak{q}}(\mathfrak{f})$ for all $\mathfrak{q} \nmid 26$, showing that for $\mathfrak{f}_9$ we can never

p	case	S	eigenform $\mathfrak{f}$	$B_S(\mathfrak{f})$
7	$7 \mid x$	$\{3\}$	$\mathfrak{f}_1$	$2^8 \times 3^5 \times 7^6$
11	$11 \nmid x$	$\{23, 43\}$	$\mathfrak{f}_2$	1
	$11 \mid x$	$\{23, 43\}$	$\mathfrak{f}_3$	1
			$\mathfrak{f}_4$	1
13	$13 \nmid x$	$\{79, 103\}$	$\mathfrak{f}_5$	$2^{6240} \times 3^{312}$
			$\mathfrak{f}_6$	$2^{12792} \times 3^{234}$
			$\mathfrak{f}_7$	$2^{10608} \times 3^{624}$
			$\mathfrak{f}_8$	$2^{18720} \times 3^{936}$
	$13 \mid x$	$\{3, 5, 31, 47\}$	$\mathfrak{f}_9$	7^2
			$\mathfrak{f}_{10}$	3^7
			$\mathfrak{f}_{11}$	7^6
			$\mathfrak{f}_{12}$	1

Table 2. Our choice of set of primes S and the value of $B_S(\mathfrak{f})$ for each of the eigenforms in Table 1.

eliminate $\ell = 7$ by enlarging the set S. We can still complete the contradiction in this case as follows. Note that $\bar{\rho}_{\mathfrak{f}_9,7} \sim \bar{\rho}_{F,7}$ which is reducible. As $\bar{\rho}_{\mathcal{E},7}$ is irreducible we have $\bar{\rho}_{\mathcal{E},7} \not\sim \bar{\rho}_{\mathfrak{f}_9,7}$, completing the contradiction for $\mathfrak{f} = \mathfrak{f}_9$. We strongly suspect that reducibility of $\bar{\rho}_{\mathfrak{f}_{11},\lambda}$ (where λ is the unique prime above 7 of $\mathbb{Q}_{\mathfrak{f}_{11}}$), 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 ℓ, for larger values of p provided the eigenforms $\mathfrak{f}$ at the relevant levels can be computed. The usual obstruction to bounding the exponent (see [Siksek 2012, Section 9]) comes from eigenforms $\mathfrak{f}$ that correspond to elliptic curves with a torsion structure that matches the Frey curve $\mathcal{E}$. Let $\mathfrak{f}$ be such an eigenform. Let $q \nmid 2p$ be a rational prime and $\mathfrak{q}_1, \ldots, \mathfrak{q}_r$ be the primes of $\mathcal{K}$ above q. Note that $\mathrm{Norm}(\mathfrak{q}_1) = \cdots = \mathrm{Norm}(\mathfrak{q}_r) = q^{d/r}$, where $d = [\mathcal{K} : \mathbb{Q}]$. We would like to estimate the "probability" that $B_q(\mathfrak{f})$ is nonzero. Observe that if $B_q(\mathfrak{f})$ is nonzero, then we obtain a bound for ℓ. Examining the definitions above shows that the ideal $B_q(\mathfrak{f})$ is 0 if and only if there is some $(\eta, \mu) \in \mathcal{A}_q$ such that $a_{\mathfrak{q}}(E_{(\eta,\mu)}) = a_{\mathfrak{q}}(\mathfrak{f})$ for $\mathfrak{q} = \mathfrak{q}_1, \mathfrak{q}_2, \ldots, \mathfrak{q}_r$. Treating $a_{\mathfrak{q}}(E_{(\eta,\mu)})$ as a random variable belonging to the Hasse interval $[-2q^{d/2r}, 2q^{d/2r}]$, we see that the "probability" that $a_{\mathfrak{q}}(E_{(\eta,\mu)}) = a_{\mathfrak{q}}(\mathfrak{f})$ is roughly $c/q^{d/2r}$, with $c = \frac{1}{4}$.

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 $\mathcal{E}$ has full 2-torsion (i.e., $\mathcal{E}$ is the Frey curve E) and take $c = \frac{1}{2}$ if $\mathcal{E}$ has just one nontrivial point of order 2 (i.e., $\mathcal{E} = E'$ and $p \equiv 1 \pmod 4$). Thus the "probability" that $a_{\mathfrak{q}}(E_{(\eta,\mu)}) = a_{\mathfrak{q}}(\mathfrak{f})$ for all $\mathfrak{q} \mid q$ simultaneously is roughly $c^r/q^{d/2}$. Since $B_q(\mathfrak{f}) = q \prod_{(\eta,\mu)\in\mathcal{A}_q} B_q(\mathfrak{f}, \eta, \mu)$, it follows that the "probability" $\mathbb{P}_q$ (say) that $B_q(\mathfrak{f})$ is nonzero satisfies

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

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 $\mathbb{P}_q \to 1$ as $q \to \infty$. Thus we certainly expect our strategy to succeed in bounding the exponent ℓ.

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

[Barnet-Lamb et al. 2012] T. Barnet-Lamb, T. Gee, and D. Geraghty, "Congruences between Hilbert modular forms: constructing ordinary lifts", *Duke Math. J.* **161**:8 (2012), 1521–1580. MR Zbl

[Barnet-Lamb et al. 2013] T. Barnet-Lamb, T. Gee, and D. Geraghty, "Congruences between Hilbert modular forms: constructing ordinary lifts, II", *Math. Res. Lett.* **20**:1 (2013), 67–72. MR Zbl

[Bennett 2006] M. A. Bennett, "The equation $x^{2n} + y^{2n} = z^5$", *J. Théor. Nombres Bordeaux* **18**:2 (2006), 315–321. MR Zbl

[Bennett and Chen 2012] M. A. Bennett and I. Chen, "Multi-Frey $\mathbb{Q}$-curves and the Diophantine equation $a^2 + b^6 = c^n$", *Algebra Number Theory* **6**:4 (2012), 707–730. MR Zbl

[Bennett and Skinner 2004] M. A. Bennett and C. M. Skinner, "Ternary Diophantine equations via Galois representations and modular forms", *Canad. J. Math.* **56**:1 (2004), 23–54. MR Zbl

[Bennett et al. 2010] M. A. Bennett, J. S. Ellenberg, and N. C. Ng, "The Diophantine equation $A^4 + 2^\delta B^2 = C^n$", *Int. J. Number Theory* **6**:2 (2010), 311–338. MR Zbl

[Bennett et al. 2015a] M. A. Bennett, I. Chen, S. R. Dahmen, and S. Yazdani, "Generalized Fermat equations: a miscellany", *Int. J. Number Theory* **11**:1 (2015), 1–28. MR Zbl

[Bennett et al. 2015b] M. A. Bennett, S. R. Dahmen, M. Mignotte, and S. Siksek, "Shifted powers in binary recurrence sequences", *Math. Proc. Cambridge Philos. Soc.* **158**:2 (2015), 305–329. MR

[Beukers 2012] F. Beukers, "The generalized Fermat equation", pp. 119–149 in *Explicit methods in number theory*, Panor. Synthèses **36**, Soc. Math. France, Paris, 2012. MR Zbl

[Blasius 2004] D. Blasius, "Elliptic curves, Hilbert modular forms, and the Hodge conjecture", pp. 83–103 in *Contributions to automorphic forms, geometry, and number theory*, edited by H. Hida et al., Johns Hopkins University Press, Baltimore, 2004. MR Zbl

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system. I. The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR Zbl

[Breuil and Diamond 2014] C. Breuil and F. Diamond, "Formes modulaires de Hilbert modulo p et valeurs d'extensions entre caractères galoisiens", *Ann. Sci. Éc. Norm. Supér.* (4) **47**:5 (2014), 905–974. MR Zbl

[Breuil et al. 2001] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, "On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises", *J. Amer. Math. Soc.* **14**:4 (2001), 843–939. MR Zbl

[Bruin and Najman 2016] P. Bruin and F. Najman, "A criterion to rule out torsion groups for elliptic curves over number fields", *Res. Number Theory* **2** (2016), Art. 3, 13. MR Zbl

[Darmon 1997] H. Darmon, "Faltings plus epsilon, Wiles plus epsilon, and the generalized Fermat equation", *C. R. Math. Rep. Acad. Sci. Canada* **19**:1 (1997), 3–14. MR Zbl

[Darmon and Granville 1995] H. Darmon and A. Granville, "On the equations $z^m = F(x, y)$ and $Ax^p + By^q = Cz^r$", *Bull. London Math. Soc.* **27**:6 (1995), 513–543. MR Zbl

[Darmon and Merel 1997] H. Darmon and L. Merel, "Winding quotients and some variants of Fermat's last theorem", *J. Reine Angew. Math.* **490** (1997), 81–100. MR Zbl

[David 2012] A. David, "Caractère d'isogénie et critères d'irréductibilité", preprint, 2012. arXiv

[Dembélé and Voight 2013] L. Dembélé and J. Voight, "Explicit methods for Hilbert modular forms", pp. 135–198 in *Elliptic curves, Hilbert modular forms and Galois deformations* (CRM Barcelona), edited by H. Darmon et al., Birkhäuser/Springer, Basel, 2013. MR Zbl

[Derickx and van Hoeij 2014] M. Derickx and M. van Hoeij, "Gonality of the modular curve $X_1(N)$", *J. Algebra* **417** (2014), 52–71. MR Zbl

[Derickx et al. $\geq$ 2016] M. Derickx, S. Kamienny, W. Stein, and M. Stoll, "Torsion points on elliptic curves over number fields of small degree", in preparation.

[Dieulefait and Freitas 2013] L. Dieulefait and N. Freitas, "Fermat-type equations of signature $(13, 13, p)$ via Hilbert cuspforms", *Math. Ann.* **357**:3 (2013), 987–1004. MR Zbl

[Ellenberg 2004] J. S. Ellenberg, "Galois representations attached to $\mathbb{Q}$-curves and the generalized Fermat equation $A^4 + B^2 = C^p$", *Amer. J. Math.* **126**:4 (2004), 763–787. MR Zbl

[Freitas 2015] N. Freitas, "Recipes to Fermat-type equations of the form $x^r + y^r = Cz^p$", *Math. Z.* **279**:3 (2015), 605–639. MR Zbl

[Freitas and Siksek 2015a] N. Freitas and S. Siksek, "The asymptotic Fermat's last theorem for five-sixths of real quadratic fields", *Compos. Math.* **151**:8 (2015), 1395–1415. MR Zbl

[Freitas and Siksek 2015b] N. Freitas and S. Siksek, "Criteria for irreducibility of mod p representations of Frey curves", *J. Théor. Nombres Bordeaux* **27**:1 (2015), 67–76. MR Zbl

[Freitas and Siksek 2015c] N. Freitas and S. Siksek, "Fermat's last theorem over some small real quadratic fields", *Algebra Number Theory* **9**:4 (2015), 875–895. MR Zbl

[Freitas et al. 2015] N. Freitas, B. V. Le Hung, and S. Siksek, "Elliptic curves over real quadratic fields are modular", *Invent. Math.* **201**:1 (2015), 159–206. MR Zbl

[Kamienny 1992] S. Kamienny, "Torsion points on elliptic curves and q-coefficients of modular forms", *Invent. Math.* **109**:2 (1992), 221–229. MR Zbl

[Kisin 2009] M. Kisin, "Moduli of finite flat group schemes, and modularity", *Ann. of Math.* (2) **170**:3 (2009), 1085–1180. MR Zbl

[Kraus 1997] A. Kraus, "Majorations effectives pour l'équation de Fermat généralisée", *Canad. J. Math.* **49**:6 (1997), 1139–1161. MR Zbl

[Kraus 2007] A. Kraus, "Courbes elliptiques semi-stables sur les corps de nombres", *Int. J. Number Theory* **3**:4 (2007), 611–633. MR Zbl

[Langlands 1980] R. P. Langlands, *Base change for* GL(2), Annals of Mathematics Studies **96**, Princeton University Press, 1980. MR Zbl

[Merel 1996] L. Merel, "Bornes pour la torsion des courbes elliptiques sur les corps de nombres", *Invent. Math.* **124**:1-3 (1996), 437–449. MR Zbl

[Miller 2014] J. C. Miller, "Class numbers of real cyclotomic fields of composite conductor", *LMS J. Comput. Math.* **17**:suppl. A (2014), 404–417. MR Zbl

[Parent 2000] P. Parent, "Torsion des courbes elliptiques sur les corps cubiques", *Ann. Inst. Fourier (Grenoble)* **50**:3 (2000), 723–749. MR Zbl

[Parent 2003] P. Parent, "No 17-torsion on elliptic curves over cubic number fields", *J. Théor. Nombres Bordeaux* **15**:3 (2003), 831–838. MR Zbl

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl

[Siksek 2012] S. Siksek, "The modular approach to Diophantine equations", pp. 151–179 in *Explicit methods in number theory*, Panor. Synthèses **36**, Soc. Math. France, Paris, 2012. MR Zbl

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR Zbl

[Sinnott 1978] W. Sinnott, "On the Stickelberger ideal and the circular units of a cyclotomic field", *Ann. of Math.* (2) **108**:1 (1978), 107–134. MR Zbl

[Skinner and Wiles 1999] C. M. Skinner and A. J. Wiles, "Residually reducible representations and modular forms", *Inst. Hautes Études Sci. Publ. Math.* 89 (1999), 5–126. MR Zbl

[Swinnerton-Dyer 1973] H. P. F. Swinnerton-Dyer, "On l-adic representations and congruences for coefficients of modular forms", pp. 1–55. Lecture Notes in Math., Vol. 350 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Springer, Berlin, 1973. MR Zbl

[Thorne 2016] J. A. Thorne, "Automorphy of some residually dihedral Galois representations", *Math. Ann.* **364**:1 (2016), 589–648. MR Zbl

[Wiles 1995] A. Wiles, "Modular elliptic curves and Fermat's last theorem", *Ann. of Math.* (2) **141**:3 (1995), 443–551. MR Zbl

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1173

Geometry and stability of tautological bundles on Hilbert schemes of points

David Stapleton

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 $\mathcal{L}$ is a line bundle on S then $\mathcal{L}^{[n]}$ is the rank-n vector bundle whose fiber at the point corresponding to a length-n subscheme $\xi \subset S$ is the vector space $H^0(S, \mathcal{L} \otimes \mathcal{O}_\xi)$. 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 $\mathcal{O}_{\mathbb{P}^2}^{[n]}$, $\mathcal{O}_{\mathbb{P}^2}(1)^{[n]}$, and $\mathcal{O}_{\mathbb{P}^2}(2)^{[n]}$ generate the cohomology of $(\mathbb{P}^2)^{[n]}$. 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 $\mathrm{Sym}^n(S)$, 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 $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$. Our second result says the sheaf $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ 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*

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

and α_n induces an isomorphism between $(T_S)^{[n]}$ and $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$.

The analogous statement also holds for smooth curves. In general, the sheaves $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ are only guaranteed to be reflexive, as B_n is not a simple normal crossing divisor. However, Theorem B shows $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ 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* $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ *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 $\mathcal{E}$ 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*

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

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 $\mathfrak{S}_n$-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 $\mathcal{E}$ is a vector bundle on X, then for any divisor class $H \in N^1(X)$ we define the *slope of $\mathcal{E}$ with respect to H* to be the rational number

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

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

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

1. Stability of tautological bundles

In this section we prove that the tautological bundle of a stable vector bundle $\mathcal{E}$ is stable with respect to natural Chow divisors on $S^{[n]}$. Thus we deduce Theorem A when $\mathcal{E}$ 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:

$$\begin{array}{ccc} S \times S^{[n]} \supset \mathcal{Z}_n & \xrightarrow{p_1} & S \\ {\scriptstyle p_2}\downarrow & & \\ S^{[n]} & & \end{array}$$

For a fixed vector bundle $\mathcal{E}$ on S of rank r, we define

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

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

The symmetric group on n elements, $\mathfrak{S}_n$, naturally acts on the cartesian product S^n, and we write σ_n for the quotient map

$$\sigma_n : S^n \to S^n/\mathfrak{S}_n =: \mathrm{Sym}^n(S).$$

There is also a Hilbert–Chow morphism,

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

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

We wish to view $\mathcal{E}^{[n]}$ as an $\mathfrak{S}_n$-equivariant sheaf on S^n. Recall that if G is a finite group that acts on a scheme X, and if $\mathcal{F}$ is a coherent sheaf on X, then a *G-equivariant structure on $\mathcal{F}$* is given by a choice of isomorphisms

$$\phi_g : \mathcal{F} \to g^*\mathcal{F}$$

for all $g \in G$ satisfying the compatibility condition $h^*(\phi_g) \circ \phi_h = \phi_{gh}$. Following [Danila 2001] and [Scala 2009], we study the tautological bundles on $S^{[n]}$ by working with $\mathfrak{S}_n$-equivariant sheaves on S^n. For our purposes it is enough to study $\mathcal{E}^{[n]}$ equivariantly on the open subset of distinct points in $S^{[n]}$.

We write $\mathrm{Sym}^n(S)_\circ$ for the open subset of $\mathrm{Sym}^n(S)$ of distinct points. Likewise, given a map $f : X \to \mathrm{Sym}^n(S)$, we write $X_\circ$ for $f^{-1}(\mathrm{Sym}^n(S)_\circ)$. 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 $\mathcal{F}$ on X we define $\mathcal{F}_\circ := \mathcal{F}|_{X_\circ}$. The map $h_{n,\circ} : S^{[n]}_\circ \to \mathrm{Sym}^n(S)_\circ$ is an isomorphism. We define

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

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

$$(\mathcal{F})_{S^n} := j_*(\bar{\sigma}^*_{n,\circ}(\mathcal{F}_\circ)),$$

where j is the inclusion $j : S^n_\circ \to S^n$. The sheaf $(\mathcal{F})_{S^n}$ can be thought of as a modification of $\mathcal{F}$ along the exceptional divisor of h_n.

The pullback $\bar{\sigma}_{n,\circ}^*(-)$ is left exact, as the map $\bar{\sigma}_{n,\circ}$ is étale; thus the functor $(-)_{S^n}$ is left exact. If $\mathcal{F}$ is reflexive, the normality of S^n implies that the natural $\mathfrak{S}_n$-equivariant structure on the reflexive sheaf $\bar{\sigma}_{n,\circ}^*(\mathcal{F}_\circ)$ pushes forward uniquely to an $\mathfrak{S}_n$-equivariant structure on $(\mathcal{F})_{S^n}$.

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

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

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

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

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

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

Proof. Consider the following fiber square:

$$\begin{tikzcd} F := \mathcal{Z}_{n,\circ} \times_{S_\circ^{[n]}} S_\circ^n \arrow[r, "p'_{2,\circ}"] \arrow[d, "\bar{\sigma}'_{n,\circ}"'] & S_\circ^n \arrow[d, "\bar{\sigma}_{n,\circ}"] \\ \mathcal{Z}_{n,\circ} \arrow[r, "p_{2,\circ}"] & S_\circ^{[n]} \end{tikzcd}$$

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

$$(p'_{2,\circ})_*\big((\bar{\sigma}'_{n,\circ})^*(p_{1,\circ}^*(\mathcal{E}))\big) \cong \mathcal{E}_\circ^{\boxplus n}.$$

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

$$(p'_{2,\circ})_*\big((\bar{\sigma}'_{n,\circ})^*(p_{1,\circ}^*(\mathcal{E}))\big) \cong \bar{\sigma}_{n,\circ}^*\big((p_{2,\circ})_*(p_{1,\circ}^*(\mathcal{E}))\big).$$

The latter sheaf is $(\mathcal{E}^{[n]})_{S^n,\circ}$. Finally, any isomorphism between vector bundles on $S_\circ^n$ uniquely extends to an isomorphism between their pushforwards along j. Therefore, there is a natural $\mathfrak{S}_n$-equivariant isomorphism $(\mathcal{E}^{[n]})_{S^n} \cong \mathcal{E}^{\boxplus n}$. $\square$

Given an ample divisor H on S, there is a natural $\mathfrak{S}_n$-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 $\mathrm{Sym}^n(S)$. 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

$$H_n = \{\xi \in S^{[n]} \mid \xi \cap \mathrm{Supp}(H) \neq \varnothing\}.$$

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 $S^{[n]}_\circ$, $\mathrm{Sym}^n(S)_\circ$, and $S^n_\circ$. □

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 $\mathcal{E} \ncong \mathcal{O}_S$ is slope-stable on S with respect to an ample divisor H then there are no $\mathfrak{S}_n$-equivariant subsheaves of $\mathcal{E}^{\boxplus n}$ that are slope-destabilizing with respect to H_{S^n}.*

Proof. Let $0 \neq \mathcal{F} \subset \mathcal{E}^{\boxplus n}$ be an $\mathfrak{S}_n$-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

$$\mathcal{F}' \to \mathcal{E}^{\boxplus n} \to q_i^*\mathcal{E}$$

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 $\mu_{H_{S^n}}(q_i^*\mathcal{E}) = \mu_{H_{S^n}}(\mathcal{E}^{\boxplus n})$, and $\mathcal{F}$ does not destabilize $\mathcal{E}^{\boxplus n}$.

In case (2), we know $\mathcal{F}' \cong q_i^*\mathcal{E}$. Because $\mathcal{E} \not\cong \mathcal{O}_S$, the pullbacks $q_i^*\mathcal{E}$ and $q_j^*\mathcal{E}$ are not isomorphic unless $i = j$. As all the $q_j^*\mathcal{E}$ have the same slope and are stable with respect to H_{S^n}, we have $\mathrm{Hom}(\mathcal{F}', q_j^*\mathcal{E}) = 0$ for $j \neq i$. In particular, all the compositions

$$\mathcal{F}' \to \mathcal{E}^{\boxplus n} \to q_j^*\mathcal{E}$$

are zero for $j \neq i$. Thus $\mathcal{F}'$ is a summand of $\mathcal{E}^{\boxplus n}$. So $\mathcal{F}$ is an $\mathfrak{S}_n$-equivariant subsheaf of $\mathcal{E}^{\boxplus n}$, which contains one of the summands. But $\mathfrak{S}_n$ acts transitively on the summands so $\mathcal{F}$ contains all the summands, hence $\mathcal{F}$ does not destabilize $\mathcal{E}^{\boxplus n}$. □

Now we prove Theorem A in full generality.

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

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

$$\mu_{H_n}(\mathcal{F}) < \mu_{H_n}(\mathcal{E}^{[n]}) \quad \Longleftrightarrow \quad \mu_{H_{S^n}}((\mathcal{F})_{S^n}) < \mu_{H_{S^n}}(\mathcal{E}^{\boxplus n}).$$

Now $(\mathcal{F})_{S^n}$ is naturally an $\mathfrak{S}_n$-equivariant subsheaf of $\mathcal{E}^{\boxplus n}$. Thus, by Lemma 1.3,

$$\mu_{H_{S^n}}((\mathcal{F})_{S^n}) < \mu_{H_{S^n}}(\mathcal{E}^{\boxplus n}).$$

Therefore, $\mu_{H_n}(\mathcal{F}) < \mu_{H_n}(\mathcal{E}^{[n]})$ for all torsion-free subsheaves of intermediate rank, and $\mathcal{E}^{[n]}$ 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 $\mathrm{Der}_{\mathbb{C}}(-\log D)$, 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, $\mathrm{Der}_{\mathbb{C}}(-\log D)$ 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, $\mathrm{Der}_{\mathbb{C}}(-\log D)$ is reflexive by definition, so it is enough to define $\mathrm{Der}_{\mathbb{C}}(-\log D)$ 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, $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ 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 α_n induces an isomorphism between $(T_S)^{[n]}$ and $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$.

At a point $[\xi] \in S^{[n]}$ the map $\alpha_n|_{[\xi]}$ can be interpreted as deformations of ξ coming from tangent vectors of S. We expect that the degeneracy loci of α_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*

$$df : H^0(X, TX) \to H^0(X, f^*TY),$$

then $\delta \in H^0(Y, \mathrm{Der}_{\mathbb{C}}(-\log B))$.

Proof. It is enough to check that δ 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

$$f^*(y_1) = x_1^m, \quad f^*(y_i) = x_i \quad \text{for } i > 1.$$

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

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

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

$$f^*\delta = f^*(g_1) f^*\Big(\frac{\partial}{\partial y_1}\Big) + \cdots + f^*(g_n) f^*\Big(\frac{\partial}{\partial y_n}\Big).$$

Thus x_1^{m-1} divides $f^*(g_1)$. So y_1 divides g_1 and δ is in $H^0(Y, \mathrm{Der}_{\mathbb{C}}(-\log B))$. □

Proof of Theorem B. As in Section 1 we use $\mathcal{Z}_n \subset S \times S^{[n]}$ 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 $T_{S^{[n]}} = (p_2)_* \mathcal{H}om(\mathcal{I}_{\mathcal{Z}_n}, \mathcal{O}_{\mathcal{Z}_n})$. The normal sequence for $\mathcal{Z}_n$ gives a map

$$p_1^* T_S \oplus p_2^* T_{S^{[n]}} \cong T_{S \times S^{[n]}}|_{\mathcal{Z}_n} \xrightarrow{\beta} (\mathcal{I}_{\mathcal{Z}_n}/\mathcal{I}_{\mathcal{Z}_n}^2)^\vee \cong \mathcal{H}om(\mathcal{I}_{\mathcal{Z}_n}, \mathcal{O}_{\mathcal{Z}_n}).$$

Thus after pushing forward the first summand we get a map

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

To prove that α_n maps $(T_S)^{[n]}$ isomorphically onto $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ we first restrict to the open set $U \subset S^{[n]}$ parametrizing subschemes $\xi \subset S$, where ξ 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

$$V := p_2^{-1} U \subset \mathcal{Z}_n$$

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

$$\begin{tikzcd}
 & & p_2^*(T_S)^{[n]}|_V \arrow[d, "p_2^*\alpha_n|_V \oplus -\phi|_V"] & \\
0 \arrow[r] & T_{\mathcal{Z}_n}|_V \arrow[r] & p_2^* T_{S^{[n]}}|_V \oplus p_1^* T_S|_V \arrow[r, "\beta"] & \mathcal{H}om(\mathcal{I}_{\mathcal{Z}_n}, \mathcal{O}_{\mathcal{Z}_n})|_V
\end{tikzcd}$$

in which ϕ 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 $T_{\mathcal{Z}_n}|_V$. It follows from Lemma 2.2 that $(T_S)^{[n]}$ is contained in $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$. Now we can think of α_n as having codomain $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$. The map is an isomorphism of $(T_S)^{[n]}$ and $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ away from B_n and they both have the same first Chern class. Therefore, α_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. □

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 $\mathcal{L}_1$ nor $\mathcal{L}_2$ 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, $\mathrm{Der}_{\mathbb{C}}(-\log B_n)$ is a direct sum of stable bundles of the same slope with respect to $K_{S^{[n]}}$, proving Corollary C. $\square$

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

$$\dim T_{[\xi]} P_\xi = 2n - \mathrm{rank}(\alpha_n|_{[\xi]}).$$

Moreover, we give an explicit combinatorial formula for computing $\mathrm{rank}(\alpha_n|_{[\xi]})$ 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 $\mathcal{O}_{\mathbb{P}^2}(1)^{[n]}$ along an embedding of the curve in $(\mathbb{P}^2)^{[n]}$. To prove the theorem we need the spectral curves of [Beauville et al. 1989]. For completeness, we recall the construction.

Let $\pi : D \to C$ be an n:1 map between smooth irreducible projective curves and let $\mathcal{E}$ be an $\mathcal{O}_C$-module. If D can be embedded into the total space

$$\mathbb{L} := \mathcal{S}\mathrm{pec}_{\mathcal{O}_C}(\mathrm{Sym}^\bullet(\mathcal{L}^\vee)) \xrightarrow{\pi_{\mathbb{L}}} C$$

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

$$\pi_*\mathcal{O}_D \cong \mathrm{Sym}^\bullet(\mathcal{L}^\vee)/(x^n + s_1 x^{n-1} + \cdots + s_n)$$

for $x^n + s_1 x^{n-1} + \cdots + s_n \in H^0(\mathbb{L}, (\pi_{\mathbb{L}}^*\mathcal{L})^{\otimes n})$. Here we write $x \in H^0(\mathbb{L}, \pi_{\mathbb{L}}^*(\mathcal{L}))$ for the *coordinate section* of $\pi_{\mathbb{L}}^*(\mathcal{L})$. To give $\mathcal{E}$ the structure of a $\pi_*\mathcal{O}_D$-module we need to specify a multiplication map $m : \mathcal{E} \otimes \mathcal{L}^{-1} \to \mathcal{E}$ (equivalently $\mathcal{E} \to \mathcal{E} \otimes \mathcal{L}$) which satisfies the relation $m^n + s_1 m^{n-1} + \cdots + s_n = 0$.

Every $\mathcal{L}$-twisted endomorphism $m : \mathcal{E} \to \mathcal{E} \otimes \mathcal{L}$ has an associated $\mathcal{L}$-twisted characteristic polynomial which is a global section $p_m(x) \in H^0(\mathbb{L}, (\pi_{\mathbb{L}}^*\mathcal{L})^{\otimes n})$. A global version of the Cayley–Hamilton theorem says that m automatically satisfies its $\mathcal{L}$-twisted characteristic polynomial. In particular, if the zero set of $p_m(x)$ is D then $\mathcal{E}$ can naturally be thought of as a $\pi_*\mathcal{O}_D$-module. Fixing $s \in H^0(\mathbb{L}, (\pi_{\mathbb{L}}^*\mathcal{L})^{\otimes n})$, which cuts out the integral curve D, [Beauville et al. 1989, Proposition 3.6] gives the beautiful correspondence

$$\{\mathcal{E} \xrightarrow{m} \mathcal{E}\otimes\mathcal{L} \mid \mathcal{E} \text{ a vector bundle and } p_m(x) = s\}$$
$$\xleftrightarrow{1:1} \{\text{invertible sheaves } \mathcal{M} \text{ on } D\}. \quad (\diamond)$$

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

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

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

The method of proof of the Key Lemma involves a standard analysis of the *discriminant locus*, where a section of $\mathcal{E}\mathrm{nd}(\mathcal{E})\otimes\mathcal{L}$ has eigenvalues with multiplicity ≥ 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 $\mathcal{E}$ a rank-n semistable vector bundle on C. Let $\mathcal{L}$ be a line bundle on C of degree $\geq 2g$. By the Key Lemma, if

$$m : \mathcal{E} \to \mathcal{E}\otimes\mathcal{L}$$

is a general $\mathcal{L}$-twisted endomorphism then the resulting $\mathcal{L}$-twisted characteristic polynomial is smooth and irreducible.

Thus, by the correspondence $(\diamond)$ there is a line bundle $\mathcal{M}$ on D such that $\pi_*\mathcal{M} \cong \mathcal{E}$. The genus of D is $g_D = \binom{r}{2}\deg(\mathcal{L}) + n(g-1) + 1$ and is independent of $\mathcal{E}$. However, the degree of $\mathcal{M}$ is $\deg(\mathcal{E}) + \binom{r}{2}\deg(\mathcal{L})$ and does depend on the degree of $\mathcal{E}$. In particular, if

$$\deg(\mathcal{E}) \geq \binom{r}{2}\deg(\mathcal{L}) + r(2g-2) + 3$$

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 $\mathcal{E}$, proving Theorem D. $\square$

We now proceed with the proof of the Key Lemma.

Lemma 3.1. *If a subvariety $X \subset \mathbb{E}$ of a globally generated vector bundle $\mathbb{E}$ over a smooth curve C has codimension ≥ 2 then a generic section of $\mathbb{E}$ avoids X. If $X \subset \mathbb{E}$ is a reduced divisor then a generic section of $\mathbb{E}$ 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 $\mathbb{E} \to C$ 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. □

If $\mathbb{H}$ is the total space of $\mathcal{E}\mathrm{nd}(\mathcal{E}) \otimes \mathcal{L}$, and $\mathcal{C} = \mathbb{L} \oplus \cdots \oplus \mathbb{L}^{\otimes n}$, then there is a map $\epsilon : \mathbb{H} \to \mathcal{C}$ which sends an $\mathcal{L}$-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 $s : C \to \mathbb{H}$ meets $\mathbb{V}$ transversely and avoids the locus in $\mathbb{V}$ with more than one repeated eigenvalue or an eigenvalue of multiplicity ≥ 3, then the corresponding spectral curve is smooth.*

Proof. First, local trivialization of $\mathbb{H}$, $\mathbb{U}$, $\mathbb{V}$ and $\mathbb{L}$ implies it is enough to check on a fiber. Over a point $x \in C$ we have $\mathbb{H}|_x \cong \mathrm{Mat}_{n\times n}(k)$ and $\mathcal{C}|_x \cong \mathbb{A}^n$. Let $\mathbb{V}|_x$ be the locus of matrices whose eigenvalues have multiplicity ≥ 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. □

Proof of Key Lemma. Semistability of $\mathcal{E}$ and the inequality $\deg \mathcal{L} \geq 2g$ imply that $\mathcal{E}\mathrm{nd}(\mathcal{E}) \otimes \mathcal{L}$ is globally generated. By Lemma 3.1 and the first part of Lemma 3.2, a generic section s of $\mathcal{E}\mathrm{nd}(\mathcal{E}) \otimes \mathcal{L}$ meets $\mathbb{V}$ transversely and avoids the locus with more than one repeated eigenvalue or an eigenvalue of multiplicity of ≥ 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

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

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

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 $\mathcal{E}$ be a torsion-free sheaf on X. For any sheaf $\mathcal{Q}$ on X, we denote by $\mathrm{Sing}(\mathcal{Q})$ the closed locus where $\mathcal{Q}$ is not locally free.

Definition 4.1. The *slope of $\mathcal{E}$ with respect to γ* is the real number

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

Remark 4.2. Fixing an ample class $H \in N^1(X)_{\mathbb{R}}$, it is true that $\mu_H(\mathcal{E}) = \mu^{H^{d-1}}(\mathcal{E})$. 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 $\mathcal{E}$ is *slope-stable* (resp. *slope-semistable*) *with respect to γ* if, for all torsion-free quotients $\mathcal{E} \to \mathcal{Q} \to 0$ of intermediate rank, we have

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

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

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

Proposition 4.5. *Let $\mathcal{E} \to \mathcal{Q} \to 0$ be a torsion-free quotient which destabilizes $\mathcal{E}$ with respect to the curve class γ. Suppose $C \subset X$ is a smooth irreducible closed curve which represents γ, avoids* $\mathrm{Sing}(\mathcal{E})$, *and avoids the singularities of X. If C is not contained in* $\mathrm{Sing}(Q)$ *then $\mathcal{E}|_C$ 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 $\mathrm{Sing}(\mathcal{Q})$ and $\mathrm{Sing}(\mathcal{E})$ 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 $\mathrm{Sing}(\mathcal{Q})$ and $\mathrm{Sing}(\mathcal{E})$ properly. Once such a surface is chosen, we have

$$c_1(\mathcal{Q})|_S = c_1(\mathcal{Q}|_S) = c_1(\mathcal{Q}|_S/\mathrm{Tors}(\mathcal{Q}|_S)), \quad c_1(\mathcal{E})|_S = c_1(\mathcal{E}|_S) = c_1(\mathcal{E}|_S/\mathrm{Tors}(\mathcal{E}|_S))$$

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

$$\mathcal{E}|_S/\mathrm{Tors}(\mathcal{E}|_S) \to \mathcal{Q}|_S/\mathrm{Tors}(\mathcal{Q}|_S) \to 0$$

is a torsion-free quotient on S which destabilizes $\mathcal{E}|_S/\mathrm{Tors}(\mathcal{E}|_S)$ 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 $c_1(\mathcal{Q}|_C) = c_1(\mathcal{Q})|_C$. The restriction $c_1(\mathcal{Q})|_C$ is computed via the derived pullback

$$c_1(\mathcal{Q})|_C = \sum_{i=0}^{\infty} (-1)^i c_1(\mathrm{Tor}_i^{\mathcal{O}_X}(\mathcal{Q}, \mathcal{O}_C)),$$

where the $\mathrm{Tor}_i^{\mathcal{O}_X}(\mathcal{Q}, \mathcal{O}_C)$ 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 $\mathcal{O}_C$ has a two-term locally free resolution. So the $\mathrm{Tor}_i^{\mathcal{O}_X}(\mathcal{Q}, \mathcal{O}_C)$ vanish for $i > 2$ and $\mathrm{Tor}_1^{\mathcal{O}_X}(\mathcal{Q}, \mathcal{O}_C) = 0$ because $\mathcal{Q}$ is torsion-free. Therefore,

$$c_1(\mathcal{Q})|_C = c_1(\mathrm{Tor}_0^{\mathcal{O}_X}(\mathcal{Q}, \mathcal{O}_C)) = c_1(\mathcal{Q}|_C).$$

So $\mathcal{E}|_C$ is not slope-stable. □

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

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

Proof. Suppose for contradiction that $\mathcal{E}$ is unstable with respect to γ. Then there exists a torsion-free quotient $\mathcal{E} \to \mathcal{Q} \to 0$ with $\mu^\gamma(\mathcal{Q}) \leq \mu^\gamma(\mathcal{E})$. As $\mathcal{Q}$ is torsion-free, $\mathrm{Sing}(\mathcal{Q})$ has codimension ≥ 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 $\mathrm{Sing}(\mathcal{Q})$. Then Proposition 4.5 guarantees that $\mathcal{E}|_{C_t}$ 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

$$F := \{(H_1, \dots, H_{d-1}) \in |kH_X|^{d-1} \mid C = H_1 \cap \cdots \cap H_{d-1} \text{ is a smooth complete intersection curve and } \mathcal{E}|_C \text{ is stable}\}.$$

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 γ.

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

$$G := \{(H_1, \dots, H_{e-1}) \in |lH_Y|^{e-1} \mid H_1 \cap \cdots \cap H_{e-1} \text{ is a smooth complete intersection curve}\}.$$

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 δ.

Finally,

$$(p_1^* H_X + p_2^* H_Y)^{d+e-1} = \binom{d+e-1}{e} \frac{(H_Y)^e}{k^{d-1}} \cdot \gamma + \binom{d+e-1}{d} \frac{(H_X)^d}{l^{e-1}} \cdot \delta.$$

Thus, by Lemma 4.4, $p_1^*(\mathcal{E})$ is slope-stable with respect to $p_1^*(H_X) + p_2^*(H_Y)$. $\square$

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 lef 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 $\mathcal{E}$ 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 $\mathcal{E}$ is stable with respect to $H + \epsilon A$ for ϵ 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 $\mathcal{E}$ is a vector bundle on a smooth projective surface S which is stable with respect to an ample divisor H, then $\mathcal{E}^{[n]}$ 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 $\mathcal{E}^{[n]}$ and H_n satisfy the conditions of Proposition 4.8. Therefore, $\mathcal{E}^{[n]}$ is stable with respect to ample divisors close to H_n. □

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 $\gamma \in U$, $\mathcal{E}$ is stable with respect to γ. More precisely, if $\delta \in N_1(X)_{\mathbb{R}}$ represents the $(d-1)$-st power of an ample divisor then $\mathcal{E}$ is stable with respect to the perturbed curve class $[H^{d-1}] + \epsilon \cdot \delta$ for ϵ 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 ϵ, $(H + \epsilon A)^{d-1}$ maps into U. Therefore, for ϵ sufficiently small, $\mathcal{E}$ is stable with respect to $H + \epsilon A$. □

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

[Arnold 1971] V. I. Arnold, "Matrices depending on parameters", *Uspehi Mat. Nauk* **26**:2(158) (1971), 101–114. In Russian; translated in *Russ. Math. Surv.* **26**:2 (1971), 29–43. MR Zbl

[Aubin 1976] T. Aubin, "Équations du type Monge–Ampère sur les variétés kähleriennes compactes", *C. R. Acad. Sci. Paris Sér. A-B* **283**:3 (1976), Aiii, A119–A121. MR Zbl

[Beauville et al. 1989] A. Beauville, M. S. Narasimhan, and S. Ramanan, "Spectral curves and the generalised theta divisor", *J. Reine Angew. Math.* **398** (1989), 169–179. MR Zbl

[Bejleri and Stapleton 2016] D. Bejleri and D. Stapleton, "The tangent space of the punctual Hilbert scheme", preprint, 2016. arXiv

[Buchweitz et al. 2009] R.-O. Buchweitz, W. Ebeling, and H.-C. Graf von Bothmer, "Low-dimensional singularities with free divisors as discriminants", *J. Algebraic Geom.* **18**:2 (2009), 371–406. MR Zbl

[de Cataldo and Migliorini 2002] M. A. A. de Cataldo and L. Migliorini, "The hard Lefschetz theorem and the topology of semismall maps", *Ann. Sci. École Norm. Sup.* (4) **35**:5 (2002), 759–772. MR Zbl

[Danila 2001] G. Danila, "Sur la cohomologie d'un fibré tautologique sur le schéma de Hilbert d'une surface", *J. Algebraic Geom.* **10**:2 (2001), 247–280. MR Zbl

[Donaldson 1985] S. K. Donaldson, "Anti self-dual Yang–Mills connections over complex algebraic surfaces and stable vector bundles", *Proc. London Math. Soc.* (3) **50**:1 (1985), 1–26. MR Zbl

[Ellingsrud and Strømme 1993] G. Ellingsrud and S. A. Strømme, "Towards the Chow ring of the Hilbert scheme of $\mathbb{P}^2$", *J. Reine Angew. Math.* **441** (1993), 33–44. MR Zbl

[Fogarty 1973] J. Fogarty, "Algebraic families on an algebraic surface, II: The Picard scheme of the punctual Hilbert scheme", *Amer. J. Math.* **95** (1973), 660–687. MR Zbl

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics **2**, Springer, Berlin, 1998. MR Zbl

[Greb and Toma 2013] D. Greb and M. Toma, "Compact moduli spaces for slope-semistable sheaves", preprint, 2013. arXiv

[Greb et al. 2016] D. Greb, S. Kebekus, and T. Peternell, "Movable curves and semistable sheaves", *Int. Math. Res. Not.* **2016**:2 (2016), 536–570. MR Zbl

[Kobayashi 1987] S. Kobayashi, *Differential geometry of complex vector bundles*, Publications of the Mathematical Society of Japan **15**, Princeton University Press, Iwanami Shoten, Tokyo, 1987. MR Zbl

[Lehn 1998] M. Lehn, "On the cotangent sheaf of Quot-schemes", *Internat. J. Math.* **9**:4 (1998), 513–522. MR Zbl

[Lehn 1999] M. Lehn, "Chern classes of tautological sheaves on Hilbert schemes of points on surfaces", *Invent. Math.* **136**:1 (1999), 157–207. MR Zbl

[Mehta and Ramanathan 1984] V. B. Mehta and A. Ramanathan, "Restriction of stable sheaves and representations of the fundamental group", *Invent. Math.* **77**:1 (1984), 163–172. MR Zbl

[Mistretta 2006] E. C. Mistretta, *Some constructions around stability of vector bundles on projective varieties*, Ph.D. thesis, Université Paris Diderot, Paris 7, 2006, available at http://www.math.unipd.it/~ernesto/pub/tesi.pdf.

[Nakajima 1999] H. Nakajima, *Lectures on Hilbert schemes of points on surfaces*, University Lecture Series **18**, American Mathematical Society, Providence, RI, 1999. MR Zbl

[Okounkov 2014] A. Y. Okounkov, "Hilbert schemes and multiple q-zeta values", *Funktsional. Anal. i Prilozhen.* **48**:2 (2014), 79–87. MR Zbl

[Scala 2009] L. Scala, "Cohomology of the Hilbert scheme of points on a surface with values in representations of tautological bundles", *Duke Math. J.* **150**:2 (2009), 211–267. MR Zbl

[Schlickewei 2010] U. Schlickewei, "Stability of tautological vector bundles on Hilbert squares of surfaces", *Rend. Semin. Mat. Univ. Padova* **124** (2010), 127–138. MR Zbl

[Wandel 2013] M. Wandel, "Tautological sheaves: Stability, moduli spaces and restrictions to generalised Kummer varieties", preprint, 2013. arXiv

[Wandel 2014] M. Wandel, "Stability of tautological bundles on the Hilbert scheme of two points on a surface", *Nagoya Math. J.* **214** (2014), 79–94. MR Zbl

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1191

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 conjectures. 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 $S_0(K, \mathcal{A}_g)$ 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 $s \in S_0(K, \mathcal{A}_g)$ gives rise to a system of ℓ-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 $\mathcal{A}_g(K)$, 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 $\overline{X}$ 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

$$1 \to \pi_1(\bar{X}) \to \pi_1(X) \to G_K \to 1. \tag{1}$$

The map $p_X : \pi_1(X) \to G_K$ from the above sequence is obtained by functoriality from the structural morphism $X \to \operatorname{Spec} K$. 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 $\pi_1(\bar{X})$ by which to define $H^1(G_K, \pi_1(\bar{X}))$, 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 $\bar{K} \subset \bar{K}_v$ induces a map $\alpha_v : G_{K_v} \to G_K$ 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, σ_{X_v/K_v} factors through $X(K_v)_\bullet$, 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 $X(K_v) = X(K_v)_\bullet$ and we have the following diagram:

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

We define the set X^f (the finite descent obstruction) as the set of points $(P_v)_v \in \prod_v X(K_v)_\bullet$ 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 $\prod X(K_v)_\bullet$ 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 $X^f = \varnothing$ if and only if $S(K, X) = \varnothing$.*

Proof. If $X^f \neq \varnothing$ 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 $S(K, X) \neq \varnothing$.

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 $(P_v) \in X^f$. □

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 $\mathrm{End}(A)$ for its ring of all K-endomorphisms and $\mathrm{End}^0(A)$ for the corresponding (finite-dimensional semisimple) $\mathbb{Q}$-algebra $\mathrm{End}(A)\otimes\mathbb{Q}$. If $n\geq 3$ is an integer that is not divisible by $\mathrm{char}(K)$ and all points of order n on A are defined over K then, by a theorem of Silverberg [1992], all $\bar{K}$-endomorphisms of A are defined over K, i.e., lie in $\mathrm{End}(A)$.

If ℓ is a prime different from $\mathrm{char}(K)$ 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

$$\rho_{\ell,A}:G_K\to \mathrm{Aut}_{\mathbb{Z}_\ell}(T_\ell(A))$$

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

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

The image of $\mathrm{End}(A)\otimes\mathbb{Z}_\ell$ 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 ℓ-adic representation

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

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

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

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

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

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

Let χ_ℓ be the *cyclotomic character* $\chi_\ell:G_K\to\mathbb{Z}_\ell^*$ that defines the Galois action on all ℓ-power roots of unity, and $\mathbb{Z}_\ell(1)$ the ℓ-adic Tate module of the *multiplicative* *group* $\mathbb{G}_m$. 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

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

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 $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(A), \mathbb{Z}_\ell(1))$ are isomorphic.

3. Moduli of abelian varieties

The moduli space of principally polarized abelian varieties of dimension g is denoted by $\mathcal{A}_g$. 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 $\mathcal{A}_g$, 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 $\mathcal{A}_g$ is defined over $\mathbb{Q}$, we can consider it over an arbitrary number field K. As per our earlier conventions, $\bar{\mathcal{A}}_g$ is the base change of $\mathcal{A}_g$ 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 $\bar{\mathcal{A}}_g$ is the symplectic group $\mathrm{Sp}_{2g}(\mathbb{Z})$ and the algebraic fundamental group is its profinite completion. When $g > 1$ (which we henceforth assume) $\mathrm{Sp}_{2g}(\mathbb{Z})$ has the congruence subgroup property [Bass et al. 1964; Mennicke 1965] and therefore its profinite completion is $\mathrm{Sp}_{2g}(\hat{\mathbb{Z}})$.

The group $\pi_1(\mathcal{A}_g)$ is essentially described by the exact sequences (3.2) and (3.3) of [Ihara and Nakamura 1997] and it follows that the set $H(K, \mathcal{A}_g)$ 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 $\chi : G_K \to \hat{\mathbb{Z}}^*$.

$$\begin{tikzcd}
1 \arrow[r] & \pi_1(\bar{\mathcal{A}}_g) \arrow[r] \arrow[d, "\cong"] & \pi_1(\mathcal{A}_g) \arrow[r] \arrow[d] & G_K \arrow[r] \arrow[d, "\chi"'] & 1 \\
1 \arrow[r] & \mathrm{Sp}_{2g}(\hat{\mathbb{Z}}) \arrow[r] & \mathrm{GSp}_{2g}(\hat{\mathbb{Z}}) \arrow[r] & \hat{\mathbb{Z}}^* \arrow[r] & 1
\end{tikzcd}$$

The coverings of $\bar{\mathcal{A}}_g$ corresponding to the congruence subgroups of $\mathrm{Sp}_{2g}(\hat{\mathbb{Z}})$ are those obtained by adding level structures. In particular, for an abelian variety A, $\sigma_{\mathcal{A}_g/K}(A) = \prod T_\ell(A)$, 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 $\sigma_{\mathcal{A}_g/K}$, 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 $\sigma_{\mathcal{A}_g/K}$ is finite. On the other hand, $\sigma_{\mathcal{A}_g/K}$ is not necessarily injective to $S_0(K, \mathcal{A}_g)$. For example, for each g there exists K with noninjective $\sigma_{\mathcal{A}_g/K}$. Regarding surjectivity, we will prove that those elements of $S_0(K, \mathcal{A}_g)$ 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 $\sigma_{\mathcal{A}_g/K}$, assuming the Fontaine–Mazur conjecture, the Grothendieck–Serre conjecture on semisimplicity of ℓ-adic cohomology of smooth projective varieties, and the Tate and Hodge conjectures. The integral Selmer set $S_0(K, \mathcal{A}_g)$, 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 $\sigma_{\mathcal{A}_g/K}$ to $S_0(K, \mathcal{A}_g)$ 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 $G_K = \mathrm{Gal}(\overline{K}/K)$. Fix a finite set of rational primes S, and consider a collection of continuous ℓ-adic representations

$$\{\rho_\ell : G_K \to \mathrm{GL}_N(\mathbb{Q}_\ell)\}_{\ell \notin S}.$$

We will say that the collection $\{\rho_\ell\}_{\ell \notin S}$ is *weakly compatible* if there exists a finite set of places Σ of K such that

(1) for all $\ell \notin S$, ρ_ℓ is unramified outside the union of Σ and the places Σ_ℓ of K dividing ℓ; and

(2) for all $v \notin \Sigma \cup \Sigma_\ell$, denoting by fr_v a (geometric) frobenius element at v, the characteristic polynomial of $\rho_\ell(\mathrm{fr}_v)$ has rational coefficients and is independent of $\ell \notin S$.[1]

Our aim is to prove the following:

Theorem 3.1. *We will assume $\{\rho_\ell\}_{\ell \notin S}$ is weakly compatible and moreover satisfies the following three conditions*:

(1) *For some prime $\ell_0 \notin S$, ρ_{ℓ_0} is de Rham at all places of K above ℓ_0.*

(2) *For some prime $\ell_1 \notin S$, ρ_{ℓ_1} is absolutely irreducible.*

[1] These systems were introduced by Serre [1989], who called them *strictly compatible*.

(3) *For some prime $\ell_2 \notin S$, 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 ℓ.*

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 ℓ-adic realization functor*

$$\mathcal{M}_{k,\mathbb{Q}_\ell} \to \mathrm{Rep}_{\mathbb{Q}_\ell}(G_k),$$

valued in the category of continuous ℓ-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). $\square$

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 $\mathrm{Rep}(\mathcal{G}_{k,E})$ for some proreductive group $\mathcal{G}_{k,E}$ 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 $\iota : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_\ell$, so that when E is a subfield of $\overline{\mathbb{Q}}$ we can speak of the ℓ-adic realization

$$H_\iota : \mathcal{M}_{k,E} \to \mathrm{Rep}_{\overline{\mathbb{Q}}_\ell}(G_k)$$

associated to ι.

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 ℓ-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 $r_\ell : G_K \to \mathrm{GL}_N(\mathbb{Q}_\ell)$ 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_ℓ is isomorphic to a subquotient of $H^i(X_{\overline{K}}, \mathbb{Q}_\ell)(j)$ 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 [SGA $4\frac{1}{2}$ 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, $H^i(X_{\overline{K}}, \mathbb{Q}_\ell)$ is a semisimple representation of G_K. Putting all of these conjectures together, we can characterize the essential image of H_ι:

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

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

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

Proof. The Fontaine–Mazur conjecture asserts that for some smooth projective variety X/k, r_ℓ is a subquotient of $H^i(X_{\overline{K}}, \mathbb{Q}_\ell)(j)$ 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

$$H_\iota : \mathrm{End}_{\mathcal{M}_K}\big(H^i(X)(j)\big) \otimes_{\mathbb{Q}} \overline{\mathbb{Q}}_\ell \xrightarrow{\sim} \mathrm{End}_{\overline{\mathbb{Q}}_\ell[G_K]}\big(H^i(X_{\overline{K}}, \overline{\mathbb{Q}}_\ell)(j)\big) \tag{2}$$

is an isomorphism.

Now, there is a projector (of $\overline{\mathbb{Q}}_\ell[G_K]$-modules) $H^i(X_{\overline{K}}, \overline{\mathbb{Q}}_\ell)(j) \twoheadrightarrow r_\ell$, which combined with Equation (2) yields a projector in $\mathrm{End}_{\mathcal{M}_K}(H^i(X)(j)) \otimes_{\mathbb{Q}} \overline{\mathbb{Q}}_\ell$ whose image has ℓ-adic realization r_ℓ. But $\mathrm{End}_{\mathcal{M}_K}(H^i(X)(j))$ is a semisimple algebra over $\mathbb{Q}$ (Lemma 3.2), which certainly splits over $\overline{\mathbb{Q}}$, so the decomposition of $H^i(X)(j)$ into simple objects of $\mathcal{M}_{K,\overline{\mathbb{Q}}_\ell}$ is already realized in $\mathcal{M}_{K,\overline{\mathbb{Q}}}$.[2]

For the final claim about the essential image (which we do not use in what follows), it suffices to show an irreducible $r_\iota : G_K \to \mathrm{GL}_N(\overline{\mathbb{Q}}_\ell)$ lies in the essential image. Such an r_ι is defined over a finite extension of $\mathbb{Q}_\ell$ and can thus be regarded as a higher-dimensional geometric representation r_ℓ with $\mathbb{Q}_\ell$-coefficients, necessarily semisimple. By the first part of the lemma, $r_\ell \otimes_{\mathbb{Q}_\ell} \overline{\mathbb{Q}}_\ell$ is isomorphic to $H_\iota(M)$ for some $M \in \mathcal{M}_{K,\overline{\mathbb{Q}}}$, and by the Tate conjecture there is a projector in $\mathrm{End}(M) \otimes_{\overline{\mathbb{Q}}} \overline{\mathbb{Q}}_\ell$ inducing the canonical (adjunction) projector $r_\ell \otimes_{\mathbb{Q}_\ell} \overline{\mathbb{Q}}_\ell \twoheadrightarrow r_\iota$. 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_ι is in the essential image of H_ι. $\square$

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 $\iota_0 : \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_{\ell_0}$, 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 $\rho : \mathcal{G}_{K,E} \to \mathrm{GL}_{N,E}$ such that $H_{\iota_0}(\rho) \cong \rho_{\ell_0} \otimes \overline{\mathbb{Q}}_{\ell_0}$. Let us denote by λ_0 the place of E induced by $E \subset \overline{\mathbb{Q}} \xrightarrow{\iota_0} \overline{\mathbb{Q}}_\ell$. Then for all finite places λ of E (say $\lambda|\ell$), and for almost all places v of K, compatibility gives us the following equality of rational numbers (note that ρ_λ denotes the λ-adic realization of the motivic Galois representation ρ, while ρ_ℓ denotes the original ℓ-adic representation in our compatible system):

$$\mathrm{tr}(\rho_\lambda(\mathrm{fr}_v)) = \mathrm{tr}(\rho_{\lambda_0}(\mathrm{fr}_v)) = \mathrm{tr}(\rho_{\ell_0}(\mathrm{fr}_v)) = \mathrm{tr}(\rho_\ell(\mathrm{fr}_v)).$$

[2] In fact, it is realized over the maximal CM subfield of $\overline{\mathbb{Q}}$: see, e.g., [Patrikis 2012, Lemma 4.1.22].

Here we use the fact that the collection of ℓ-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 $\rho_\ell^{\mathrm{ss}} \otimes_{\mathbb{Q}_\ell} E_\lambda \cong \rho_\lambda$; this holds for all λ for which ρ_ℓ makes sense, i.e., for all λ above $\ell \notin S$.

Recall that for some $\ell_1 \notin S$, we have assumed ρ_{ℓ_1} is absolutely irreducible; hence for any place λ_1 of E above ℓ_1, the previous paragraph shows that $\rho_{\lambda_1} \cong \rho_{\ell_1} \otimes E_{\lambda_1}$ is absolutely irreducible. *A fortiori*, ρ is absolutely irreducible, and then by the Tate conjecture all ρ_λ 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 $\ell \notin S$.

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

Lemma 3.4. *With the notation above*, obs_ρ *in fact belongs to*

$$\ker\Bigl(H^1\bigl(\mathrm{Gal}(E/\mathbb{Q}), \mathrm{PGL}_N(E)\bigr) \to \prod_{\ell \notin S} H^1\bigl(\mathrm{Gal}(E_\lambda/\mathbb{Q}_\ell), \mathrm{PGL}_N(E_\lambda)\bigr)\Bigr).$$

In particular, *if S is empty*, *then ρ can be defined over* $\mathbb{Q}$.

Proof. We know that each of the λ-adic realizations ρ_λ (for $\lambda | \ell \notin S$) can be defined over $\mathbb{Q}_\ell$; to prove the lemma, we need to verify that the canonical localizations of obs_ρ (which arise by extending scalars on the motivic Galois representation) are in fact given by the corresponding obstruction classes for the λ-adic realizations. Thus, we have to recall how these realizations are constructed from ρ itself. The surjection $\mathcal{G}_K \twoheadrightarrow G_K$ admits a continuous section on $\mathbb{Q}_\ell$-points, $s_\ell : G_K \to \mathcal{G}_K(\mathbb{Q}_\ell)$; composition with $\rho \otimes_E E_\lambda$ yields ρ_λ:

$$\begin{tikzcd} G_K \arrow[r, "s_\ell"'] \arrow[rrr, bend left, "\rho_\lambda"] & \mathcal{G}_K(\mathbb{Q}_\ell) \arrow[r, hook] & \mathcal{G}_{K,E}(E_\lambda) \arrow[r, "\rho \otimes_E E_\lambda"'] & \mathrm{GL}_N(E_\lambda). \end{tikzcd}$$

By construction of the respective obstruction classes, the canonical map from endomorphisms of $\rho \otimes_E E_\lambda$ to those of ρ_λ realizes the obstruction class for ρ_λ as the localization of obs_ρ at $\mathrm{Gal}(E_\lambda/\mathbb{Q}_\ell)$. But we have seen that ρ_λ can be defined over $\mathbb{Q}_\ell$, so we conclude that obs_ρ has trivial restriction to each $\mathrm{Gal}(E_\lambda/\mathbb{Q}_\ell)$, as desired.

For the final claim, note that by Hilbert 90 we can regard obs_ρ as an element of

$$\ker\left(H^2(\mathrm{Gal}(E/\mathbb{Q}), E^\times) \to \prod_{\ell \notin S} H^2(\mathrm{Gal}(E_\lambda/\mathbb{Q}_\ell), E_\lambda^\times)\right).$$

If S is empty, then the structure of the Brauer group of $\mathbb{Q}$ (which has only one infinite place!) then forces obs_ρ to be trivial. □

Proof of Theorem 3.1. From now on we assume $S = \varnothing$, so that our compatible system $\{\rho_\ell\}_\ell$ arises from a rational representation

$$\rho : \mathcal{G}_K \to \mathrm{GL}_{N,\mathbb{Q}}.$$

Let M be the rank N object of $\mathcal{M}_K$ corresponding to ρ via the Tannakian equivalence. Recall that we are given a prime ℓ_2 and a place $v|\ell_2$ of K for which we are given that $\rho_{\ell_2}|_{G_{K_v}}$ 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 "ℓ_2-adic Hodge theory": denoting Fontaine's period ring over K_v by $\mathrm{B}_{\mathrm{dR},K_v}$, and the de Rham realization functor by $H_{\mathrm{dR}} : \mathcal{M}_K \to \mathrm{Fil}_K$ (the category of filtered K-vector spaces), we have the comparison (respecting filtration and G_{K_v}-action)

$$H_{\mathrm{dR}}(M) \otimes_K \mathrm{B}_{\mathrm{dR},K_v} \xrightarrow{\sim} H_{\ell_2}(M) \otimes_{\mathbb{Q}_{\ell_2}} \mathrm{B}_{\mathrm{dR},K_v},$$

hence

$$H_{\mathrm{dR}}(M) \otimes_K K_v \cong \mathrm{D}_{\mathrm{dR},K_v}(H_{\ell_2}(M)).$$

The Hodge filtration on $H_{\mathrm{dR}}(M)$ therefore satisfies

$$\dim_K \mathrm{gr}^0(H_{\mathrm{dR}}(M)) = \dim_K \mathrm{gr}^{-1}(H_{\mathrm{dR}}(M)) = \frac{N}{2} \tag{3}$$

and $\mathrm{gr}^i(H_{\mathrm{dR}}(M)) = 0$ 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

$$H_{\mathrm{dR}}(M) \otimes_K \mathbb{C} \xrightarrow{\sim} H_{\mathrm{B}}(M|_{\mathbb{C}}) \otimes_{\mathbb{Q}} \mathbb{C}. \tag{4}$$

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 $H_{\mathrm{B}}(M|_{\mathbb{C}})$ 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

$H_1(A(\mathbb{C}), \mathbb{Q}) \cong H_{\mathrm{B}}(M|_{\mathbb{C}})$. The Hodge conjecture implies that this isomorphism comes from an isomorphism $H_1(A) \xrightarrow{\sim} M|_{\mathbb{C}}$ in $\mathcal{M}_{\mathbb{C}}$.

For any $\sigma \in \mathrm{Aut}(\mathbb{C}/\overline{\mathbb{Q}})$, we deduce an isomorphism

$$^{\sigma}H_1(A) \xrightarrow{\sim} {}^{\sigma}M|_{\mathbb{C}} = M|_{\mathbb{C}} \xleftarrow{\sim} H_1(A),$$

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 $\mathcal{K}$ be a countable subfield of the field $\mathbb{C}$ and $\bar{\mathcal{K}}$ the algebraic closure of $\mathcal{K}$ in $\mathbb{C}$. Let $\mathcal{Y}$ be a complex abelian variety of dimension g such that for each field automorphism $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathcal{K})$ the complex abelian variety $\mathcal{Y}$ and its "conjugate" $^{\sigma}\mathcal{Y} = \mathcal{Y} \times_{\mathbb{C},\sigma} \mathbb{C}$ are isogenous. Then there exists an abelian variety $\mathcal{Y}_0$ over $\bar{\mathcal{K}}$ such that $\mathcal{Y}_0 \times_{\bar{\mathcal{K}}} \mathbb{C}$ is isomorphic to $\mathcal{Y}$.*

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

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

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 $\mathrm{Res}_{L/K}(A_L)$; under the *fully faithful* embedding

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

we can think of $H_1(\mathrm{Res}_{L/K}(A_L))$ as $\mathrm{Ind}_L^K(H_1(A_L))$, where the induction is taken in the sense of motivic Galois representations (note that the quotient $\mathcal{G}_K/\mathcal{G}_L$ is canonically $\mathrm{Gal}(L/K)$, so this is just the usual induction from a finite-index subgroup). Frobenius reciprocity then implies the existence of a nonzero map $M \to \mathrm{Ind}_L^K(H_1(A_L))$ 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 $\mathrm{Res}_{L/K}(A_L)$ 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 $\bar{\mathcal{K}}$ is also countable, we may replace $\mathcal{K}$ by $\bar{\mathcal{K}}$, i.e., assume that $\mathcal{K}$ is algebraically closed. Since the isogeny class of $\mathcal{Y}$ consists of a countable set of (complex) abelian varieties (up to an isomorphism), we conclude that the set $\mathrm{Aut}(\mathbb{C}/\mathcal{K})(\mathcal{Y})$ of isomorphism classes of complex abelian varieties of the form $\{^{\sigma}\mathcal{Y} \mid \sigma \in \mathrm{Aut}(\mathbb{C}/\mathcal{K})\}$ is either finite or countable.

Our plan is as follows. Let us consider a *fine* moduli space $\mathcal{A}_{g,?}$ 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 $P \in \mathcal{A}_{g,?}(\mathbb{C})$ be the corresponding point of our moduli space. We need to prove that

$$P \in \mathcal{A}_{g,?}(\mathcal{K}).$$

Suppose that it is not true. Then the orbit $\mathrm{Aut}(\mathbb{C}/\mathcal{K})(P)$ of P is *uncountable*. Indeed, P lies in one of the $(N+1)$ affine charts/spaces $\mathbb{A}^N$ that do cover $\mathbb{P}^N$. This implies that P does *not* belong to $\mathbb{A}^N(\mathcal{K})$ and therefore (at least) one of its coordinates is transcendental over $\mathcal{K}$. But the $\mathrm{Aut}(\mathbb{C}/\mathcal{K})$-orbit of this coordinate coincides with uncountable $\mathbb{C} \setminus \mathcal{K}$ and therefore the $\mathrm{Aut}(\mathbb{C}/\mathcal{K})$-orbit $\mathrm{Aut}(\mathbb{C}/\mathcal{K})(P)$ of P is uncountable in $\mathcal{A}_{g,?}(\mathbb{C})$. However, for each $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathcal{K})$ 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 $\mathrm{Aut}(\mathbb{C}/\mathcal{K})(\mathcal{Y})$ of $\mathcal{Y}$, is, at most, countable, we conclude that the orbit $\mathrm{Aut}(\mathbb{C}/\mathcal{K})(P)$ 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 δ that was introduced and studied by D. Mumford [1966]. In order to choose (define) a suitable δ, let us pick a totally symmetric ample invertible sheaf $\mathcal{L}_0$ on $\mathcal{Y}$ [Mumford 1966, Section 2] and consider its 8th power $\mathcal{L} := \mathcal{L}_0^8$ in $\mathrm{Pic}(\mathcal{Y})$. Then $\mathcal{L}$ is a very ample invertible sheaf that defines a polarization $\Lambda(\mathcal{L})$ on $\mathcal{Y}$ [Mumford 1966, Part I, Section 1] that is an isogeny from $\mathcal{Y}$ to its dual; the kernel $H(\mathcal{L})$ of $\Lambda(\mathcal{L})$ is a finite commutative subgroup of $\mathcal{Y}(\mathbb{C})$ (that contains all points of order 8). The order of $H(\mathcal{L})$ is the degree of the polarization. The type δ is essentially the isomorphism class of the group $H(\mathcal{L})$ [Mumford 1966, Part I, Section 1, p. 294]. The resulting moduli space $\mathcal{A}_{g,?} := M_\delta$ [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 ℓ-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_ℓ for the ℓ-adic representation associated to s; thus s_ℓ is a representation of G_K on a free $\mathbb{Z}_\ell$-module $\mathcal{T}_\ell$ 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 ℓ-adic representations) to the given $s_\ell \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ (for all ℓ). 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 *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 $\mathcal{T}_\ell$ (for all ℓ), i.e., such that $\sigma_{\mathcal{A}_g/K}(B) = s$. 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 ℓ, there is an isomorphism $T_\ell(A) \cong \mathcal{T}_\ell$. Since $\mathrm{End}(A) = \mathbb{Z}$, [Zarhin 1985, Corollary 5.4.5] implies that $A[\ell]$ is an absolutely simple Galois module for almost all ℓ, and hence that for almost all ℓ, 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 $\mathcal{T}_\ell$ for almost all ℓ. 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 $\mathcal{T}_\ell$ are isomorphic for all ℓ.

In order to prove the uniqueness of such a B up to an isomorphism, first, notice that $\mathrm{End}(B) = \mathbb{Z}$. 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 ℓ. 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],

$$\mathrm{Hom}(B, C) = \mathrm{Hom}_{G_K}(T_\ell(B), T_\ell(C)).$$

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

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

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

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

$$\mathcal{T}_\ell \times \mathcal{T}_\ell \to \mathbb{Z}_\ell(1).$$

Since $\mathcal{T}_\ell$ 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

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

This implies that the Galois modules $T_\ell(B)$ and $\mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(B), \mathbb{Z}_\ell(1))$ 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 ℓ. This implies that the abelian varieties $\hat{B}$ and B are isomorphic. Since $\mathrm{End}(B) = \mathbb{Z}$, there is an *isomorphism* $\mu : B \to \hat{B}$ such that $\mathrm{Hom}(B, \hat{B}) = \mathbb{Z} \cdot \mu$. 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 μ by $-\mu$, we may and will assume that n is a *positive* integer. It follows from [Mumford 1970, Section 23, Theorem 3] that μ is a polarization, which is obviously principal. (Clearly, there is exactly one principal polarization on B, namely μ.) So, $\sigma_{\mathcal{A}_g/K}(B)$ is defined and obviously coincides with s. $\square$

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

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

which is an *alternating* Galois-equivariant $\mathbb{Z}_\ell$-bilinear perfect/unimodular form on the free $\mathbb{Z}_\ell$-module $T_\ell(B)$. Since $\mathrm{End}(B) = \mathbb{Z}$, the already cited result of Faltings implies that $\mathrm{End}_{G_K}(T_\ell(A)) = \mathbb{Z}_\ell$. 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 $\mathcal{T}_\ell$ 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 ℓ 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 $\mathbb{M}_2(k\otimes_{\mathbb{Q}}\mathbb{Q}_\ell)$ 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

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

that sends 1 to 1. This gives us the embedding

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

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

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

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 $\mathrm{End}_{G_L}(V_\ell(Y))$ coincides with the matrix algebra $\mathbb{M}_2\big(\mathrm{End}_{G_L}(W_\ell(Y))\big)$ of size 2 over the centralizer $\mathrm{End}_{G_L}(W_\ell(Y))$.

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 $D=\mathrm{End}^0(Y)$. 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

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

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

$$\mathbb{M}_2(\mathrm{End}_{G_L}(W_\ell(Y)))\cong\mathbb{M}_2(k\otimes_{\mathbb{Q}}\mathbb{Q}_\ell).$$

It follows that

$$\mathrm{End}_{G_L}(W_\ell(Y)) \cong k \otimes_{\mathbb{Q}} \mathbb{Q}_\ell.$$

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 $\mathrm{End}^0(X)$ 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 $\bar{K}$-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 $\bar{K}$-endomorphisms of A are defined over K. Now Raynaud's criterion [SGA 7_{I} 1972] (see also [Silverberg and Zarhin 1995]), implies that A has everywhere semistable reduction. On the other hand,

$$\dim_{\mathbb{Q}} \mathrm{End}^0(A) = \dim_{\mathbb{Q}} D = 8 > 4 = \dim(A).$$

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 $\mathrm{char}(\kappa(v)) \neq 2$ then all points of order 4 on $A(v)$ are defined over $\kappa(v)$; if $\mathrm{char}(\kappa(v)) \neq 3$ then all points of order 3 on $A(v)$ are defined over $\kappa(v)$. It follows from the theorem of Silverberg that all $\overline{\kappa(v)}$-endomorphisms of $A(v)$ are defined over $\kappa(v)$. For each v we get an embedding of $\mathbb{Q}$-algebras

$$D \cong \mathrm{End}^0(A) \hookrightarrow \mathrm{End}^0(A(v)).$$

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

Theorem 4.1. *If* $\ell := \mathrm{char}(\kappa(v)) \neq p$ *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 ℓ.

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

$$\pi\bar{\pi} = q_v,$$

where $\bar{\pi}$ is the complex conjugate of π.

Let $E = \mathbb{Q}(\pi)$ be the number field generated by π and let $\mathcal{O}_E$ be the ring of integers in E. Then E contains $\bar{\pi}$ and is isomorphic to the center of $\mathrm{End}^0(A(v))$ [Tate 1966; 1971]; one may view $\mathrm{End}^0(A(v))$ 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],

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

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

Clearly,

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

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

First assume that E has degree 4 and $\mathrm{End}^0(A(v))$ is a quaternion algebra. Then (thanks to the theorem of Hasse–Brauer–Noether) there exists a place w of E with residual characteristic ℓ and such that the localization $\mathrm{End}^0(A(v)) \otimes_E E_w$ 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

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

This implies that $\mathrm{End}^0(A(v)) \otimes_E E_w$ 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

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

In particular, the order of the class of $\mathrm{End}^0(A(v))$ 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 $\mathrm{End}^0(A(v))$ is denoted by E.

We have already seen that there exist, at least, two maximal ideals in $\mathcal{O}_E$ that lie above ℓ. 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

$$\mathrm{ord}_{w_i} : E^* \twoheadrightarrow \mathbb{Z}$$

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

$$\mathrm{ord}_{w_i}(\ell) = 1, \quad \mathrm{ord}_{w_i}(\pi) + \mathrm{ord}_{w_i}(\bar{\pi}) = \mathrm{ord}_{w_i}(q_v) = N.$$

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

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

In addition, the sum in $\mathbb{Q}/\mathbb{Z}$ of local invariants of $\mathrm{End}^0(A(v))$ 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 $\mathrm{End}^0(A(v))$ in the Brauer group of E is either 2 or 4, we conclude that the local invariants of $\mathrm{End}^0(A(v))$ at $\{w_1, w_2\}$ are either $\left\{\frac{1}{4} \bmod 1, \frac{3}{4} \bmod 1\right\}$ or $\left\{\frac{3}{4} \bmod 1, \frac{1}{4} \bmod 1\right\}$ (and in both cases the order of $\mathrm{End}^0(A(v))$ in the Brauer group of E is 4) or $\left\{\frac{1}{2} \bmod 1, \frac{1}{2} \bmod 1\right\}$. In the latter case it follows from the formula for the w_i-adic invariant of $\mathrm{End}^0(A(v))$ that

$$\mathrm{ord}_{w_i}(\pi) = \frac{N}{2} = \mathrm{ord}_{w_i}(\bar{\pi})$$

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

$$\frac{\pi^2}{q_v} = \frac{\pi^2}{\pi\bar{\pi}} = \frac{\pi}{\bar{\pi}}$$

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 π^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 $\mathrm{End}^0(A(v)) \otimes_E E_{w_1}$ in the Brauer group of the $E_{w_1} \cong \mathbb{Q}_\ell$ is 4. This implies that $\mathrm{End}^0(A(v)) \otimes_E E_{w_1}$ 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 $\mathrm{End}^0(A(v)) \otimes_E E_{w_1}$ 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

$$D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \to \mathrm{End}^0(A(v)) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \twoheadrightarrow \mathrm{End}^0(A(v)) \otimes_E E_{w_1}.$$

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

Theorem 4.2. *If $\ell := \mathrm{char}(\kappa(v)) \neq p$ 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

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

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

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

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

$$D \to \mathrm{End}^0(A_i)$$

that must be injective, since D is a *simple* $\mathbb{Q}$-algebra. This implies that each $\mathrm{End}^0(A_i)$ 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 $\dim(A_i) = 3$ then $\dim(A_j) = 1$, which could not be the case. This implies that $A(v)$ itself is isotypic. It follows that if $\ell = \mathrm{char}(\kappa(v)) \neq p$ 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 $\mathrm{char}(\kappa(v))$)

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

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* $\mathrm{char}(\kappa(v)) \neq r$ *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 $D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ 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_ℓ does *not* come from an abelian variety over K. Indeed, if the G_K-module W_ℓ 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_ℓ is de Rham. It is also clear that W_ℓ 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_ℓ 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_ℓ 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 ℓ-adic representations of G_K.*

If $v|\ell$ then W_ℓ 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_ℓ is also semisimple. On the other hand, we know that the centralizer

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

in particular, none of W_ℓ 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 $\mathbb{Q}$ that is ramified exactly at p and ∞. In particular, for each prime $\ell \neq p$ we have a $\mathbb{Q}_\ell$-algebra isomorphism

$$H \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \cong \mathbb{M}_2(\mathbb{Q}_\ell).$$

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 $\mathrm{End}^0(X)$ 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 $\mathrm{End}^0(B) \cong H$. In particular, B is absolutely simple. By the theorem of Faltings, if ℓ is a prime then the G_K-module $V_\ell(B)$ is semisimple and

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

In particular, if $\ell \neq p$ then $\mathrm{End}_{G_K}(V_\ell(B)) \cong \mathbb{M}_2(\mathbb{Q}_\ell)$ 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_ℓ 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_ℓ is semisimple and

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

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

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

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_ℓ. This implies that if v is an 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_ℓ has rational coefficients and does not depend on ℓ. In other words, U_ℓ is a weakly compatible system of (absolutely irreducible) ℓ-adic representations. As above, locally for each $v|\ell$ the G_{K_v}-module $V_\ell(B)$ is de Rham with Hodge weights 0 and -1 with weights g, which implies that U_ℓ 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 $v|\ell$ then U_ℓ 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_ℓ 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_ℓ 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. □

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) $\rho : \pi_1(\mathcal{M}_g) \to \mathrm{Out}(\pi_1(\bar{X}))$. This follows by considering the universal curve $\mathcal{C}_g$ of genus g together with the map $\mathcal{C}_g \to \mathcal{M}_g$, so X can be viewed as a fiber of this map. This gives rise to the fibration exact sequence

$$1 \to \pi_1(\bar{X}) \to \pi_1(\mathcal{C}_g) \to \pi_1(\mathcal{M}_g) \to 1$$

and the action of $\pi_1(\mathcal{C}_g)$ on $\pi_1(\bar{X})$ gives ρ. 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 $G_K \to \mathrm{Out}(\pi_1(\bar{X}))$ which is none other than the map obtained from the exact sequence (1) by letting $\pi_1(X)$ act on $\pi_1(\bar{X})$ 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

[André 2004] Y. André, *Une introduction aux motifs (motifs purs, motifs mixtes, périodes)*, Panoramas et Synthèses **17**, Société Mathématique de France, Paris, 2004. MR 2115000

[Bass et al. 1964] H. Bass, M. Lazard, and J.-P. Serre, "Sous-groupes d'indice fini dans $\mathbf{SL}(n, \mathbf{Z})$", *Bull. Amer. Math. Soc.* **70** (1964), 385–392. MR 0161913 Zbl 0232.20086

[Brinon and Conrad 2009] O. Brinon and B. Conrad, "CMI Summer School Notes on p-adic Hodge Theory", preprint, Clay Mathematics Institute, 2009, Available at http://www.claymath.org/galois-representations. Galois Representations (Honolulu, 2009).

[Deligne 1971] P. Deligne, "Formes modulaires et représentations l-adiques", exposé no. 355, 139–172 in *Séminaire Bourbaki* 1968/69, Lecture Notes in Math. **175**, Springer, Berlin, 1971. MR 3077124

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. MR 718935 Zbl 0588.14026

[Faltings 1984] G. Faltings, "Complements to Mordell", pp. 203–227 in *Rational points* (Bonn, 1983/1984), edited by G. Faltings and G. Wüstholz, Aspects Math. **E6**, Friedr. Vieweg & Sohn, Braunschweig, 1984. MR 766574

[Faltings 1989] G. Faltings, "Crystalline cohomology and p-adic Galois-representations", pp. 25–80 in *Algebraic analysis, geometry, and number theory* (Baltimore, MD, 1988), edited by J.-I. Igusa, Johns Hopkins Univ. Press, Baltimore, MD, 1989. MR 1463696 Zbl 0805.14008

[Fontaine and Mazur 1995] J.-M. Fontaine and B. Mazur, "Geometric Galois representations", pp. 41–78 in *Elliptic curves, modular forms, & Fermat's last theorem*, 1st ed. (Hong Kong, 1993), edited by J. Coates, Int. Press, Cambridge, MA, 1995. MR 1363495 Zbl 0839.14011

[Fontaine and Ouyang 2007] J.-M. Fontaine and Y. Ouyang, "Theory of p-adic representations", preprint, 2007, Available at http://www.math.u-psud.fr/∼fontaine/recherche.html.

[Grothendieck 1997a] A. Grothendieck, "Brief an G. Faltings", pp. 49–58 in *Geometric Galois actions, 1*, edited by L. Schneps and P. Lochak, London Math. Soc. Lecture Note Ser. **242**, Cambridge Univ. Press, 1997. MR 1483108 Zbl 0901.14002

[Grothendieck 1997b] A. Grothendieck, "Esquisse d'un programme", pp. 5–48 in *Geometric Galois actions, 1*, edited by L. Schneps and P. Lochak, London Math. Soc. Lecture Note Ser. **242**, Cambridge Univ. Press, 1997. MR 1483107 Zbl 0901.14001

[Hain 2011] R. Hain, "Rational points of universal curves", *J. Amer. Math. Soc.* **24**:3 (2011), 709–769. MR 2784328 Zbl 1225.14016

[Harari and Stix 2012] D. Harari and J. Stix, "Descent obstruction and fundamental exact sequence", pp. 147–166 in *The arithmetic of fundamental groups: PIA 2010*, edited by J. Stix, Contrib. Math. Comput. Sci. **2**, Springer, Heidelberg, 2012. MR 3220518 Zbl 1315.14032

[Harari and Voloch 2010] D. Harari and J. F. Voloch, "The Brauer–Manin obstruction for integral points on curves", *Math. Proc. Cambridge Philos. Soc.* **149**:3 (2010), 413–421. MR 2726726 Zbl 1280.11038

[Helm and Voloch 2011] D. Helm and J. F. Voloch, "Finite descent obstruction on curves and modularity", *Bull. Lond. Math. Soc.* **43**:4 (2011), 805–810. MR 2820165 Zbl 1235.11062

[Ihara and Nakamura 1997] Y. Ihara and H. Nakamura, "Some illustrative examples for anabelian geometry in high dimensions", pp. 127–138 in *Geometric Galois actions, 1*, edited by L. Schneps and P. Lochak, London Math. Soc. Lecture Note Ser. **242**, Cambridge Univ. Press, 1997. MR 1483114 Zbl 0919.14011

[Jannsen 1992] U. Jannsen, "Motives, numerical equivalence, and semi-simplicity", *Invent. Math.* **107**:3 (1992), 447–452. MR 1150598 Zbl 0762.14003

[Lang 1959] S. Lang, *Abelian varieties*, Interscience Tracts in Pure and Appl. Math. **7**, Interscience, New York, 1959. Reprinted Springer 1983. MR 0106225 Zbl 0099.16103

[Matsumoto and Tamagawa 2000] M. Matsumoto and A. Tamagawa, "Mapping-class-group action versus Galois action on profinite fundamental groups", *Amer. J. Math.* **122**:5 (2000), 1017–1026. MR 1781929 Zbl 0993.12002

[Mazur 1999] B. Mazur, "Open problems in number theory", pp. 199–203 in *Current developments in mathematics* (Cambridge, MA, 1997), edited by R. Bott et al., Int. Press, Boston, 1999. MR 1700300

[Mennicke 1965] J. L. Mennicke, "Finite factor groups of the unimodular group", *Ann. of Math.* (2) **81** (1965), 31–37. MR 0171856 Zbl 0135.06504

[Mochizuki 1996] S. Mochizuki, "The profinite Grothendieck conjecture for closed hyperbolic curves over number fields", *J. Math. Sci. Univ. Tokyo* **3**:3 (1996), 571–627. MR 1432110 Zbl 0889.11020

[Mochizuki 2003] S. Mochizuki, "Topics surrounding the anabelian geometry of hyperbolic curves", pp. 119–165 in *Galois groups and fundamental groups*, edited by L. Schneps, Math. Sci. Res. Inst. Publ. **41**, Cambridge Univ. Press, 2003. MR 2012215 Zbl 1053.14029

[Mumford 1966] D. Mumford, "On the equations defining abelian varieties, I", *Invent. Math.* **1** (1966), 287–354. MR 0204427 Zbl 0219.14024

[Mumford 1970] D. Mumford, *Abelian varieties*, Tata Inst. Fund. Res. Stud. Math. **5**, Oxford University Press, London, 1970. MR 0282985 Zbl 0223.14022

[Noot 1995] R. Noot, "Abelian varieties — Galois representation and properties of ordinary reduction", *Compositio Math.* **97**:1-2 (1995), 161–171. MR 1355123 Zbl 0868.14021

[Oort 1987] F. Oort, "Lifting algebraic curves, abelian varieties, and their endomorphisms to characteristic zero", pp. 165–195 in *Algebraic geometry: Bowdoin, 1985* (Brunswick, ME, 1985), edited by S. J. Bloch, Proc. Sympos. Pure Math. **46**, Amer. Math. Soc., Providence, RI, 1987. MR 927980 Zbl 0645.14017

[Oort 1988] F. Oort, "Endomorphism algebras of abelian varieties", pp. 469–502 in *Algebraic geometry and commutative algebra*, vol. II, edited by H. Hijikata et al., Kinokuniya, Tokyo, 1988. MR 977774 Zbl 0697.14029

[Oort 1992] F. Oort, "CM-liftings of abelian varieties", *J. Algebraic Geom.* **1**:1 (1992), 131–146. MR 1129842 Zbl 0803.14025

[Oort and Zarhin 1995] F. Oort and Y. Zarhin, "Endomorphism algebras of complex tori", *Math. Ann.* **303**:1 (1995), 11–29. MR 1348352 Zbl 0858.14024

[Pál 2011] A. Pál, "The real section conjecture and Smith's fixed-point theorem for pro-spaces", *J. Lond. Math. Soc.* (2) **83**:2 (2011), 353–367. MR 2776641 Zbl 1263.14056

[Patrikis 2012] S. Patrikis, *Variations on a theorem of Tate*, Ph.D. thesis, Princeton University, 2012, Available at http://search.proquest.com/docview/1040698945. MR 3078437 arXiv 1207.6724

[Ribet 1976] K. A. Ribet, "Galois action on division points of Abelian varieties with real multiplications", *Amer. J. Math.* **98**:3 (1976), 751–804. MR 0457455 Zbl 0348.14022

[Serre 1989] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, 2nd ed., Addison-Wesley, Redwood City, CA, 1989. MR 1043865 Zbl 0709.14002

[SGA $4\frac{1}{2}$ 1977] P. Deligne, *Cohomologie étale* (Séminaire de Géométrie Algébrique du Bois Marie), Lecture Notes in Math. **569**, Springer, Berlin, 1977. MR 0463174 Zbl 0345.00010

[SGA 7_{I} 1972] A. Grothendieck, "Exposé IX: modèles de Néron et monodromie", pp. 313–523 in *Groupes de monodromie en géométrie algébrique, I* (Séminaire de Géométrie Algébrique du Bois Marie 1967–1969), edited by A. Grothendieck et al., Lecture Notes in Math. **288**, Springer, Berlin, 1972. MR 0354656 Zbl 0237.00013

[Shimura 1963] G. Shimura, "On analytic families of polarized abelian varieties and automorphic functions", *Ann. of Math.* (2) **78** (1963), 149–192. MR 0156001 Zbl 0142.05402

[Silverberg 1992] A. Silverberg, "Fields of definition for homomorphisms of abelian varieties", *J. Pure Appl. Algebra* **77**:3 (1992), 253–262. MR 1154704 Zbl 0808.14037

[Silverberg and Zarhin 1995] A. Silverberg and Y. G. Zarhin, "Semistable reduction and torsion subgroups of abelian varieties", *Ann. Inst. Fourier* (*Grenoble*) **45**:2 (1995), 403–420. MR 1343556 Zbl 0818.14017

[Stix 2013] J. Stix, *Rational points and arithmetic of fundamental groups: evidence for the section conjecture*, Lecture Notes in Math. **2054**, Springer, Heidelberg, 2013. MR 2977471 Zbl 1272.14003

[Stoll 2007] M. Stoll, "Finite descent obstructions and rational points on curves", *Algebra Number Theory* **1**:4 (2007), 349–391. MR 2368954 Zbl 1167.11024

[Tate 1966] J. Tate, "Endomorphisms of abelian varieties over finite fields", *Invent. Math.* **2** (1966), 134–144. MR 0206004 Zbl 0147.20303

[Tate 1971] J. Tate, "Classes d'isogénie des variétés abéliennes sur un corps fini (d'après T. Honda)", exposé no. 352, 95–110 in *Séminaire Bourbaki* 1968/69, Lecture Notes in Math. **175**, Springer, Berlin, 1971. MR 3077121 Zbl 0212.25702

[Tate 1994] J. Tate, "Conjectures on algebraic cycles in l-adic cohomology", pp. 71–83 in *Motives* (Seattle, WA, 1991), edited by U. Jannsen et al., Proc. Sympos. Pure Math. **55**, Amer. Math. Soc., Providence, RI, 1994. MR 1265523 Zbl 0814.14009

[Voloch 2012] J. F. Voloch, "Finite descent obstruction for curves over function fields", *Bull. Braz. Math. Soc.* (*N.S.*) **43**:1 (2012), 1–6. MR 2909919 Zbl 1300.11073

[Zarhin 1985] Y. G. Zarhin, "A finiteness theorem for unpolarized abelian varieties over number fields with prescribed places of bad reduction", *Invent. Math.* **79**:2 (1985), 309–321. MR 778130 Zbl 0557.14024

[Zarhin 2015] Y. G. Zarhin, "Eigenvalues of Frobenius endomorphisms of abelian varieties of low dimension", *J. Pure Appl. Algebra* **219**:6 (2015), 2076–2098. MR 3299720 Zbl 06398749

[Zoonekynd 2001] V. Zoonekynd, "The fundamental group of an algebraic stack", preprint, 2001. arXiv math/0111071

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1221

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 (φ, Γ)-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 $G_K := \mathrm{Gal}(\overline{K}/K)$. The local Tamagawa number conjecture is a statement describing a certain $\mathbb{Q}_p$-basis of the determinant line $\det_{\mathbb{Q}_p} R\Gamma(K, V)$ 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 ϵ-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 $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ [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 $\mathbb{Z}_p[\mathrm{Gal}(K/\mathbb{Q}_p)]$, 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

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

Here $\zeta = (\zeta_{p^n})_n \in \Gamma(\overline{\mathbb{Q}}_p, \mathbb{Z}_p(1))$ is a compatible system of p^n-th roots of unity which we fix throughout this paper. The conjectures for a triple (Λ, T, ζ) and its dual $(\Lambda^{\mathrm{op}}, T^*(1), \zeta)$ 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]

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

In a first approximation one may say that the local Tamagawa number conjecture describes the relation between the two $\mathbb{Z}_p$-lattices $\exp(\mathcal{O}_K)$ and $\mathrm{im}(H^1(K, \mathbb{Z}_p(r)))$ 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 $\mathrm{Det}_{\mathbb{Z}_p} \exp(\mathcal{O}_K)$ to $\mathrm{Det}_{\mathbb{Z}_p}(\mathrm{im}(H^1(K, \mathbb{Z}_p(r))))$ inside the $\mathbb{Q}_p$-line $\mathrm{Det}_{\mathbb{Q}_p} H^1(K, \mathbb{Q}_p(r))$. 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 $\mathrm{im}(H^1(K, \mathbb{Z}_p(r)))$ 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 $K_1(\mathbb{Q}_p^{\mathrm{ur}}[G]))/K_1(\mathbb{Z}_p^{\mathrm{ur}}[G]))$ 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 $\mathrm{im}(H^1(K, \mathbb{Z}_p(r)))$ by the entire perfect complex $R\Gamma(K, \mathbb{Z}_p(r))$. There still is an isomorphism

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

since the groups $H^1(K, \mathbb{Z}_p(r))_{\mathrm{tor}}$ 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 $\mathrm{im}(H^1(K, \mathbb{Z}_p(r)))$ or $\mathcal{O}_K$ 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 $\exp(\mathcal{O}_K)$ directly. Instead one uses the "period isomorphism"

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

and tries to compare $\mathrm{Det}_{\mathbb{Z}_p} R\Gamma(K, \mathbb{Z}_p(r))$ to a suitable lattice in this last space. The left-$\mathbb{Z}_p[G]$-module $\mathrm{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \mathbb{Z}_p$ is always free of rank one whereas $\mathcal{O}_K$ need not be. After choosing an embedding $K \to \overline{\mathbb{Q}}_p$ one gets an isomorphism $\psi : G_{\mathbb{Q}_p}/G_K \cong G$

and an isomorphism

$$\operatorname{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \mathbb{Z}_p \cong \mathbb{Z}_p[G] \tag{2}$$

so that the $\mathbb{Z}_p[G]$-linear left action of $\gamma \in G_{\mathbb{Q}_p}$ is given by

$$\mathbb{Z}_p[G] \ni x \mapsto x\psi(\gamma^{-1}). \tag{3}$$

The period isomorphism is then given for $x \in K$ by

$$\operatorname{per}(x) := \operatorname{per}(1 \otimes x) = \sum_{g \in G} g(x) \cdot g^{-1} \in \overline{\mathbb{Q}}_p[G].$$

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_β 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_β 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 $K_0(\mathbb{Z}_p[G], \mathbb{Q}_p)$ 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 $K_1(\mathbb{Q}_p[G])/\operatorname{im}(K_1(\mathbb{Z}_p[G]))$. Extending scalars to $\overline{\mathbb{Q}}_p$ we get an isomorphism of free rank-one $\overline{\mathbb{Q}}_p[G]$-modules

$$H^1(K, \mathbb{Q}_p(1-r)) \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\exp^* \otimes \overline{\mathbb{Q}}_p} K \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\operatorname{per}} \overline{\mathbb{Q}}_p[G]$$

sending the $\overline{\mathbb{Q}}_p[G]$-basis β to a unit $\operatorname{per}(\exp^*(\beta)) \in \overline{\mathbb{Q}}_p[G]^\times$. As such it has a class

$$[\operatorname{per}(\exp^*(\beta))] \in K_1(\overline{\mathbb{Q}}_p[G])$$

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 $R^\times \to \mathrm{GL}(R) \to \mathrm{GL}(R)^{\mathrm{ab}} =: K_1(R)$). In Section 2.2 below we shall define an ϵ-factor $\epsilon(K/\mathbb{Q}_p, 1-r) \in K_1(\overline{\mathbb{Q}}_p[G])$ such that

$$\epsilon(K/\mathbb{Q}_p, 1-r) \cdot [\operatorname{per}(\exp^*(\beta))] \in K_1(\mathbb{Q}_p^{\mathrm{ur}}[G]).$$

Let $F \subseteq K$ denote the maximal unramified subfield, $\Sigma = \operatorname{Gal}(F/\mathbb{Q}_p)$ 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*

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

is equivalent to the identity

$$[(r-1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1-r) \cdot [\operatorname{per}(\exp^*(\beta))] \cdot [C_\beta]^{-1} \cdot \left[\frac{1-p^{r-1}\sigma}{1-p^{-r}\sigma^{-1}}\right]_F = 1 \quad (4)$$

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

Before we begin the proof of the proposition we explain what we mean by the local Tamagawa number conjecture for $(\mathbb{Z}_p[G], \operatorname{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \mathbb{Z}_p(1-r), \zeta)$. The local Tamagawa number conjecture [Fukaya and Kato 2006, Conjecture 3.4.3] claims the existence of ϵ-isomorphisms $\epsilon_{\Lambda,\zeta}(T)$ for all triples (Λ, T, ζ), where Λ is a semilocal pro-p ring satisfying a certain finiteness condition [Fukaya and Kato 2006, 1.4.1], T a finitely generated projective Λ-module with continuous $G_{\mathbb{Q}_p}$-action and ζ a basis of $\Gamma(\overline{\mathbb{Q}}_p, \mathbb{Z}_p(1))$, 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], $S \otimes_R - : C_R \to C_S$ is the Picard functor induced by a ring homomorphism $R \to S$ and $\widetilde{R} = W(\bar{\mathbb{F}}_p) \otimes_{\mathbb{Z}_p} R$ 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 (Λ, T, ζ) if

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

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

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

$$P^* = \operatorname{Hom}_{\mathbb{Q}_p[G]}(P, \mathbb{Q}_p[G]),$$

which is a perfect complex of $\mathbb{Q}_p[G]^{\mathrm{op}}$-modules. Fix $r \geq 2$ and set

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

which are free of rank one over $\mathbb{Q}_p[G]$ and $\mathbb{Q}_p[G]^{\mathrm{op}}$, 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 ζ determines an element $t = \log(\zeta)$ of B_{dR}. We have

$$D_{\mathrm{cris}}(V) = F \cdot t^{r-1}, \quad D_{\mathrm{dR}}(V)/D^0_{\mathrm{dR}}(V) = 0,$$
$$D_{\mathrm{cris}}(V^*(1)) = F \cdot t^{-r}, \quad D_{\mathrm{dR}}(V^*(1))/D^0_{\mathrm{dR}}(V^*(1)) = K,$$
$$C_f(\mathbb{Q}_p, V) : F \xrightarrow{1-p^{r-1}\sigma} F,$$
$$C_f(\mathbb{Q}_p, V^*(1)) : F \xrightarrow{(1-p^{-r}\sigma,\subseteq)} F \oplus K,$$

and commutative diagrams

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[r, "\eta(\mathbb{Q}_p,V)"] & \mathrm{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p,V)\cdot \mathrm{Det}_{\mathbb{Q}_p[G]} D_{\mathrm{dR}}(V)/D^0_{\mathrm{dR}}(V) \\
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[u, "{[1-p^{r-1}\sigma]_F^{-1}}"] \arrow[r, "\eta'(\mathbb{Q}_p,V)"] & \mathrm{Det}_{\mathbb{Q}_p[G]}(0)\cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(0)^{-1}\cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[u, "c"']
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[r, "\eta(\mathbb{Q}_p,V^*(1))^{*,-1}"] & \begin{array}{l}\mathrm{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p,V^*(1))^* \\ \times(\mathrm{Det}_{\mathbb{Q}_p[G]} D_{\mathrm{dR}}(V^*(1))/D^0_{\mathrm{dR}}(V^*(1)))^*\end{array} \\
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[u, "{[1-p^{-r}\sigma^{-1}]_F}"'] \arrow[r, "\eta'(\mathbb{Q}_p,V^*(1))^{*,-1}"] & \mathrm{Det}_{\mathbb{Q}_p[G]}(0)\cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(K^*)^{-1}\cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(K^*) \arrow[u, "c"']
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p[G]} C_f(\mathbb{Q}_p,V^*(1))^* \arrow[r, "\mathrm{Det}_{\mathbb{Q}_p[G]} \Psi_f(\mathbb{Q}_p,V^*(1))^{*,-1}"] & \mathrm{Det}_{\mathbb{Q}_p[G]}\big(C(\mathbb{Q}_p,V)/C_f(\mathbb{Q}_p,V)\big) \\
\mathrm{Det}_{\mathbb{Q}_p[G]}(K^*)^{-1} \arrow[u, "c"] \arrow[r, "\Psi'"] & \mathrm{Det}_{\mathbb{Q}_p[G]} H^\bullet(\mathbb{Q}_p,V) \arrow[u, "c"]
\end{tikzcd}$$

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

$$\theta_{\mathbb{Q}_p[G]}(V) = \eta(\mathbb{Q}_p, V)\cdot \big(\mathrm{Det}_{\mathbb{Q}_p[G]} \Psi_f(\mathbb{Q}_p, V^*(1))^{*,-1} \circ \eta(\mathbb{Q}_p, V^*(1))^{*,-1}\big)$$

we obtain a commutative diagram

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[r, "\theta_{\mathbb{Q}_p[G]}(V)"] & \mathrm{Det}_{\mathbb{Q}_p[G]} C(\mathbb{Q}_p, V) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]} D_{\mathrm{dR}}(V) \\
\mathrm{Det}_{\mathbb{Q}_p[G]}(0) \arrow[r, "\theta'"] \arrow[u, "{\left[\frac{1-p^{-r}\sigma^{-1}}{1-p^{r-1}\sigma}\right]_F}"] & \mathrm{Det}_{\mathbb{Q}_p[G]} H^\bullet(\mathbb{Q}_p, V) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(K) \arrow[u, "c"']
\end{tikzcd}$$

where θ' is induced by the dual exponential map

$$H^1(\mathbb{Q}_p, V) \xrightarrow{\exp^*_{V^*(1)}} K.$$

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

$$[(-1)^{r-1}(r-1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1-r) \cdot \mathrm{Det}_{\overline{\mathbb{Q}}_p[G]}(\mathrm{per})$$

and the isomorphism

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

fits into a commutative diagram

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p^{\mathrm{ur}}[G]}(0) \arrow[r, "\epsilon_{\mathbb{Q}_p[G],\zeta}(V)"] & \mathbb{Q}_p^{\mathrm{ur}}[G] \underset{\mathbb{Q}_p[G]}{\otimes} \big(\mathrm{Det}_{\mathbb{Q}_p[G]} R\Gamma(K, \mathbb{Q}_p(1-r)) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(V)\big) \\
\mathrm{Det}_{\mathbb{Q}_p^{\mathrm{ur}}[G]}(0) \arrow[r, "\theta''"] \arrow[u, "{\left[\frac{1-p^{-r}\sigma^{-1}}{1-p^{r-1}\sigma}\right]_F}"'] & \mathbb{Q}_p^{\mathrm{ur}}[G] \underset{\mathbb{Q}_p[G]}{\otimes} \big(\mathrm{Det}^{-1}_{\mathbb{Q}_p[G]} H^1(K, \mathbb{Q}_p(1-r)) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(\mathbb{Q}_p[G])\big) \arrow[u, "c"']
\end{tikzcd}$$

where

$$\theta'' = [(-1)^{r-1}(r-1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1-r) \cdot \mathrm{Det}_{\overline{\mathbb{Q}}_p[G]}(\mathrm{per}) \cdot \theta'$$

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

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

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

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

combined with the acyclicity isomorphism

$$\mathrm{can} : \mathrm{Det}_{\mathbb{Q}_p[G]}(0) \cong \mathrm{Det}_{\mathbb{Q}_p[G]}(C_{\beta,\mathbb{Q}_p}).$$

Since the class of C_β in $K_0(\mathbb{Z}_p[G])$ vanishes, we can choose an isomorphism

$$a : \mathrm{Det}_{\mathbb{Z}_p[G]}(0) \cong \mathrm{Det}_{\mathbb{Z}_p[G]}(C_\beta),$$

which leads to another isomorphism

$$c' : \mathrm{Det}^{-1}_{\mathbb{Z}_p[G]}(\mathbb{Z}_p[G]\cdot\beta) \cong \mathrm{Det}_{\mathbb{Z}_p[G]} R\Gamma(K, \mathbb{Z}_p(1-r))$$

defined over $\mathbb{Z}_p[G]$. Setting

$$\lambda := (c'_{\mathbb{Q}_p})^{-1} c \in \mathrm{Aut}\big(\mathrm{Det}^{-1}_{\mathbb{Q}_p[G]} H^1(K, \mathbb{Q}_p(1-r))\big) = K_1(\mathbb{Q}_p[G]),$$

we obtain a commutative diagram

$$\begin{tikzcd}
\mathrm{Det}_{\mathbb{Q}_p^{\mathrm{ur}}[G]}(0) \arrow[r, "\epsilon_{\mathbb{Q}_p[G],\zeta}(V)"] & \mathbb{Q}_p^{\mathrm{ur}}[G] \underset{\mathbb{Q}_p[G]}{\otimes} \big(\mathrm{Det}_{\mathbb{Q}_p[G]} R\Gamma(K, \mathbb{Q}_p(1-r)) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(V)\big) \\
\mathrm{Det}_{\mathbb{Q}_p^{\mathrm{ur}}[G]}(0) \arrow[u, "\left[\frac{1-p^{-r}\sigma^{-1}}{1-p^{r-1}\sigma}\right]_F"] \arrow[r, "\theta'''"] & \mathbb{Q}_p^{\mathrm{ur}}[G] \underset{\mathbb{Q}_p[G]}{\otimes} \big(\mathrm{Det}^{-1}_{\mathbb{Q}_p[G]} H^1(K, \mathbb{Q}_p(1-r)) \cdot \mathrm{Det}_{\mathbb{Q}_p[G]}(\mathbb{Q}_p[G])\big) \arrow[u, "c'_{\mathbb{Q}_p}"']
\end{tikzcd}$$

where

$$\theta''' = \lambda \circ \theta'' = \lambda \cdot [(-1)^{r-1}(r-1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1-r) \cdot \mathrm{Det}_{\overline{\mathbb{Q}}_p[G]}(\mathrm{per}) \cdot \theta'.$$

The local Tamagawa number conjecture claims that $\epsilon_{\mathbb{Q}_p[G],\zeta}(V)$ is induced by an isomorphism

$$\mathrm{Det}_{\mathbb{Z}_p^{\mathrm{ur}}[G]}(0) \xrightarrow{\epsilon_{\mathbb{Z}_p[G],\zeta}(T)} \mathbb{Z}_p^{\mathrm{ur}}[G] \underset{\mathbb{Z}_p[G]}{\otimes} \big(\mathrm{Det}_{\mathbb{Z}_p[G]} R\Gamma(K, \mathbb{Z}_p(1-r)) \cdot \mathrm{Det}_{\mathbb{Z}_p[G]}(T)\big)$$

and this will be the case if and only if

$$\theta^{\mathrm{iv}} := \theta''' \cdot \left[\frac{1-p^{r-1}\sigma}{1-p^{-r}\sigma^{-1}}\right]_F$$

is induced by an isomorphism

$$\mathrm{Det}_{\mathbb{Z}_p^{\mathrm{ur}}[G]}(0) \xrightarrow{\theta^{\mathrm{iv}}_{\mathbb{Z}_p[G]}} \mathbb{Z}_p^{\mathrm{ur}}[G] \underset{\mathbb{Z}_p[G]}{\otimes} \big(\mathrm{Det}^{-1}_{\mathbb{Z}_p[G]}(\mathbb{Z}_p[G]\cdot\beta) \cdot \mathrm{Det}_{\mathbb{Z}_p[G]}(\mathbb{Z}_p[G])\big).$$

The isomorphism of $\overline{\mathbb{Q}}_p[G]$-modules

$$\tau : H^1(K, \mathbb{Q}_p(1-r)) \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\exp^* \otimes \overline{\mathbb{Q}}_p} K \otimes_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p \xrightarrow{\mathrm{per}} \overline{\mathbb{Q}}_p[G] \xrightarrow{\cdot\, \mathrm{per}(\exp^*(\beta))^{-1}} \overline{\mathbb{Q}}_p[G]$$

is clearly induced by an isomorphism of $\mathbb{Z}_p[G]$-modules

$$\tau_{\mathbb{Z}_p[G]} : \mathbb{Z}_p[G]\cdot\beta \xrightarrow{\sim} \mathbb{Z}_p[G]$$

and we have

$$\theta^{\mathrm{iv}} = \left[\frac{1-p^{r-1}\sigma}{1-p^{-r}\sigma^{-1}}\right]_F \cdot \lambda \cdot [(-1)^{r-1}(r-1)!] \cdot \epsilon(K/\mathbb{Q}_p, 1-r)\cdot[\mathrm{per}(\exp^*(\beta))]\cdot \mathrm{Det}_{\overline{\mathbb{Q}}_p[G]}(\tau).$$

Hence θ^{iv} is induced by an isomorphism $\theta^{\mathrm{iv}}_{\mathbb{Z}_p[G]}$ if and only if the class in $K_1(\mathbb{Q}_p^{\mathrm{ur}}[G])$ of

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

lies in $K_1(\mathbb{Z}_p^{\mathrm{ur}}[G])$. Now note that $[(-1)] \in K_1(\mathbb{Z}) \subset K_1(\mathbb{Z}_p^{\mathrm{ur}}[G])$ and that $\lambda = [C_\beta]^{-1}$, so we do indeed obtain identity (4). In order to see this last identity, note that we have

$$\lambda^{-1} = a^{-1}\cdot \mathrm{can}$$

and that $a^{-1}\cdot\mathrm{can} \in K_1(\mathbb{Q}_p^{\mathrm{ur}}[G])$ is a lift of $[C_\beta] \in K_0(\mathbb{Z}_p^{\mathrm{ur}}[G], \mathbb{Q}_p)$ according to the conventions of [Fukaya and Kato 2006, 1.3.8, Theorem 1.3.15(ii)]. □

2.1. ***Description of K_1.*** For any finite group G we have the Wedderburn decomposition

$$\overline{\mathbb{Q}}_p[G] \cong \prod_{\chi\in\widehat{G}} M_{d_\chi}(\overline{\mathbb{Q}}_p),$$

where $\widehat{G}$ is the set of irreducible $\overline{\mathbb{Q}}_p$-valued characters of G and $d_\chi = \chi(1)$ is the degree of χ. Hence there is a corresponding decomposition

$$K_1(\overline{\mathbb{Q}}_p[G]) \cong \prod_{\chi\in\widehat{G}} K_1(M_{d_\chi}(\overline{\mathbb{Q}}_p)) \cong \prod_{\chi\in\widehat{G}} \overline{\mathbb{Q}}_p^\times, \tag{5}$$

which allows one to think of $K_1(\overline{\mathbb{Q}}_p[G])$ as a collection of nonzero p-adic numbers indexed by $\widehat{G}$. 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 $\chi \in \widehat{G}$ take values in $\mathbb{Z}_p^{\mathrm{ur}}$, the Wedderburn decomposition is already defined over $\mathbb{Z}_p^{\mathrm{ur}}$ and so is the decomposition of K_1. One has

$$K_1(\mathbb{Z}_p^{\mathrm{ur}}[G]) \cong \prod_{\chi\in\widehat{G}} K_1(M_{d_\chi}(\mathbb{Z}_p^{\mathrm{ur}})) \cong \prod_{\chi\in\widehat{G}} \mathbb{Z}_p^{\mathrm{ur},\times}$$

and

$$K_1(\mathbb{Q}_p^{\mathrm{ur}}[G])/\operatorname{im}(K_1(\mathbb{Z}_p^{\mathrm{ur}}[G])) \cong \prod_{\chi\in\widehat{G}} \mathbb{Q}_p^{\mathrm{ur},\times}/\mathbb{Z}_p^{\mathrm{ur},\times} \cong \prod_{\chi\in\widehat{G}} p^{\mathbb{Z}}, \tag{6}$$

which allows one to think of elements in $K_1(\mathbb{Q}_p^{\mathrm{ur}}[G])/\operatorname{im}(K_1(\mathbb{Z}_p^{\mathrm{ur}}[G]))$ as a collection of integers (p-adic valuations) indexed by $\widehat{G}$.

2.2. *Definition of the ϵ-factor.* If L is a local field, E an algebraically closed field of characteristic 0 with the discrete topology, μ_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 ϵ-factor $\epsilon(r, \psi_L, \mu_L) \in E^\times$.

We shall take $E = \overline{\mathbb{Q}}_p$ and always fix μ_L and ψ_L so that $\mu_L(\mathcal{O}_L) = 1$ and $\psi_L = \psi_{\mathbb{Q}_p} \circ \mathrm{Tr}_{L/\mathbb{Q}_p}$ where $\psi_{\mathbb{Q}_p}(p^{-n}) = \zeta_{p^n}$ for our fixed $\zeta = (\zeta_{p^n})_n \in \Gamma(\overline{\mathbb{Q}}_p, \mathbb{Z}_p(1))$. Setting

$$\epsilon(r) := \epsilon(r, \psi_L, \mu_L) \in E^\times$$

and leaving the dependence on ζ 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 π be a uniformizer of $\mathcal{O}_L$, δ_L the exponent of the different of $L/\mathbb{Q}_p$ and $q = |\mathcal{O}_L/\pi|$.

(a) If $r : W_L \to E^\times$ is a homomorphism, set

$$r_\sharp : L^\times \xrightarrow{\text{rec}} W_L^{\text{ab}} \xrightarrow{r} E^\times$$

where rec is normalized as in [Deligne 1973, (2.3)] and sends a uniformizer to a *geometric* Frobenius automorphism in W_L^{ab}. Then we have

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

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

$$\tau(r_\sharp, \psi_\pi) = \sum_{u \in (\mathcal{O}_L/\pi^c)^\times} r_\sharp^{-1}(u) \psi_\pi(u) \tag{7}$$

is the Gauss sum associated to the restriction of $r_\sharp$ to $(\mathcal{O}_L/(\pi^c))^\times$ and the additive character

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

of $\mathcal{O}_L/(\pi^c)$.

(b) If L/K is unramified then $\epsilon(r) = \epsilon(\mathrm{Ind}_{W_L}^{W_K} r)$ 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 $D_{\mathrm{pst}}(V)$, a finite dimensional $\widehat{\mathbb{Q}}_p^{\mathrm{ur}}$-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 $\widehat{\mathbb{Q}}_p^{\mathrm{ur}}$ and discrete on the inertia subgroup. Moreover, $D_{\mathrm{pst}}(V)$ has a Frob-semilinear automorphism φ. The associated linear representation r_V of $W_{\mathbb{Q}_p}$ over $E = \widehat{\mathbb{Q}}_p^{\mathrm{ur}}$ is the space $D_{\mathrm{pst}}(V)$ with action

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

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

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

$$D_{\mathrm{pst}}(V) = (\mathrm{Ind}_{G_K}^{G_{\mathbb{Q}_p}} \widehat{\mathbb{Q}}_p^{\mathrm{ur}}) \cdot t^{r-1}, \qquad r_V = (\mathrm{Ind}_{W_K}^{W_{\mathbb{Q}_p}} \widehat{\mathbb{Q}}_p^{\mathrm{ur}})(p^{1-r}),$$

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

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

For each $\chi \in \widehat{G}$ define a representation r_χ of $W_{\mathbb{Q}_p}$ over $E = \overline{\mathbb{Q}}_p$ by

$$W_{\mathbb{Q}_p} \xrightarrow{\iota} G_{\mathbb{Q}_p} \xrightarrow{\psi} G \xrightarrow{\rho_\chi} \mathrm{GL}_{d_\chi}(E), \tag{8}$$

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

$$r_{V,\chi} = E^{d_\chi} \otimes_{\mathbb{Q}_p^{\mathrm{ur}}[G]} r_V = E^{d_\chi} \otimes_{\mathbb{Q}_p^{\mathrm{ur}}[G]} (\mathrm{Ind}_{W_K}^{W_{\mathbb{Q}_p}} \mathbb{Q}_p^{\mathrm{ur}})(p^{1-r}) \cong E^{d_\chi}.$$

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

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

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

$$\epsilon(K/\mathbb{Q}, 1-r) := (\epsilon(r_{V,\chi}))_{\chi \in \widehat{G}} = (\epsilon(r_{\bar{\chi}})p^{(r-1)c(r_{\bar{\chi}})})_{\chi \in \widehat{G}} \tag{9}$$

as an element of $K_1(\overline{\mathbb{Q}}_p[G])$ 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 $|G| = [K : \mathbb{Q}_p]$. In particular $K/\mathbb{Q}_p$ is tamely ramified with maximal unramified subfield F. Although our methods probably extend to an arbitrary tamely ramified extension $K/\mathbb{Q}_p$ (i.e., where p is allowed to divide $[F : \mathbb{Q}_p]$) this would add an extra layer of notational complexity which we have preferred to avoid. The group $G = \mathrm{Gal}(K/\mathbb{Q}_p)$ is an extension of two cyclic groups

$$\Sigma := \mathrm{Gal}(F/\mathbb{Q}_p) \cong \mathbb{Z}/f\mathbb{Z},$$
$$\Delta := \mathrm{Gal}(K/F) \cong \mathbb{Z}/e\mathbb{Z},$$

where the action of $\sigma \in \Sigma$ on Δ is given by $\delta \mapsto \delta^p$ and we have $e \mid p^f - 1$. By Kummer theory $K = F(\sqrt[e]{p_0})$, 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 $F'(\sqrt[e]{p_0'})$, where $F' := F(\sqrt[e]{\lambda'})$ 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

$$K = F(\sqrt[e]{p_0}), \qquad p_0 \in \mu_{p-1} \cdot p \subseteq \mathbb{Q}_p.$$

We then have

$$G = \mathrm{Gal}(K/\mathbb{Q}_p) \cong \Sigma \ltimes \Delta$$

since $\mathrm{Gal}(K/\mathbb{Q}_p(\sqrt[e]{p_0}))$ is a complement of Δ. If $(e, p-1) = 1$, then the fields $K = F(\sqrt[e]{p_0})$ 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 $F(\sqrt[e]{p})$.

The choice of p_0 (in fact just the valuation of p_0) determines a character

$$\eta_0 : \Delta \xrightarrow{\sim} \mu_e \subset F^\times \subset \mathbb{Q}_p^{\mathrm{ur},\times} \subset \overline{\mathbb{Q}}_p^\times \tag{10}$$

by the usual formula $\delta(\sqrt[e]{p_0}) = \eta_0(\delta) \cdot \sqrt[e]{p_0}$. Let

$$\eta : \Delta \to F^\times$$

be any character of Δ and

$$\Sigma_\eta := \{g \in \Sigma \mid \eta(g\delta g^{-1}) = \eta(\delta) \quad \text{for all } \delta \in \Delta\}$$

the stabilizer of η. Then for any character $\eta' : \Sigma_\eta \to \mathbb{Q}_p^{\mathrm{ur},\times}$ we obtain a character

$$\eta'\eta : G_\eta := \Sigma_\eta \ltimes \Delta \to \mathbb{Q}_p^{\mathrm{ur},\times}$$

and an induced character

$$\chi := \mathrm{Ind}_{G_\eta}^G(\eta'\eta)$$

of G. By [Lang 2002, Exercise XVIII.7], all irreducible characters of G are obtained by this construction, and in fact each $\chi \in \widehat{G}$ is parametrized by a unique pair $([\eta], \eta')$ where $[\eta]$ denotes the Σ-orbit of η. The degree of χ is given by

$$d_\chi = \chi(1) = f_\eta := [\Sigma : \Sigma_\eta] = [F_\eta : \mathbb{Q}_p], \tag{11}$$

where $F_\eta \subseteq F$ is the fixed field of Σ_η.

We have

$$r_\chi = \operatorname{Ind}_{W_{F_\eta}}^{W_{\mathbb{Q}_p}}(r_{\eta'\eta}),$$

where r_χ and $r_{\eta'\eta}$ are the representations of $W_{\mathbb{Q}_p}$ and W_{F_η}, respectively, defined as in (8). By [Serre 1979, Chapter VI, Corollary to Propoposition 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

$$\begin{aligned} \epsilon(r_\chi) &= \epsilon(r_{\eta'\eta}) \\ &= \begin{cases} 1, & \eta = 1, \\ \epsilon(r_\eta) r_{\eta'}(\mathrm{Frob}_{F_\eta})^{-c(r_\eta)} = \eta(\mathrm{rec}(p))\tau(r_{\eta,\sharp}, \psi_p)\eta'(\sigma^{f_\eta})^{-1}, & \eta \neq 1. \end{cases} \end{aligned} \tag{12}$$

3.1. *Gauss sums.* If k_η denotes the residue field of F_η, we have a canonical character

$$\omega : k_\eta^\times \xleftarrow{\sim} \mu_{p^{f_\eta}-1} \subseteq F_\eta^\times \subseteq K^\times \subseteq \overline{\mathbb{Q}}_p^\times,$$

where the first arrow is reduction mod p. On the other hand we have our character

$$r_{\eta,\sharp} : F_\eta^\times \xrightarrow{\mathrm{rec}} W_{F_\eta}^{\mathrm{ab}} \xrightarrow{\iota} G_{F_\eta}^{\mathrm{ab}} \xrightarrow{\psi} G_\eta^{\mathrm{ab}} \xrightarrow{\eta} \overline{\mathbb{Q}}_p^\times$$

of order dividing e. So there exists a unique $m_\eta \in \mathbb{Z}/e\mathbb{Z}$ such that

$$r_{\eta,\sharp}|_{\mu_{p^{f_\eta}-1}} = \omega^{m_\eta(p^{f_\eta}-1)/e} \tag{13}$$

and formula (7) gives

$$\tau(r_{\eta,\sharp}, \psi_p) = \tau(\omega^{-m_\eta(p^{f_\eta}-1)/e}),$$

where

$$\tau(\omega^{-i}) := \sum_{a \in k_\eta^\times} \omega(a)^{-i} \zeta_p^{\mathrm{Tr}_{k_\eta/\mathbb{F}_p}(a)}$$

is a Gauss sum associated to the finite field k_η. The p-adic valuation of these sums is known:

Lemma 3 [Washington 1997, Proposition 6.13 and Lemma 6.14]. *For $0 \le i \le p^{f_\eta} - 1$, let $i = i_0 + p i_1 + p^2 i_2 + \cdots + i_{f_\eta - 1} p^{f_\eta - 1}$ be the p-adic expansion with digits $0 \le i_j \le p - 1$. Then*

$$v_p(\tau(\omega^{-i})) = \frac{i_0 + i_1 + \cdots + i_{f_\eta - 1}}{p - 1} = \sum_{j=0}^{f_\eta - 1} \left\langle \frac{i p^j}{p^{f_\eta} - 1} \right\rangle,$$

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 $0 \le \langle x \rangle < 1$ is the fractional part of the real number x.

Corollary 4. *For all $\eta \in \hat{\Delta}$ we have*

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

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 ϵ-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*

$$(\epsilon(r_{\bar{\chi}}))_{\chi \in \widehat{G}} \cdot [\mathrm{per}(b)] \in \mathrm{im}(K_1(\mathbb{Z}_p^{\mathrm{ur}}[G]))$$

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

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 σ will temporarily denote a generic element of Σ rather than the Frobenius.

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

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

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

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

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

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

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 $(\sqrt[e]{p_0})^{-(e-1)}\mathcal{O}_K$ fixed by Σ and ξ 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

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

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

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

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

$$\xi_{\eta'} \in \mathbb{Z}_p^{\mathrm{ur},\times}. \tag{14}$$

We now analyze the second factor

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

which is the product over the projections of x into the η^{p^i}-eigenspaces for $i = 0, \dots, f_\eta - 1$ (up to the unit $|\Delta| = e$). For $0 \le j < e$ the η_0^{-j}-eigenspace of the inverse different is generated over $\mathcal{O}_F$ by $(\sqrt[e]{p_0})^{-j}$ and since x was a $\mathcal{O}_F[\Delta]$-basis of the inverse different its projection lies in $\mathcal{O}_F^\times\cdot(\sqrt[e]{p_0})^{-j}$. So by Lemma 6 below we have

$$x_\eta \in \mathcal{O}_F^\times\cdot\prod_{i=0}^{f_\eta-1}(\sqrt[e]{p_0})^{-e\langle p^i(-m_\eta)/e\rangle}\subset K$$

and hence

$$v_p(x_\eta) = -\sum_{i=0}^{f_\eta-1}\left\langle\frac{-m_\eta p^i}{e}\right\rangle = -v_p(\tau(r_{\bar{\eta},\sharp},\psi_p)), \tag{15}$$

using Corollary 4 and the fact that $\bar{\eta} = \eta_0^{-m_\eta}$. One checks that $\tau(r_{\bar{\eta},\sharp},\psi_p)\in\mathbb{Q}_p^{\mathrm{ur}}(\zeta_p)$ is an eigenvector for the character

$$\varrho = \eta_0^{-m_\eta(p^{f_\eta}-1)/(p-1)}$$

of the group $\mathrm{Gal}(\mathbb{Q}_p^{\mathrm{ur}}(\zeta_p) \cap K^{\mathrm{ur}}/\mathbb{Q}_p^{\mathrm{ur}})$. Also, since x_η is an eigenvector for ϱ^{-1}, Equation (15) implies

$$\tau(r_{\bar{\eta},\sharp}, \psi_p) \cdot x_\eta \in \mathbb{Z}_p^{\mathrm{ur},\times}.$$

Combining this with (14) and (12) we find

$$\epsilon(r_{\bar{\chi}}) \cdot [\mathrm{per}(b)]_\chi = \bar{\eta}(\mathrm{rec}(p))\tau(r_{\bar{\eta},\sharp}, \psi_p)\bar{\eta}'(\sigma^{f_\eta}) \cdot x_\eta \cdot \xi_{\eta'} \in \mathbb{Z}_p^{\mathrm{ur},\times}$$

and hence

$$(\epsilon(r_{\bar{\chi}}))_{\chi\in\widehat{G}} \cdot [\mathrm{per}(b)] \in \mathrm{im}(K_1(\mathbb{Z}_p^{\mathrm{ur}}[G])). \qquad \square$$

Lemma 6. *We have $\eta = \eta_0^{m_\eta}$, where η_0 is the character (10) associated to the element p_0 of valuation 1 and m_η was defined in (13).*

Proof. It suffices to show that the composite map

$$\omega' : \mu_{p^{f_\eta}-1} \subset F^\times \xrightarrow{\mathrm{rec}} G_F^{\mathrm{ab}} \to \mathrm{Gal}(K/F) \xrightarrow{\eta_0^{m_\eta}} \mu_e$$

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 rec is the inverse of that used in [Neukirch 1999], we have

$$\omega'(\zeta) = \left(\frac{\zeta^{-1}, p_0^{m_\eta}}{F}\right),$$

which by [Neukirch 1999, Theorem V.3.4] equals

$$\left(\frac{\zeta^{-1}, p_0^{m_\eta}}{F}\right) = \left((-1)^{\alpha\beta}\frac{p_0^\beta}{\zeta^{-\alpha}}\right)^{(p^{f_\eta}-1)/e} = \zeta^{m_\eta(p^{f_\eta}-1)/e},$$

where $\alpha = v_p(p_0^{m_\eta}) = m_\eta$ and $\beta = v_p(\zeta^{-1}) = 0$. $\square$

Denote by γ a topological generator of

$$\Gamma := \mathrm{Gal}(\mathbb{Q}_p(\zeta_{p^\infty})/\mathbb{Q}_p)$$

and by

$$\chi^{\mathrm{cyclo}} : \mathrm{Gal}(\mathbb{Q}_p(\zeta_{p^\infty})/\mathbb{Q}_p) \cong \mathbb{Z}_p^\times$$

the cyclotomic character. As in the proof of Proposition 5 choose b such that

$$\mathbb{Z}_p[G] \cdot b = (\sqrt[e]{p_0})^{-(e-1)}\mathcal{O}_K.$$

Denote by $e_1 = \dfrac{1}{|\Sigma|}\sum_{g\in\Sigma} g \in \mathbb{Z}_p[\Sigma]$ the idempotent for the trivial character of Σ.

Proposition 7. *If $p \nmid |G|$ then one can choose $\beta \in H^1(K, \mathbb{Z}_p(1-r))$ such that*

$$H^1(K, \mathbb{Z}_p(1-r)) = H^1(K, \mathbb{Z}_p(1-r))_{\mathrm{tor}} \oplus \mathbb{Z}_p[G] \cdot \beta$$

and the local Tamagawa number conjecture (4) is equivalent to the identity

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

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

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

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

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

this identity is equivalent to

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

Proof. If $p\nmid|G|$ then the module $H^1(K,\mathbb{Z}_p(1-r))/\mathrm{tor}$ 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

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

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

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

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

$$\begin{aligned}[C_\beta]_F&=[H^1(F,\mathbb{Z}_p(1-r))_{\mathrm{tor}}]^{-1}\cdot[H^2(F,\mathbb{Z}_p(1-r))]\\&=[H^0(F,\mathbb{Q}_p/\mathbb{Z}_p(1-r))]^{-1}\cdot[H^0(F,\mathbb{Q}_p/\mathbb{Z}_p(r))]\\&=[(\chi^{\mathrm{cyclo}}(\gamma)^{r-1}-1)e_1+1-e_1]\cdot[(\chi^{\mathrm{cyclo}}(\gamma)^r-1)e_1+1-e_1]^{-1}\\&=\left[\frac{\chi^{\mathrm{cyclo}}(\gamma)^{r-1}-1}{\chi^{\mathrm{cyclo}}(\gamma)^r-1}e_1+1-e_1\right].\end{aligned}$$

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

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 (φ, Γ_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

$$K_n = K(\zeta_{p^n}), \quad K_\infty = \bigcup_{n \in \mathbb{N}} K_n,$$
$$\Gamma_K := \mathrm{Gal}(K_\infty/K), \quad \Lambda_K = \mathbb{Z}_p[\![\mathrm{Gal}(K_\infty/\mathbb{Q}_p)]\!]$$

and

$$H^m_{Iw}(K, \mathbb{Z}_p(1)) = \varprojlim_n H^m(K_n, \mathbb{Z}_p(1)) \cong \varprojlim_n H^m(K, \mathrm{Ind}_{G_{K_n}}^{G_K} \mathbb{Z}_p(1)) \cong H^m(K, T),$$

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

$$T := \varprojlim_n \mathrm{Ind}_{G_{K_n}}^{G_K} \mathbb{Z}_p(1) \cong \varprojlim_n \mathbb{Z}_p[\mathrm{Gal}(K_n/K)](1) \cong \mathbb{Z}_p[\![\Gamma_K]\!](1)$$

is a free rank-one $\mathbb{Z}_p[\![\Gamma_K]\!]$-module with G_K-action given by $\psi^{-1}\chi^{\mathrm{cyclo}}$, 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

$$0 \to T \xrightarrow{\gamma_K \cdot \chi^{\mathrm{cyclo}}(\gamma_K)^{r-1} - 1} T \longrightarrow \mathbb{Z}_p(r) \to 0 \tag{17}$$

where $\gamma_K \in \Gamma_K$ is a topological generator (our assumption that p is odd assures that Γ_K is procyclic for any K). It is clear from the definition that

$$H^m_{Iw}(K, \mathbb{Z}_p(1)) \cong H^m_{Iw}(K_n, \mathbb{Z}_p(1)) \tag{18}$$

for any $n \geq 0$. So $H^m_{Iw}(K, \mathbb{Z}_p(1))$ only depends on the field K_∞, and it is naturally a Λ_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 $T \cong \mathrm{Ind}_{G_{K_n}}^{G_K} T_n$. In view of (18) we obtain induced maps

$$\mathrm{pr}_{n,r} : H^1_{Iw}(K, \mathbb{Z}_p(1)) \to H^1(K_n, \mathbb{Z}_p(r)) \tag{19}$$

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 $\mathrm{pr}_{n,r}$ induces an isomorphism*

$$H^1_{Iw}(K, \mathbb{Z}_p(1))/(\gamma_n - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r})H^1_{Iw}(K, \mathbb{Z}_p(1)) \cong H^1(K_n, \mathbb{Z}_p(r)).$$

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

$$\begin{tikzcd}
0 \arrow[r] & H^0_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r, "\gamma_n - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r}"] & H^0_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r] & H^0(K_n, \mathbb{Z}_p(r)) \arrow[dll] & \\
& H^1_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r, "\gamma_n - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r}"'] & H^1_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r, "\mathrm{pr}_{n,r}"] & H^1(K_n, \mathbb{Z}_p(r)) \arrow[dll] & \\
& H^2_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r, "\gamma_n - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r}"'] & H^2_{Iw}(K, \mathbb{Z}_p(1)) \arrow[r] & H^2(K_n, \mathbb{Z}_p(r)) \arrow[r] & 0.
\end{tikzcd}$$

By Tate local duality there is a canonical isomorphism of $\mathrm{Gal}(K_n/K)$-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 Γ_{K_n}. This implies that for $r \neq 1$ multiplication by

$$\gamma_n - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r} = 1 - \chi^{\mathrm{cyclo}}(\gamma_n)^{1-r}$$

is injective on $H^2_{Iw}(K, \mathbb{Z}_p(1))$. Hence $\mathrm{pr}_{n,r}$ is surjective and we obtain the desired isomorphism. $\square$

4.2. ***The ring A_K and the reciprocity law.*** The theory of (φ, Γ_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 π_K is (for now) a formal variable and $F' \supseteq F$ is the maximal unramified subfield of K_∞. (The notation $(-)^\wedge$ means $\widehat{-}$.) The ring A_K carries an operator φ extending the Frobenius on $\mathcal{O}_{F'}$ and an action of Γ_K commuting with φ, which are somewhat hard to describe in terms of π_K. However, on the subring

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

one has

$$\varphi(1+\pi) = (1+\pi)^p, \quad \gamma(1+\pi) = (1+\pi)^{\chi^{\mathrm{cyclo}}(\gamma)} \tag{20}$$

for $\gamma \in \Gamma_K$.

The ring A_K is a complete, discrete valuation ring with uniformizer p. We denote by $E_K \cong k((\bar{\pi}_K))$ 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

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

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

By [Cherbonnier and Colmez 1999, III.2] the field B_K is contained in a field $\widetilde{B}$ on which φ is bijective and $\widetilde{B}$ contains a G_K-stable subring $\widetilde{B}^{\dagger,n}$ 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

$$\varphi^{-n} : \widetilde{B}^{\dagger,n} \to B_{\mathrm{dR}},$$

which again is rather inexplicit in general but is given by

$$\varphi^{-n}(\pi) = \zeta_{p^n} e^{t/p^n} - 1$$

on the element π.

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*

$$\Lambda_K := \mathbb{Z}_p[[\mathrm{Gal}(K_\infty/\mathbb{Q}_p)]]$$

its Iwasawa algebra.

(a) *There is an isomorphism of Λ_K-modules*

$$\mathrm{Exp}^*_{\mathbb{Z}_p} : H^1_{Iw}(K, \mathbb{Z}_p(1)) \cong A_K^{\psi=1}(1).$$

(b) *There is $n_0 \in \mathbb{Z}$ such that for $n \geq n_0$ the following hold:*

(b1) $A_K^{\psi=1} \subseteq \widetilde{B}^{\dagger,n}$.

(b2) *The G_K-equivariant map $\varphi^{-n} : A_K^{\psi=1} \to B_{\mathrm{dR}}$ factors through*

$$\varphi^{-n} : A_K^{\psi=1} \to K_n[[t]] \subseteq B_{\mathrm{dR}}.$$

(b3) *One has*

$$p^{-n}\varphi^{-n}(\mathrm{Exp}^*_{\mathbb{Z}_p}(u)) = \sum_{r=1}^{\infty} \exp^*_{\mathbb{Q}_p(r)}(\mathrm{pr}_{n,1-r}(u)) \cdot t^{r-1}$$

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 $\mathrm{Exp}^*_{\mathbb{Z}_p}$, 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 π:

$$\nabla(\pi) = 1 + \pi.$$

We set

$$\nabla \log(x) = \frac{\nabla(x)}{x}$$

and denote by

$$\widehat{M} := \varprojlim_n M/p^n M$$

the p-adic completion of an abelian group M.

Proposition 10. *There is a commutative diagram of Λ_K-modules, where the maps labeled by $\cong$ are isomorphism.*

$$\begin{tikzcd}
& H^1_{Iw}(K, \mathbb{Z}_p(1)) \arrow[drr, "\mathrm{Exp}^*_{\mathbb{Z}_p}", "\cong"'] & & \\
A(K_\infty) := \varprojlim_{m,n} K_n^\times/(K_n^\times)^{p^m} \arrow[ur, "\cong", "\delta"'] \arrow[r, "\iota_K", "\cong"'] & (E_K^\times)^\wedge & (A_K^{\mathcal{N}=1})^\wedge \arrow[l, "\cong"', "\mathrm{mod}\ p"] \arrow[r, "\cong", "\nabla\log"'] & A_K^{\psi=1}(1) \\
U := \varprojlim_{m,n} \mathcal{O}_{K_n}^\times/(\mathcal{O}_{K_n}^\times)^{p^m} \arrow[u, hook] \arrow[r, "\iota_K|_U", "\cong"'] & 1 + \bar{\pi}_K k[\![\bar{\pi}_K]\!] \arrow[u, hook] & &
\end{tikzcd}$$

Proof. The isomorphism δ 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]

$$\varprojlim_n \mathcal{O}_{K_n} \xrightarrow{\cong} k[\![\bar{\pi}_K]\!],$$

which induces our isomorphism $\iota_K|_U$ after restricting to units and passing to p-adic completions and our isomorphism ι_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 $(A_K^{\mathcal{N}=1})^\wedge \to (E_K^\times)^\wedge$ 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. □

4.3. ***Specialization to the tamely ramified case.*** We now resume our assumption that p does not divide the degree of $[K : \mathbb{Q}_p]$ together with (most of) the notation from Section 3. In addition we assume that

$$\zeta_p \in K,$$

which implies that K_∞/K is totally ramified and hence that $F = F'$ is the maximal unramified subfield of K_∞. 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

$$\mathrm{Gal}(E_K/E_F) \cong \mathrm{Gal}(K_\infty/F_\infty) \cong \mathrm{Gal}(K/F(\zeta_p)).$$

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

$$K = F(\sqrt[e(p-1)]{-p}).$$

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

$$\zeta_p - 1 = \sqrt[(p-1)]{-p} \cdot u' \tag{21}$$

with $u' \in 1 + (\zeta_p - 1)\mathbb{Z}_p[\zeta_p]$. By Kummer theory we then also have

$$K = F(\sqrt[e]{\zeta_p - 1})$$

and $B_K = B_F(\sqrt[e]{\pi})$. Any choice of $\pi_K = \sqrt[e]{\pi}$ fixes a choice of

$$\sqrt[e]{\zeta_p - 1} = \varphi^{-1}(\pi_K)|_{t=0}$$

and of

$$\sqrt[e(p-1)]{-p} = \sqrt[e]{\zeta_p - 1} \cdot (u')^{-1/e}.$$

We have

$$G \cong \Sigma \ltimes \Delta$$

with Σ cyclic of order f and Δ 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 Γ_K.

Proposition 11. *There is an isomorphism of Λ_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 Λ_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 Δ of [loc. cit.]). □

Corollary 12. *There is an isomorphism of $\mathbb{Z}_p[G]$-modules*

$$H^1(K, \mathbb{Z}_p(1-r)) \cong \mathbb{Z}_p[G] \cdot \beta \oplus H^1(K, \mathbb{Z}_p(1-r))_{\mathrm{tor}},$$

where $\beta = \mathrm{pr}_{0,1-r}(\beta_{Iw}) = \mathrm{pr}_{1,1-r}(\beta_{Iw})$.

Proof. This is clear from Proposition 11 and Lemma 8 (with r replaced by $1-r$) in view of the isomorphisms

$$\mathbb{Z}_p[G] \xrightarrow{\sim} \Lambda_K/(\gamma_1 - \chi^{\mathrm{cyclo}}(\gamma_1)^r)\Lambda_K$$

and

$$\begin{aligned}\mathbb{Z}_p(1)/(\gamma_1 - \chi^{\mathrm{cyclo}}(\gamma_1)^r)\mathbb{Z}_p(1) &= \mathbb{Z}_p/(\chi^{\mathrm{cyclo}}(\gamma_1) - \chi^{\mathrm{cyclo}}(\gamma_1)^r)\mathbb{Z}_p \\ &\cong H^0(K, \mathbb{Q}_p/\mathbb{Z}_p(1-r))) \\ &\cong H^1(K, \mathbb{Z}_p(1-r))_{\mathrm{tor}}. \end{aligned}$$ □

If we choose the element β of Corollary 12 to verify the identity in Proposition 7 it remains to get an explicit hold on some Λ_K-basis β_{Iw}, or rather of its image

$$\alpha = \mathrm{Exp}^*_{\mathbb{Z}_p}(\beta_{Iw}) \in A_K^{\psi=1}(1). \tag{22}$$

Since α is a (infinite) Laurent series in π_K it will be amenable to somewhat explicit analysis. In the unramified components of Proposition 7 ($\eta = 1$) we can compute α 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 α as much as we can in Section 6.

In order to compute $\exp^*_{\mathbb{Q}_p(r)}(\beta)$ we also need to be able to apply Theorem 9 for $n = 1$.

Proposition 13. *Part* (*b*) *of Theorem 9 holds with* $n_0 = 1$.

Proof. It will follow from an explicit analysis of elements in $A_K^{\psi=1}$ in Corollary 37 below that $\varphi^{-1}(a)$ converges for $a \in A_K^{\psi=1}$, which shows (b1). Since $\pi_K^e = \pi$ and $\varphi^{-n}(\pi) = \zeta_{p^n} e^{t/p^n} - 1$ it is also clear that the values of φ^{-n} on A_K, if convergent, lie in $F(\sqrt[e]{\zeta_{p^n} - 1})[\![t]\!] = K_n[\![t]\!]$. This shows (b2). By [Cherbonnier and Colmez 1999, théorème IV.2.1] the right-hand side of (b3) is given by $T_n\varphi^{-m}(\mathrm{Exp}^*_{\mathbb{Z}_p}(u))$ 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. □

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 ∇ applied to the left-hand side. First we have

Lemma 14 [Cherbonnier and Colmez 1999, lemme III.2.3]. *Suppose $\varphi^{-n}f$ and $\varphi^{-n}(\nabla f)$ both converge in B_{dR}. 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 $p^n\frac{d}{dt}\circ\varphi^{-n}$ both agree on $1+\pi$, since they are both derivations. We see that

$$\varphi^{-n}\nabla(1+\pi)=\varphi^{-n}(1+\pi)=\zeta_{p^n}e^{t/p^n}$$
$$p^n\frac{d}{dt}\varphi^{-n}(1+\pi)=p^n\frac{d}{dt}\zeta_{p^n}e^{t/p^n}=\zeta_{p^n}e^{t/p^n}.$$ □

The next Lemma shows that ∇ 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) $\nabla\gamma f=\chi^{\mathrm{cyclo}}(\gamma)\cdot\gamma\nabla f$,

(b) $\nabla\varphi f=p\cdot\varphi\nabla f$,

(c) $\nabla\operatorname{Tr}_{B/\varphi B}f=\operatorname{Tr}_{B/\varphi B}\nabla f$,

(d) $\nabla\psi f=p^{-1}\cdot\psi\nabla f$.

Proof. This is a straightforward computation. For example, to see (c) note that $(1+\pi)^i$, $i=0,\dots,p-1$ is a φB-basis of B and

$$\operatorname{Tr}_{B/\varphi B}(x)=\operatorname{Tr}_{B/\varphi B}\left(\sum_{i=0}^{p-1}\varphi x_i\cdot(1+\pi)^i\right)=p\cdot\varphi x_0.$$

Hence

$$\begin{aligned}\operatorname{Tr}_{B/\varphi B}(\nabla x)&=\operatorname{Tr}_{B/\varphi B}\left(\sum_{i=0}^{p-1}\nabla\varphi x_i\cdot(1+\pi)^i+\varphi x_i\cdot i\cdot(1+\pi)^i\right)\\&=\operatorname{Tr}_{B/\varphi B}\left(\sum_{i=0}^{p-1}\varphi(p\nabla x_i+x_i\cdot i)\cdot(1+\pi)^i\right)\\&=p^2\varphi\nabla x_0=\nabla(p\cdot\varphi x_0)=\nabla\operatorname{Tr}_{B/\varphi B}(x).\end{aligned}$$

See [Daigle 2014, Lemma 3.1.3] for more details. □

Recall the normalized trace maps

$$T_n:K_\infty\to K_n$$

from [Cherbonnier and Colmez 1999, p. 259] which are given by

$$T_n(x)=p^{-m}\operatorname{Tr}_{K_m/K_n}x$$

for any $m\geq n$ such that $x\in K_m$, and extend to a map

$$T_n:K_\infty[\![t]\!]\to K_n[\![t]\!]$$

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 $f=\operatorname{Exp}^*_{\mathbb{Z}_p}(u)\in A_K^{\psi=1}$ 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 $P \in A_K^{\psi=p^{r-1}}$ 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

$$(T_0\varphi^{-m}P)(0) = p^{(r-1)m}(1 - p^{-r}\sigma^{-1})(\varphi^{-0}P)(0). \tag{24}$$

Proof. Since $P \in A_K^{\psi=p^{r-1}}$, we know that $\psi(P) = p^{r-1}P$ and thus that

$$p^{-r}\,\mathrm{Tr}_{B/\varphi B}(P) = \varphi(P).$$

Recall that we can choose π_K such that $\pi_K^e = \pi$. Then $\{((1+\pi)\zeta - 1)^{1/e} : \zeta \in \mu_p\}$ is the set of conjugates of π_K over $\varphi(B)$ in an algebraic closure of B, so this gives

$$p^{-r}\sum_{\zeta\in\mu_p} P\big(((1+\pi)\zeta - 1)^{1/e}\big) = P^{\sigma}\big(((1+\pi)^p - 1)^{1/e}\big).$$

Whenever $\varphi^{-(l+1)}P$ converges for some $l \in \mathbb{N}$, the operator $\varphi^{-(l+1)}P|_{t=0}$ corresponds to setting $\pi = \zeta_{p^{l+1}} - 1$ and applying $\sigma^{-(l+1)}$ to each coefficient. We get

$$p^{-r}\sum_{\zeta\in\mu_p} P^{\sigma^{-(l+1)}}((\zeta\cdot\zeta_{p^{l+1}} - 1)^{1/e}) = P^{\sigma^{-l}}((\zeta_{p^l} - 1)^{1/e}). \tag{25}$$

If $l \geq 1$, this simplifies to

$$p^{-r}\,\mathrm{Tr}_{K_{l+1}/K_l}\,P^{\sigma^{-(l+1)}}((\zeta_{p^{l+1}} - 1)^{1/e}) = P^{\sigma^{-l}}((\zeta_{p^l} - 1)^{1/e}),$$

and by induction, we see that for any $1 \leq n < m$,

$$p^{m-r(m-n)}T_n P^{\sigma^{-m}}((\zeta_{p^m} - 1)^{1/e}) = P^{\sigma^{-n}}((\zeta_{p^n} - 1)^{1/e}). \tag{26}$$

Since $P^{\sigma^{-m}}((\zeta_{p^m}-1)^{1/e}) = (\varphi^{-m}P)(0)$, this proves Equation (23). If $l=0$ then Equation (25) becomes

$$p^{-r}\sum_{\zeta\in\mu_p} P^{\sigma^{-1}}((\zeta\cdot\zeta_p-1)^{1/e}) = (\varphi^{-0}P)(0).$$

The left-hand side is now equal to

$$p^{-r}P^{\sigma^{-1}}(0) + p^{-r}\operatorname{Tr}_{K_1/K_0} P^{\sigma^{-1}}((\zeta_p-1)^{1/e})$$

and we have

$$p^{-r}\operatorname{Tr}_{K_1/K_0}(P^{\sigma^{-1}}((\zeta_p-1)^{1/e})) = (1-p^{-r}\sigma^{-1})(\varphi^{-0}P)(0).$$

By induction we get

$$p^{m-rm}T_0P^{\sigma^{-m}}((\zeta_{p^m}-1)^{1/e}) = (1-p^{-r}\sigma^{-1})(\varphi^{-0}P)(0),$$

which proves Equation (24). □

Corollary 17. *If $P\in A_K^{\psi=1}$ 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 Λ_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

$$\begin{aligned}
\mathcal{P}_F &:= \Big\{\sum_{n\geq 0} a_n\pi^n \in F[\![\pi]\!] : na_n \in \mathcal{O}_F\Big\},\\
\overline{\mathcal{P}}_F &:= \mathcal{P}_F/p\mathcal{O}_F[\![\pi]\!],\\
\overline{\mathcal{P}}_{F,\log} &:= \{f \in \overline{\mathcal{P}}_F : (p-\varphi)(f) = 0\},\\
\mathcal{P}_{F,\log} &:= \{f \in \mathcal{P}_F : \bar{f} \in \overline{\mathcal{P}}_{F,\log}\} = \{f \in \mathcal{P}_F : (p-\varphi)(f) \in p\mathcal{O}_F[\![\pi]\!]\},\\
\mathcal{O}_F[\![\pi]\!]_{\log} &:= \{f \in \mathcal{O}_F[\![\pi]\!]^\times : f \mod p\mathcal{O}_F[\![\pi]\!] \in 1+\pi k[\![\pi]\!]\}\\
&= 1+(\pi, p).
\end{aligned}$$

Note that $\mathcal{P}_F$ is the space of power series in F whose derivative with respect to π lies in $\mathcal{O}_F[\![\pi]\!]$. Observe that the map $d\log$ is given by an integral power series, and therefore $\log \mathcal{O}_F[\![\pi]\!]_{\log} \subseteq \mathcal{P}_F$ 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 φ reduces modulo p to the Frobenius, i.e., to the p-th power map, the logarithm series in fact induces a map

$$\log : \mathcal{O}_F[\![\pi]\!]_{\log} \to \mathcal{P}_{F,\log}.$$

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*

$$f \in 1+\pi k[\![\pi]\!] = \widehat{\mathbb{G}}_m(k[\![\pi]\!])$$

and let $\hat{f}$ be any lift of f to $\mathcal{O}_F[\![\pi]\!]_{\log}$. Then

$$\log(\hat{f}) \bmod p\mathcal{O}_F[\![\pi]\!] \in \overline{\mathcal{P}}_{F,\log}$$

does not depend on the choice of $\hat{f}$, and the map $f \mapsto \log(\hat{f}) \bmod p\mathcal{O}_F[\![\pi]\!]$ is an isomorphism $\log_k : 1+\pi k[\![\pi]\!] \xrightarrow{\sim} \overline{\mathcal{P}}_{F,\log}$.

Lemma 19 [Perrin-Riou 1990, lemme 2.2]. *Let $f \in \mathcal{P}_{F,\log}$. Then the sequence $p^m\psi^m(f)$ converges to a limit $f^\infty \in \mathcal{P}_{F,\log}$, and we have*

(1) $f^\infty \equiv f \mod p\mathcal{O}_F[\![\pi]\!]$,

(2) $\psi(f^\infty) = p^{-1}f^\infty$,

(3) $(1-p^{-1}\varphi)f^\infty \in \mathcal{O}_F[\![\pi]\!]$,

(4) $f^\infty = 0 \quad$ *if* $f \in \mathcal{O}_F[\![\pi]\!]$,

(5) $f^\infty = g^\infty \quad$ *if* $f \equiv g \bmod p\mathcal{O}_F[\![\pi]\!]$.

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*

$$\begin{tikzcd}
\mathcal{O}_F[\![\pi]\!]_{\log}^{\mathcal{N}=1} \arrow[r, "\log", "\cong"'] \arrow[d, "\mathrm{mod}\ p", "\cong"'] & \mathcal{P}_{F,\log}^{\psi=p^{-1}} \arrow[d, "\mathrm{mod}\ p", "\cong"'] \\
1+\pi k[\![\pi]\!] \arrow[r, "\log_k", "\cong"'] & \overline{\mathcal{P}}_{F,\log}
\end{tikzcd}$$

Proof. To see the first part, note that we have a commutative diagram

$$\begin{tikzcd}
1 \arrow[d] & 0 \arrow[d] \\
1+p\mathcal{O}_F[\![\pi]\!] \arrow[r, "\log", "\cong"'] \arrow[d] & p\mathcal{O}_F[\![\pi]\!] \arrow[d] \\
\mathcal{O}_F[\![\pi]\!]_{\log} \arrow[r, "\log"] \arrow[d] & \mathcal{P}_{F,\log} \arrow[d] \\
1+\pi k[\![\pi]\!] \arrow[r, "\log_k", "\cong"'] \arrow[d] & \overline{\mathcal{P}}_{F,\log} \arrow[d] \\
1 & 0
\end{tikzcd}$$

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 $\mathcal{P}_{F,\log}^{\psi=p^{-1}}$ 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* Λ_F*-modules*:

$$\begin{tikzcd}
A(F_\infty) = \varprojlim_{m,n} F_n^\times/(F_n^\times)^{p^m} & (E_F^\times)^\wedge \arrow[l, "\cong"'] & (A_F^{\mathcal{N}=1})^\wedge \arrow[l, "\cong"', "\mathrm{mod}\ p"] \arrow[r, "\cong", "\nabla\log"'] & A_F^{\psi=1}(1) \\
U = \varprojlim_{m,n} \mathcal{O}_{F_n}^\times/(\mathcal{O}_{F_n}^\times)^{p^m} \arrow[u, hook] & 1+\pi k[\![\pi]\!] \arrow[l, "\cong"'] \arrow[u, hook] & \mathcal{O}_F[\![\pi]\!]_{\log}^{\mathcal{N}=1} \arrow[l, "\cong"', "\mathrm{mod}\ p"] \arrow[u, hook] \arrow[r, "\cong", "\log"'] & \mathcal{P}_{F,\log}^{\psi=p^{-1}} \arrow[u, hook, "\nabla"]
\end{tikzcd}$$

Proof. This is immediate from Corollary 20(2). □

This diagram allows us to determine the exact relationship between $\mathcal{P}_{F,\log}^{\psi=p^{-1}}$ and $A_F^{\psi=1}(1)$ since the relationship between $A(F_\infty)$ and U is quite transparent. There is an exact sequence of Λ_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

$$A(F_\infty) \cong \Lambda_F \oplus \mathbb{Z}_p(1) \tag{27}$$

and the torsion submodule $\mathbb{Z}_p(1)$ is clearly contained in U. Hence we obtain an exact sequence

$$0 \to U_{\mathrm{tf}} \to A(F_\infty)_{\mathrm{tf}} \xrightarrow{v} \mathbb{Z}_p \to 0,$$

where $M_{\mathrm{tf}} := M/M_{\mathrm{tors}}$. The module $A(F_\infty)_{\mathrm{tf}}$ is free of rank one and since the $(\Sigma \times \Gamma)$-action on $\mathbb{Z}_p$ is trivial we find

$$U_{\mathrm{tf}} = I \cdot A(F_\infty)_{\mathrm{tf}},$$

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 Σ.

Proof. This hinges on our assumption that p does not divide the order of Σ, which implies that e_1 has coefficients in $\mathbb{Z}_p$. Using $e_1^2 = e_1$ we then find immediately

$$\begin{aligned}\sigma - 1 &= (\sigma - 1)(1 - e_1) = (\sigma - 1)(1 - e_1) \cdot [(1 - e_1) + (\gamma - 1)e_1],\\ \gamma - 1 &= ((\gamma - 1)(1 - e_1) + e_1) \cdot [(1 - e_1) + (\gamma - 1)e_1].\end{aligned}$$ □

Lemma 23. *There are elements $\alpha \in A_F^{\psi=1}(1)$, $\tilde{\alpha} \in \mathcal{P}_{F,\log}^{\psi=p^{-1}}$ such that*

(1) $A_F^{\psi=1}(1) = \Lambda_F \cdot \alpha \oplus \mathbb{Z}_p(1) \cdot 1$,

(2) $\mathcal{P}_{F,\log}^{\psi=p^{-1}} = \Lambda_F \cdot \tilde{\alpha} \oplus \mathbb{Z}_p \cdot \log(1 + \pi)$,

(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 $\mathcal{P}_{F,\log}^{\psi=p^{-1}}$ and that $(\mathcal{P}_{F,\log}^{\psi=p^{-1}})_{\mathrm{tf}}$ is free of rank one over Λ_F, since it is isomorphic under ∇ 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 α here as an element of $A_F(1)$, i.e., the action of γ is $\chi^{\mathrm{cyclo}}(\gamma)$ times the standard action (20) of γ on A_F. Setting

$$\tilde{\alpha} := \nabla^{-1}((1-e_1)+(\gamma-1)e_1) \cdot \alpha$$

we obtain (3). □

5.2. ***The Coleman exact sequence and the Perrin-Riou basis.*** Lemma 23 tells us that $(\mathcal{P}_{F,\log}^{\psi=p^{-1}})_{\mathrm{tf}}$ is generated over Λ_F by a single element $\tilde{\alpha}$, but not what this $\tilde{\alpha}$ is. By studying one more space, $\mathcal{O}_F[\![\pi]\!]^{\psi=0}$, we are able to describe $\tilde{\alpha}$ and hence α.

Proposition 24. (1) *There is an exact sequence of Λ_F-modules*

$$0 \to \mathbb{Z}_p \cdot \log(1+\pi) \to \mathcal{P}_{F,\log}^{\psi=p^{-1}} \xrightarrow{1-\varphi/p} \mathcal{O}_F[\![\pi]\!]^{\psi=0} \to \mathbb{Z}_p(1) \to 0. \tag{28}$$

(2) $\mathcal{O}_F[\![\pi]\!]^{\psi=0}$ *is a free Λ_F-module of rank one generated by $\xi(1+\pi)$, where $\xi \in \mathcal{O}_F$ is a basis of $\mathcal{O}_F$ 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]. □

Corollary 25. *The bases α and $\tilde{\alpha}$ in Lemma 23 can be chosen such that*

$$(1-\varphi/p) \cdot \tilde{\alpha} = \big((1-e_1)+(\gamma-\chi^{\mathrm{cyclo}}(\gamma))e_1\big) \cdot \xi(1+\pi). \tag{29}$$

Proof. The cokernel of $(1-\varphi/p)$ in (28) is isomorphic to

$$\mathbb{Z}_p(1) \cong \Lambda_F/(\sigma-1, \gamma-\chi^{\mathrm{cyclo}}(\gamma))$$

so the image of $(1-\varphi/p)$ must be $(\sigma-1, \gamma-\chi^{\mathrm{cyclo}}(\gamma)) \cdot \xi(1+\pi)$. As in Lemma 22 we can show that this ideal is principal, and is generated by

$$(1-e_1)+(\gamma-\chi^{\mathrm{cyclo}}(\gamma))e_1.$$

□

5.3. ***Proof of the conjecture for unramified fields.*** We now have the tools we need to explicitly compute $\exp^*_{\mathbb{Q}_p(r)}(H^1(F, \mathbb{Z}_p(1-r)))$ and prove the equality of Proposition 7 for $K = F$ (i.e., $e = 1$). By Lemma 8 we can take

$$\beta := \mathrm{pr}_{0,1-r}(\beta_{Iw}),$$

where β_{Iw} satisfies

$$\begin{aligned}\alpha &= \operatorname{Exp}^*_{\mathbb{Z}_p}(\beta_{Iw}),\\ \nabla\tilde{\alpha} &= ((1-e_1)+(\gamma-1)e_1)\cdot\alpha,\\ (1-\varphi/p)\cdot\tilde{\alpha} &= \big((1-e_1)+(\gamma-\chi^{\mathrm{cyclo}}(\gamma))e_1\big)\cdot\xi(1+\pi),\end{aligned}\tag{30}$$

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

$$\sum_{r=1}^{\infty}\exp^*_{\mathbb{Q}_p(r)}(\mathrm{pr}_{0,1-r}(u))\cdot t^{r-1} = T_0\varphi^{-m}\operatorname{Exp}^*_{\mathbb{Z}_p}(u).$$

Applying this to

$$u = ((1-e_1)+(\gamma-1)e_1)\cdot\beta_{Iw}\tag{31}$$

assures that

$$\operatorname{Exp}^*_{\mathbb{Z}_p}(u) = \nabla\tilde{\alpha}\in\mathcal{O}_F[\![\pi]\!]$$

and therefore

$$\varphi^{-0}P := \varphi^{-0}\nabla^{r-1}\operatorname{Exp}^*_{\mathbb{Z}_p}(u) = \varphi^{-0}\nabla^r\tilde{\alpha}$$

converges in B_{dR} for any $r\ge 1$. Lemma 16 then implies

$$\begin{aligned}\exp^*_{\mathbb{Q}_p(r)}(\mathrm{pr}_{0,1-r}(u)) &= \frac{1}{(r-1)!}\Big(\frac{d}{dt}\Big)^{r-1}T_0\varphi^{-m}\operatorname{Exp}^*_{\mathbb{Z}_p}(u)\Big|_{t=0}\\ &= \frac{1}{(r-1)!}T_0p^{-(r-1)m}\varphi^{-m}\nabla^{r-1}\operatorname{Exp}^*_{\mathbb{Z}_p}(u)\Big|_{t=0}\\ &= \frac{1}{(r-1)!}(1-p^{-r}\sigma^{-1})\varphi^{-0}\nabla^r\tilde{\alpha}\Big|_{t=0}\\ &= \frac{1}{(r-1)!}(1-p^{-r}\sigma^{-1})\nabla^r\tilde{\alpha}\Big|_{\pi=0}.\end{aligned}$$

Applying ∇^r to (29) and using Lemma 15 we have

$$\begin{aligned}(1-p^{r-1}\varphi)\cdot\nabla^r\tilde{\alpha} &= \big((1-e_1)+(\chi^{\mathrm{cyclo}}(\gamma)^r\gamma-\chi^{\mathrm{cyclo}}(\gamma))e_1\big)\cdot\nabla^r\xi(1+\pi)\\ &= \big((1-e_1)+(\chi^{\mathrm{cyclo}}(\gamma)^r\gamma-\chi^{\mathrm{cyclo}}(\gamma))e_1\big)\cdot\xi(1+\pi)\end{aligned}$$

and so we find

$$\begin{aligned}&\exp^*_{\mathbb{Q}_p(r)}(\mathrm{pr}_{0,1-r}(u))\\ &\qquad= \frac{1}{(r-1)!}\cdot\frac{1-p^{-r}\sigma^{-1}}{1-p^{r-1}\sigma}\cdot\big((1-e_1)+(\chi^{\mathrm{cyclo}}(\gamma)^r-\chi^{\mathrm{cyclo}}(\gamma))e_1\big)\cdot\xi.\end{aligned}$$

By Lemma 8 the action of $\gamma \in \Lambda_F$ on $H^1(F, \mathbb{Z}_p(1-r))$ is via the character $\chi^{\mathrm{cyclo}}(\gamma)^r$. Hence, for our choice (31) of u, we have

$$\begin{aligned}\mathrm{pr}_{0,1-r}(u) &= ((1-e_1) + (\chi^{\mathrm{cyclo}}(\gamma)^r - 1)e_1) \cdot \mathrm{pr}_{0,1-r}(\beta_{Iw}) \\ &= ((1-e_1) + (\chi^{\mathrm{cyclo}}(\gamma)^r - 1)e_1) \cdot \beta\end{aligned}$$

and we can finally compute

$$\exp^*_{\mathbb{Q}_p(r)}(\beta) = \frac{1}{(r-1)!} \cdot \frac{1 - p^{-r}\sigma^{-1}}{1 - p^{r-1}\sigma} \cdot \frac{(1-e_1) + (\chi^{\mathrm{cyclo}}(\gamma)^r - \chi^{\mathrm{cyclo}}(\gamma))e_1}{(1-e_1) + (\chi^{\mathrm{cyclo}}(\gamma)^r - 1)e_1} \cdot \xi.$$

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 $\eta|_{\mathrm{Gal}(K/F(\zeta_p))} = 1$ this can easily be done (for any r) using computations similar to those in Section 5.3 with

$$\beta_1 := \mathrm{pr}_{1,1-r}(\beta_{Iw})$$

and β_{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

$$\eta|_{\mathrm{Gal}(K/F(\zeta_p))} \neq 1.$$

In this case we need to verify Equation (16). The main problem is that we do not have any closed formula for a Λ_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

$$\begin{aligned} b_{m,n} &:= p^{-1} \sum_{\zeta \in \mu_p} \zeta^m (1-\zeta^{-1})^n \\ &= p^{-1} \operatorname{Tr}_{\mathbb{Q}(\zeta_p)/\mathbb{Q}} \zeta_p^m (1-\zeta_p^{-1})^n, \quad \text{if } n \geq 1. \end{aligned}$$

Clearly $b_{m,n}$ only depends on $m \pmod p$.

Lemma 27. *One has $b_{m,n} \in \mathbb{Z}$ and*

$$b_{m,n} = \begin{cases} (-1)^{\bar{m}} \binom{n}{\bar{m}}, & 0 \leq n < p, \\ (-1)^{\bar{m}} \binom{n}{\bar{m}} - (-1)^{\bar{m}} \binom{n}{\bar{m}+p}, & p \leq n < 2p, \end{cases} \tag{32}$$

where $0 \leq \bar{m} < p$ is the representative for $m \pmod p$. Moreover,

$$p^{\left\lfloor \frac{n+p-2}{p-1} \right\rfloor - 1} \mid b_{m,n}$$

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

$$\begin{aligned} &\operatorname{Tr}_{\mathbb{Q}(\zeta_p)/\mathbb{Q}}\left(\zeta_p^m (1-\zeta_p^{-1})^n\right) \subseteq p^N \mathbb{Z} \\ &\qquad \iff ((1-\zeta_p)^n) \subseteq \left(p^N (1-\zeta_p)^{2-p}\right) = \left((1-\zeta_p)^{N(p-1)+2-p}\right) \\ &\qquad \iff n \geq N(p-1)+2-p \iff N \leq \frac{n+p-2}{p-1}. \end{aligned}$$ □

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*

$$\sum_{n=0}^{\infty} a_{N+en} \binom{\frac{N}{e}+n}{n} b_{(N/e)+n,n} = \sum_{0 \leq n \leq j \leq n(p-1)} a^{\sigma}_{(N+je)/p} \binom{\frac{N+je}{pe}}{n} \beta_{n,j} \cdot p^n \tag{33}$$

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 $y = \sum_{j=1}^{p-1} \beta_{1,j}\pi^{-j}$ 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

$$\begin{aligned}\varphi(\pi_K^m) = \pi_K^{pm}(1 + p \cdot y)^{m/e} &= \pi_K^{pm} \sum_{n=0}^{\infty} \binom{\frac{m}{e}}{n} y^n \cdot p^n \\ &= \sum_{n=0}^{\infty} \binom{\frac{m}{e}}{n} \sum_{j=n}^{n(p-1)} \beta_{n,j}\pi_K^{pm-ej} \cdot p^n\end{aligned}$$

and the coefficient of π_K^N in $\varphi(a) = \sum_m a_m^\sigma \varphi(\pi_K^m)$ is

$$\sum_{m,n,j,N=pm-ej} a_m^\sigma \binom{\frac{m}{e}}{n} \beta_{n,j} \cdot p^n,$$

which is the right-hand side of (33). The conjugates of π over $\varphi(B)$ are

$$(1+\pi)\zeta - 1 = \pi \cdot \zeta \cdot (1 + (1-\zeta^{-1})\pi^{-1}),$$

hence the conjugates of π_K^m are

$$\pi_K^m \cdot \zeta^{m/e} \cdot \left(1 + (1-\zeta^{-1})\pi^{-1}\right)^{m/e} = \pi_K^m \cdot \zeta^{m/e} \cdot \sum_{n=0}^{\infty} \binom{\frac{m}{e}}{n} (1-\zeta^{-1})^n \pi^{-n}$$

and

$$p^{-1}\operatorname{Tr}_{B/\varphi(B)}(\pi_K^m) = \pi_K^m \cdot \sum_{n=0}^{\infty} \binom{\frac{m}{e}}{n} b_{m/e,n}\pi^{-n} = \sum_{n=0}^{\infty} \binom{\frac{m}{e}}{n} b_{m/e,n}\pi_K^{m-en},$$

and the coefficient of π_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 $\bar{a} \in E_K^{\psi=1}$, 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 $a = b + c \cdot p$. □

Lemma 30. *An element $a=\sum_i a_i\pi_K^i \in E_K$ lies in $E_K^{\psi=1}$ if and only if for all $k\in\mathbb{Z}$ one has*

$$\sum_{n=0}^{p-1} a_{kp+ne}(-1)^n = a_k^\sigma. \tag{34}$$

Proof. The only nonzero term on the right-hand side of (33)$_1$ is $a_{N/p}^\sigma$, corresponding to $n=j=0$, and the nonzero terms on the left-hand side are for $n\le p-1$ by Lemma 27. For $m\in\mathbb{Z}_{(p)}$ one has

$$\binom{m}{n}\binom{n}{\bar m} = \frac{m(m-1)\cdots(m-n+1)}{n!}\cdot\frac{n!}{\bar m!(n-\bar m)!} \equiv \begin{cases}0, & \bar m<n,\\ 1, & \bar m = n,\end{cases}$$

since for $\bar m<n$ one of the factors in $m(m-1)\cdots(m-n+1)$ is divisible by p, whereas for $\bar m = n$ this product is congruent to $\bar m!$ modulo p. For $\bar m>n$ one has $\binom{n}{\bar m}=0$, so $\binom{m}{n}\binom{n}{\bar m}\equiv 0$ whenever $\bar m\ne n$. Using (32) the left-hand side of (33)$_1$ is

$$\sum_{n=0}^{p-1} a_{N+en}\binom{m}{n}\binom{n}{\bar m}(-1)^{\bar m}$$

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*

$$\sum_{n=0}^{p-1} a_{kp+ne}(-1)^n\Big(1+\frac{kp}{e}H_n\Big)+\sum_{n=p+1}^{2(p-1)} a_{kp+ne}(-1)^{n-p}\cdot p\cdot H_{n-p}\Big(1+\frac{k}{e}\Big)\equiv a_k^\sigma. \tag{35}$$

Proof. The only nonzero term on the right-hand side of (33)$_2$ for $N=kp$ is a_k^σ, corresponding to $n=j=0$, since for $n=1$ there is no $1\le j\le (p-1)$ with $p\mid (N+je)=kp+je$. The nonzero terms on the left-hand side are for $n\le 2(p-1)$ by Lemma 27. Note that for $1\le j\le n<2p$ only $j=p$ is divisible by p. So computing modulo p^2 we have

$$\begin{aligned}\binom{\frac{kp}{e}+n}{n} &= \frac{\prod_{j=1}^n\big(\frac{kp}{e}+j\big)}{n!} \equiv \frac{n!+\frac{kp}{e}\sum_{j=1}^n\frac{n!}{j}+\big(\frac{kp}{e}\big)^2\sum_{1\le j_1<j_2\le n}\frac{n!}{j_1j_2}}{n!}\\ &\equiv 1+\frac{kp}{e}H_n+\Big(\frac{kp}{e}\Big)^2\sum_{1\le j_1<j_2\le n}\frac{1}{j_1j_2}\\ &\equiv \begin{cases}1+\frac{kp}{e}H_n, & n<p,\\ 1+\frac{k}{e}+\frac{kp}{e}H_{n-p}+\big(\frac{k}{e}\big)^2\cdot p\cdot H_{n-p}, & p\le n<2p.\end{cases}\end{aligned}$$

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

$$b_{(kp/e)+n,n} = \begin{cases} \binom{n}{n}(-1)^n = (-1)^n, & n < p, \\ 0, & n = p, \\ (-1)^{n-p}\left(\binom{n}{n-p} - \binom{n}{n}\right), & p < n < 2p \end{cases}$$

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

$$\begin{aligned} \binom{\frac{kp}{e}+n}{n} b_{(kp/e)+n,n} &\equiv \left(1 + \frac{k}{e} + \frac{kp}{e} H_{n-p} + \left(\frac{k}{e}\right)^2 \cdot p \cdot H_{n-p}\right)(-1)^{n-p} \cdot p \cdot H_{n-p} \\ &\equiv (-1)^{n-p}\left(1 + \frac{k}{e}\right) \cdot p \cdot H_{n-p}. \end{aligned}$$

□

Lemma 32. *The map $A_K^{\psi=1} \to E_K^{\psi=1}$ is surjective.*

Proof. This follows from the snake lemma applied to

$$\begin{tikzcd} 0 \arrow[r] & A_K \arrow[r, "p"] \arrow[d, "\psi-1"'] & A_K \arrow[r] \arrow[d, "\psi-1"'] & E_K \arrow[r] \arrow[d, "\psi-1"'] & 0 \\ 0 \arrow[r] & A_K \arrow[r, "p"] & A_K \arrow[r] & E_K \arrow[r] & 0 \end{tikzcd}$$

and the fact that $A_K/(\psi - 1)A_K \cong H^2_{Iw}(K, \mathbb{Z}_p(1)) \cong \mathbb{Z}_p$ (see [Cherbonnier and Colmez 1999, remarque II.3.2.]) is p-torsion free. □

Definition 33. For $a = \sum_i a_i \pi_K^i \in A_K$ and $\nu \geq 1$ we set

$$l_\nu(a) := \min\{i \mid p^\nu \nmid a_i\}.$$

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

$$l(a) > -\left(2 - \frac{1}{p}\right)e + e(p-1),$$

while part (c) improves (a) for $\nu = 3$ and a with

$$l(a) > -\left(3 - \frac{2}{p}\right)e + 2e(p-1)$$

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

$$i < -\frac{\nu(p-1)+1}{p} \cdot e \Rightarrow p^\nu \mid a_i, \tag{36}$$

which we denote by $(36)_\nu$ if we want to emphasize dependence on ν. We shall prove $(36)_\nu$ by induction on ν, the statement $(36)_0$ being trivial. Now assume $(36)_{\nu'}$ for $\nu' \leq \nu$ 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 $N = pi$

$$\sum_{n=0}^{\infty} a_{pi+en} \binom{\frac{pi}{e}+n}{n} b_{(pi/e)+n,n} = a_i^\sigma + \sum_{1 \leq n \leq p\lambda \leq n(p-1)} a_{i+\lambda e}^\sigma \binom{\frac{i}{e}+\lambda}{n} \beta_{n,p\lambda} \cdot p^n \tag{37}$$

and first notice that

$$p^{\nu+1-n} \mid a_{i+\lambda e}$$

for $n/p \le \lambda \le n(p-1)/p$. This is because of

$$i + \lambda e < -\frac{(\nu+1)(p-1)+1}{p} \cdot e + \frac{n(p-1)}{p} \cdot e = -\frac{(\nu+1-n)(p-1)+1}{p} \cdot e$$

and the induction assumption. Since $\binom{(i/e)+\lambda}{n}\beta_{n,p\lambda}$ is a p-adic integer we conclude that $p^{\nu+1}$ divides the sum over λ, 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

$$p^j \mid b_{(pi/e)+n,n} \tag{38}$$

for $j(p-1) < n \le (j+1)(p-1)$. For n in this range we have

$$\begin{aligned} pi+ne &\le pi+(j+1)(p-1)e < -\big((\nu+1)(p-1)+1\big)e+(j+1)(p-1)e \\ &= -\big((\nu+1-j)(p-1)+1\big)e+(p-1)e \\ &\le -\frac{(\nu+1-j)(p-1)+1}{p} \cdot e \end{aligned} \tag{39}$$

provided this last inequality holds which is equivalent to

$$\begin{aligned} p\big((\nu+1-j)(p-1)+1\big) - p(p-1) &\ge (\nu+1-j)(p-1)+1 \\ \iff (p-1)\big((\nu+1-j)(p-1)+1\big) &\ge p(p-1) \\ \iff \big((\nu+1-j)(p-1)+1\big) &\ge p \\ \iff (\nu+1-j) \ge 1 &\iff \nu \ge j. \end{aligned}$$

So for $1 \le j \le \nu$ inequality (39) holds, and the induction assumption implies

$$p^{\nu+1-j} \mid a_{pi+ne}.$$

Using (38) we conclude that $p^{\nu+1}$ divides all summands in the left-hand side of (37) except perhaps those with $n < p$ (corresponding to $j = 0$). Since $p^{\nu+1}$ 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 $p^{\nu+1}$ and hence some $i' := pi + en$ with $n \le p - 1$ such that $p^{\nu+1} \nmid a_{i'}$. It remains to remark that

$$i' = pi + en \le pi + e(p-1) < pi - i(p-1) = i \tag{40}$$

since $i < -e$.

To prove (b) we use the same argument. Assuming the existence of

$$i \le \min\{l(a) - e(p-1), -e-1\}$$

with $p^2 \nmid a_i$ we find another $i' < i$ with $p^2 \nmid a_{i'}$. On the right-hand side of (37), apart from a_i^σ, all summands are divisible by p^2 (note there are none with $n = 1$ since λ

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 $p \le n \le 2(p-1)$ we have, assuming $l(a) < -e + e(p-1)$,

$$\begin{aligned} pi + en &\le \big(l(a) - e(p-1)\big) + 2(p-1)e = l(a) + (p-1)l(a) - (p-2)(p-1)e \\ &< l(a) + (p-1)(-e + e(p-1)) - (p-2)(p-1)e = l(a) \end{aligned}$$

and therefore $p \mid a_{pi+en}$. If $l(a) \ge -e + e(p-1)$ we have

$$pi + en < p(-e) + 2(p-1)e = -e + e(p-1) \le l(a)$$

and again conclude $p \mid a_{pi+en}$. So all summands on the left-hand side with $n \ge p$ are divisible by p^2. Hence some $i' := pi + en$ with $n \le p-1$ satisfies $p^2 \nmid a_{i'}$. Moreover, (40) holds since $i < -e$.

For (c) we use this argument yet another time. Assume

$$i \le \min\{l(a) - 2e(p-1), -e-1\}$$

and $p^3 \nmid a_i$. On the right-hand side of (37) we need $p \mid a_{i+\lambda e}$ for $2/p \le \lambda \le 2(p-1)/p$, i.e., $\lambda = 1$. But

$$i + e \le \min\{l(a) - 2e(p-1) + e, -1\} < l(a),$$

so $p \mid a_{i+e}$. Assume first $l(a) < -e + 2e(p-1)$. On the left-hand side we have for $p \le n \le 2(p-1)$

$$\begin{aligned} pi + en &\le p\big(l(a) - 2e(p-1)\big) + 2(p-1)e \\ &= l(a) - e(p-1) + (p-1)l(a) + e(p-1) - (2p-2)(p-1)e \\ &< l(a) - e(p-1) + (p-1)(-e + 2e(p-1)) - (2p-3)(p-1)e \\ &= l(a) - e(p-1) \le l_2(a) \end{aligned}$$

and therefore $p^2 \mid a_{pi+en}$. For $2p-1 \le n \le 3(p-1)$ we just add $(p-1)e$ to this last estimate to conclude

$$\begin{aligned} pi + en \le{} & p\big(l(a) - 2e(p-1)\big) + 3(p-1)e \\ & < l(a) - e(p-1) + e(p-1) = l(a) \end{aligned}$$

and hence $p \mid a_{pi+en}$. Now assume $l(a) \ge -e + 2e(p-1)$. For $p \le n \le 2(p-1)$ we have

$$pi + en \le p(-e) + 2(p-1)e \le l(a) - e(p-1) \le l_2(a)$$

and therefore $p^2 \mid a_{pi+en}$. For $2p-1 \le n \le 3(p-1)$ 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 $n \le p-1$, we have $p^3 \nmid a_{i'}$. Moreover (40) holds since $i < -e$. $\square$

Before drawing consequences of Proposition 34 we make the following definition.

Definition 36. Let ϖ be the uniformizer of K given by

$$\varpi = \sqrt[e]{\zeta_p - 1} = \varphi^{-1}(\pi_K)|_{t=0}$$

and denote by v_ϖ the unnormalized valuation of the field K, i.e.,

$$v_\varpi(p) = e(p-1).$$

For $a \in B_K^{\dagger,1}$ 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., $A_K^{\psi=1} \subseteq B_K^{\dagger,1}$.*

Proof. By (a) we have $p^\nu \mid a_i$ for

$$-\frac{(\nu+1)(p-1)+1}{p} \cdot e \le i < -\frac{\nu(p-1)+1}{p} \cdot e$$

and hence

$$v_p(a_i) \ge \nu \ge -\frac{ip+e}{e(p-1)} - 1$$

and

$$v_\varpi(a_i \varpi^i) \ge -(ip+e) - e(p-1) + i = -(p-1)i - pe. \tag{41}$$

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 $K \subseteq \widehat{\overline{\mathbb{Q}}}_p$. 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 $l(a) \ge -e$. 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 $l(a) \ge -e$ 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 $i = 0, .., p-2$ 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

$$\nabla\log(1+c\pi_K^n) = \frac{\nabla(1+c\pi_K^n)}{1+c\pi_K^n} = \frac{cn/e\cdot(\pi_K^{n-e}+\pi_K^n)}{1+c\pi_K^n} = \frac{cn}{e}\cdot\pi_K^{n-e}+\cdots$$

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.,

$$\nabla\log(\pi^j) = j\cdot\pi^{-1} + j = j\cdot\pi_K^{-e} + j,$$

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 $v_\varpi(a) = l(a)$.

Proof. Since $v_\varpi(a_{l(a)}\varpi^{l(a)}) = l(a)$ we need to show

$$v_\varpi(a_i\varpi^i) > l(a)$$

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 $v_\varpi(a_i\varpi^i) \geq e(p-1) + i > l(a)$. 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 $v_\varpi(a_i\varpi^i) \geq 2e(p-1) + i > l(a)$. Finally for

$$i \leq l(a) - 2e(p-1) < -e - e(p-1) = -ep < -2e$$

we have by (41)

$$v_\varpi(a_i\varpi^i) \geq -(p-1)i - pe > (p-1)2e - pe = (p-2)e > l(a),$$

using the assumption on $l(a)$. $\square$

In order to study $v_\varpi(a)$ 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 $v_\varpi(a)$ cannot only depend on $l(a)$ in this case. In the situation of Proposition 41(b) one can have $v_\varpi(a) = l(a)$ 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

$$l(a+pb) = l(a), \quad v_\varpi(a+pb) \leq l(b) + e(p-1) < l(a) = v_\varpi(a).$$

Proposition 41. *Let $a' \in A_K^{\psi=1}$ with*

$$l(a') = \mu p - e + e(p-1)$$

for some $\mu \in \mathbb{Z}$ with $1 \le \mu < \frac{e(p-1)}{p}$.

(a) *There exists $a \equiv a' \mod p$ with*

$$l_2(a) \ge \mu p - e = l(a) - e(p-1).$$

(b) *For a as in* (a) *we have $v_\varpi(a) \ge l(a)$ with equality if $p \nmid \mu - e$. This last condition is automatic for $e < p$.*

Proof. First note that $l_2(a') \ge -e$ by Proposition 34(b). If $l_2(a') = -e$ then Equation (35) for $k := -e$ reads

$$a'^{\sigma}_{-e} \equiv a'_{kp+e(p-1)} = a'_{-e}$$

since $i = kp + en < l_2(a')$ for $n < p-1$ and $i = kp + en \le -e + e(p-1) < l(a')$ for $p+1 \le n \le 2(p-1)$. 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 $\mu' \ge 1$. Equation (35) for $k := \mu' - e$ then reads

$$0 \equiv a'_{kp+e(p-1)} + \sum_{n=p+1}^{2(p-1)} a'_{kp+ne} \cdot (-1)^{n-p} \cdot p \cdot H_{n-p}\Big(1 + \frac{k}{e}\Big) \tag{42}$$

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 $p+1 \le n \le 2(p-1)$

$$kp + ne < (\mu - e)p + 2(p-1)e = l(a')$$

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 $\mu' \ge \mu$ 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

$$\begin{aligned} 0 &\equiv a_{kp+e(p-1)} + a_{l(a)} \cdot (-1) \cdot p \cdot H_{p-2}\Big(1 + \frac{\mu - e}{e}\Big) \\ &\equiv a_{kp+e(p-1)} - a_{l(a)} \cdot p \cdot \frac{\mu}{e} \pmod{p^2} \end{aligned} \tag{43}$$

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 $v_{\varpi}(a_i\varpi^i) \geq l(a)$ 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 $v_{\varpi}(a_i\varpi^i) > l(a)$. By (43) we have

$$\begin{aligned} a_{l(a)-e(p-1)}\varpi^{l(a)-e(p-1)} + a_{l(a)}\varpi^{l(a)} &\equiv \Big(\frac{p\mu}{\varpi^{e(p-1)}e}+1\Big)a_{l(a)}\varpi^{l(a)} \\ &= \Big(-\frac{\mu}{e}+1\Big)a_{l(a)}\varpi^{l(a)} + O(\varpi^{l(a)+1}) \end{aligned} \tag{44}$$

since

$$\varpi^{e(p-1)} = (\zeta_p-1)^{p-1} \equiv -p \pmod{(\zeta_p-1)^p}.$$

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 $v_{\varpi}(a_i\varpi^i) \geq 2e(p-1)+i > l(a)$. 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 $v_{\varpi}(a_i\varpi^i) \geq 3e(p-1)+i > l(a)$. Finally for

$$i \leq l(a)-3e(p-1) < -e-e(p-1) = -ep$$

we have by (41)

$$v_{\varpi}(a_i\varpi^i) \geq -(p-1)i - pe > (p-1)pe - pe = (p-2)pe \geq (2p-3)e > l(a)$$

using the assumption on $l(a)$. □

6.2. ***Isotypic components.*** We introduce some notation for isotypic components. Recall that

$$G \cong \Sigma \ltimes \Delta$$

with Σ cyclic of order f and Δ cyclic of order $e(p-1)$. For any Σ-orbit $[\eta]$ we define the idempotent

$$e_{[\eta]} = \sum_{\eta' \in \widehat{\Sigma}_\eta} e_\chi \in \mathbb{Z}_p[G],$$

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 a again a $\mathbb{Z}_p[G]$-module. The Σ-orbit

$$[\eta] = \{\eta, \eta^p, \eta^{p^2}, \dots, \eta^{p^{f_\eta - 1}}\} = \{\eta_0^{n_1}, \dots, \eta_0^{n_{f_\eta}}\} \tag{45}$$

corresponds to an orbit $\{n_1, \dots, n_{f_\eta}\} \subseteq \mathbb{Z}/e(p-1)\mathbb{Z}$ of residue classes modulo $e(p-1)$ under the multiplication-by-p map, i.e., we have $n_{i+1} \cong n_i p \mod e(p-1)$ where we view the index i as a class in $\mathbb{Z}/f_\eta\mathbb{Z}$. We shall use the notation

$$[\eta] = \{n_1, \dots, n_{f_\eta}\} = [n_i]$$

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

$$\Delta_e := \mathrm{Gal}(K/F(\zeta_p))$$

acts on $\sqrt[e]{\zeta_p - 1} = \varphi^{-1}(\pi_K)|_{t=0}$ via the character η_0 defined in Section 3 and acts on π_K via η_0^p. The $[\eta] = \{n_1, \dots, n_{f_\eta}\}$-isotypic component of the $\mathbb{Z}_p[\Sigma \ltimes \Delta_e]$-module A_K is

$$\Big\{a = \sum a_n \pi_K^n \;\Big|\; a_n = 0 \quad \text{for } n \bmod e \notin \{n_1, \dots, n_{f_\eta}\}\Big\},$$

but $A_K^{[\eta]}$ is much harder to describe since π_K is not an eigenvector for the full group Δ. 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*

$$p \cdot l_\nu(a) \equiv n \mod e(p-1), \tag{46}$$

then

$$l_\nu(e_\eta a) = l_\nu(a)$$

and the leading coefficients modulo p^ν of $e_\eta a$ and a agree. If $a = e_\eta a$ is an eigenvector for Δ then (46) holds.

Proof. Denote by

$$\omega : \Delta \to \mathrm{Gal}(F(\zeta_p)/F) \to \mathbb{Z}_p^\times$$

the Teichmüller character. For $\delta \in \Delta$ we have

$$\delta(\pi_K) = \big((1+\pi)^{\omega(\delta)} - 1\big)^{1/e} = \left(\sum_{i=1}^{\infty} \binom{\omega(\delta)}{i} \pi^i\right)^{\frac{1}{e}}$$

$$= \lambda(\delta)\pi_K \left(1 + \sum_{i=2}^{\infty} \frac{1}{\omega(\delta)} \binom{\omega(\delta)}{i} \pi^{i-1}\right)^{\frac{1}{e}}$$

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 $\varphi^{-1}|_{t=0}$ we find

$$\delta(\sqrt[e]{\zeta_p - 1}) \equiv \lambda(\delta)^{1/p} \cdot \sqrt[e]{\zeta_p - 1} \mod \varpi^2$$

and since $\sqrt[e]{\zeta_p - 1} \equiv \sqrt[e(p-1)]{-p} \mod \varpi^2$ we obtain $\lambda(\delta) = \eta_0(\delta)^p$. In particular, for any $a \in A_K$

$$\delta(a) \equiv \eta_0(\delta)^{p \cdot l_\nu(a)} \cdot a_{l_\nu(a)} \cdot \pi_K^{l_\nu(a)} \mod (p^\nu, \pi_K^{l_\nu(a)+1})$$

and

$$\begin{aligned} e_\eta a = \frac{1}{e(p-1)} \sum_{\delta \in \Delta} \eta^{-1}(\delta)\delta(a) &\equiv \frac{1}{e(p-1)} \sum_{\delta \in \Delta} \eta_0(\delta)^{p \cdot l_\nu(a) - n} \cdot a_{l_\nu(a)} \cdot \pi_K^{l_\nu(a)} \\ &\equiv \begin{cases} a_{l_\nu(a)} \cdot \pi_K^{l_\nu(a)} & \text{if } p \cdot l_\nu(a) \equiv n \mod e(p-1), \\ 0 & \text{if } p \cdot l_\nu(a) \not\equiv n \mod e(p-1), \end{cases} \end{aligned}$$

where the congruences are modulo $(p^\nu, \pi_K^{l_\nu(a)+1})$. This implies both statements in the lemma. □

Remark 43. With the notation introduced in this section we have

$$e_{[\eta]} = \sum_{i=1}^{f_\eta} e_{\eta p^i}.$$

6.3. ***The main result.*** We view Σ as a subgroup of G such that $\sqrt[e(p-1)]{-p} \in K^\Sigma$, where $\sqrt[e(p-1)]{-p}$ is the choice of root corresponding to our choice of root π_K of π. Then the $\mathbb{Z}_p[\Sigma]$-algebra $\mathbb{Z}_p[G]$ is finite free of rank $e(p-1)$. For each choice of η the $[\eta]$-isotypic component of $\mathbb{Z}_p[G]$ is free of rank f_η 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_η 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 $\{\alpha_i \mid i = 1, \ldots, f_\eta\}$ 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*

$$L_r := \bigoplus_{i=1}^{f_\eta} \mathbb{Z}_p[\Sigma] \cdot (\nabla^{r-1}\alpha_i^{\sigma^{-1}})(\sqrt[e]{\zeta_p - 1})$$

and

$$\mathcal{O}_K^{[n_1-re]} = \bigoplus_{i=1}^{f_\eta} \mathcal{O}_F \cdot (\sqrt[e(p-1)]{-p})^{n_{i,r}}$$

in the $[n_1 - re]$*-isotypic component*

$$K^{[n_1-re]} = \bigoplus_{i=1}^{f_\eta} F \cdot (\sqrt[e(p-1)]{-p})^{n_{i,r}} = \bigoplus_{i=1}^{f_\eta} F \cdot (\sqrt[e(p-1)]{-p})^{n_i-re}$$

of K. *Then the conjunction of* (16) (in Proposition 7) *for* $\chi = ([n_1 - re], \eta')$ *over all* η' *holds if and only if* L_r *and* $\mathcal{O}_K^{[n_1-re]}$ *have the same* $\mathbb{Z}_p[\Sigma]$*-volume, i.e.,*

$$\mathrm{Det}_{\mathbb{Z}_p[\Sigma]} L_r = \mathrm{Det}_{\mathbb{Z}_p[\Sigma]} \mathcal{O}_K^{[n_1-re]} \tag{47}$$

inside $\mathrm{Det}_{\mathbb{Q}_p[\Sigma]} K^{[n_1-re]}$.

Proof. Let α be a $\Lambda_K e_{[n_1]}$-basis of $(A_K^{\psi=1}(1))^{[n_1]}$. Then

$$\beta_{Iw} := (\mathrm{Exp}^*_{\mathbb{Z}_p})^{-1}(\alpha)$$

is a $\Lambda_K e_{[n_1]}$-basis of $H^1_{Iw}(K, \mathbb{Z}_p(1))^{[n_1]}$ and the element

$$\beta = \mathrm{pr}_{1,1-r}(\beta_{Iw})$$

of Corollary 12 is a $\mathbb{Z}_p[G]e_{[n_1-re]}$-basis of $(H^1(K, \mathbb{Z}_p(1-r))/\mathrm{tor})^{[n_1-re]}$. This follows from the fact that the isomorphism $\mathrm{pr}_{1,1-r}$ of Lemma 8 is not Λ_K-linear but Λ_K-κ_{-r}-semilinear, where κ_j is the automorphism of Λ_K given by $g \mapsto g\chi^{\mathrm{cyclo}}(g)^j$ for $g \in G \times \Gamma_K$. Theorem 9 and Proposition 13 imply

$$\begin{aligned}\exp^*_{\mathbb{Q}_p(r)}(\beta) &= \frac{1}{(r-1)!}\left(\frac{d}{dt}\right)^{r-1} p^{-1}\varphi^{-1}(\alpha)|_{t=0} \\ &= \frac{p^{-r}}{(r-1)!}(\nabla^{r-1}\alpha^{\sigma^{-1}})(\sqrt[e]{\zeta_p - 1}).\end{aligned}$$

Hence the $\mathbb{Z}_p[G]e_{[n_1-re]}$-lattice

$$\mathbb{Z}_p[G] \cdot (r-1)! \cdot p^{r-1} \cdot \exp^*_{\mathbb{Q}_p(r)}(\beta) \subset K^{[n_1-re]} \tag{48}$$

is free over $\mathbb{Z}_p[\Sigma]$ with basis

$$(r-1)! \cdot p^{r-1} \cdot \frac{p^{-r}}{(r-1)!}(\nabla^{r-1}\alpha_i^{\sigma^{-1}})\sqrt[e]{\zeta_p - 1} = p^{-1} \cdot (\nabla^{r-1}\alpha_i^{\sigma^{-1}})(\sqrt[e]{\zeta_p - 1}),$$

where $i=1,\dots,f_\eta$. Now the conjunction of (16) for $\chi=([n_1-re],\eta')$ over all η' is equivalent to the statement that the lattice (48) and the $[n_1-re]$-isotypic component of the inverse different

$$\big(\sqrt[e]{\zeta_p-1}\big)^{-(e(p-1)-1)}\mathcal{O}_K$$

have the same $\mathbb{Z}_p[\Sigma]$-volume. Since $e\nmid n_1$ we have

$$\Big(\big(\sqrt[e]{\zeta_p-1}\big)^{-(e(p-1)-1)}\mathcal{O}_K\Big)^{[n_1-re]}=(p^{-1}\mathcal{O}_K)^{[n_1-re]}$$

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 ξ a $\mathbb{Z}_p[\Sigma]$-basis of $\mathcal{O}_F$.

Proposition 45. *There exists a $\mathbb{Z}_p[\Sigma][[\gamma_1-1]]$-basis*

$$\alpha_i=\xi\cdot\pi_K^{l(\alpha_i)}+\cdots\in A_K^{\psi=1},\quad i=1,\dots,f_\eta$$

of $(A_K^{\psi=1})^{[n_1-e]}$ 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

$$(A_K^{\psi=1})^{[n_1-e]}/(p,\gamma_1-1)\cong\big(A_K^{\psi=1}/(p,\gamma_1-1)\big)^{[n_1-e]}. \tag{49}$$

By Lemma 32 we have $A_K^{\psi=1}/pA_K^{\psi=1}=E_K^{\psi=1}$. By Proposition 38 (reductions mod p of) elements α_i as described in Proposition 45 exist in $E_K^{\psi=1}$. 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 α_i and assume

$$a'\equiv(\gamma_1-1)a\mod p$$

for some $a\in A_K^{\psi=1}$. By Lemma 46 below we have $l(a')\geq -e+e(p-1)$. Since $l(a')=l(\alpha_i)$ for some i, this implies

$$l(a')\equiv -e+e(p-1)\equiv -2e\mod p.$$

Using Lemma 46 again we have $l(a)\leq l(a')-e(p-1)\equiv -e\mod p$. Since $l(a)\not\equiv -e\mod p$ by Proposition 38 we have strict inequality. Lemma 46 then shows $p\mid l(a)$ and hence $p\mid l(a')$, contradicting $l(a')\equiv -2e\mod p$. We conclude that the α_i are linearly independent in (49). Since the $\mathbb{F}_p[\Sigma]$-rank of (49) is f_η this finishes the proof. □

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 $\chi^{\mathrm{cyclo}}(\gamma_1) = 1 + p$ we find from (20) that (in E_K)

$$\gamma_1(\pi) = \pi + \pi^p + \pi^{p+1}$$

and hence for $n = jp^\kappa$

$$\begin{aligned}(\gamma_1 - 1)\pi_K^n &= (\pi + \pi^p + \pi^{p+1})^{n/e} - \pi^{n/e} = \pi_K^n\big((1 + \pi^{p-1} + \pi^p)^{n/e} - 1\big)\\ &= \pi_K^n\big((1 + \pi^{p^\kappa(p-1)} + \pi^{p^{\kappa+1}})^{j/e} - 1\big)\\ &= \frac{j}{e} \cdot \pi_K^{n+ep^\kappa(p-1)} + \cdots\end{aligned}$$

and this is indeed the leading term since $p \nmid j$. The last assertion follows from Proposition 34(a). □

Proposition 47. *If $e < p$, the identity (47) holds for $r = 1$.*

Proof. We first remark that for each i we have

$$v_\varpi(\alpha_i) = l(\alpha_i) = \begin{cases} n_i - e & \text{if } p \nmid n_i, \\ n_i - e + e(p-1) & \text{if } p \mid n_i \end{cases}$$

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

$$q_i := \alpha_i^{\sigma^{-1}}\big(\sqrt[e]{\zeta_p - 1}\big) \in K$$

is already a basis of $\mathcal{O}_K^{[n_1-e]}$. Otherwise

$$p \cdot q_1,\, p^{-1} \cdot q_2, q_3, \ldots, q_{f_\eta}$$

is a basis of $\mathcal{O}_K^{[n_1-e]}$. 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

$$\exp : p\cdot\mathcal{O}_K \cong 1+p\cdot\mathcal{O}_K,$$

the (equivariant) relative volume of $\exp(\mathcal{O}_K)$ and $(\mathcal{O}_K^\times)^\wedge \subseteq H^1(K,\mathbb{Z}_p(1))$ 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 $v_\varpi(\nabla\alpha_i)$, i.e., prove the analogues of Corollary 40 and Proposition 41 for $\nabla a\in A_K^{\psi=p}$.

Lemma 49. *Assume $e<p/2$. For $a\in A_K^{\psi=1}$ 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

$$\nabla\pi_K^j=\frac{j}{e}\pi_K^{j-e}+\frac{j}{e}\pi_K^j, \tag{50}$$

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

$$v_\varpi\left(a_j\frac{j}{e}\varpi^{j-e}\right)>l(a)-e,\quad v_\varpi\left(a_j\frac{j}{e}\varpi^{j}\right)>l(a)>l(a)-e \tag{51}$$

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

$j = l(a) - e(p-1)$. Normalizing the leading coefficient to be ξ (as in the α_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

$$\nabla a \equiv \xi \cdot \frac{\mu p}{e} \cdot \frac{\mu p - e}{e} \cdot \pi_K^{l(a)-e-e(p-1)} + \cdots + \xi \cdot \frac{l(a)}{e} \cdot \pi_K^{l(a)-e} + \cdots \mod p^2$$

and hence

$$\frac{\mu p}{e} \cdot \frac{\mu p - e}{e} \cdot \varpi^{l(a)-e-e(p-1)} + \frac{l(a)}{e} \cdot \varpi^{l(a)-e} \equiv \left(-\frac{\mu}{e} \cdot \frac{\mu p - e}{e} + \frac{l(a)}{e}\right) \cdot \varpi^{l(a)-e} \mod p^2.$$

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 $p \nmid \mu - 2e$. In the proof of Proposition 41(b) we showed $v_\varpi(a_j \varpi^j) > l(a)$ for $j \neq l(a), l(a) - e(p-1)$ and as above this implies that the corresponding terms in ∇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 $a \in (A_K^{\psi=1})^{[\mu p]}$ with $l(a) = \mu p$ and*

$$v_\varpi(\nabla a) = l(\nabla a) = \mu p - e + e(p-1).$$

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 $a' \in A_K^{\psi=1}$ with

$$a' \equiv \pi_K^{\mu p} - \pi_K^{\mu p + e(p-1)} + \cdots \pmod{p^2},$$

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 $H_{p-1} \equiv 0 \pmod p$ (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 $a'_i \not\equiv 0$ 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

$$a = e_{[\mu p]}a' \in (A_K^{\psi=1})^{[\mu p]}$$

so that $l(a) = l_2(a) = \mu p$ by Lemma 42. We have

$$\nabla a' \equiv \frac{\mu p}{e} \cdot \pi_K^{\mu p-e} + \frac{\mu p}{e} \cdot \pi_K^{\mu p} - \frac{\mu p+e(p-1)}{e} \cdot \pi_K^{\mu p-e+e(p-1)} + \cdots \pmod{p^2}$$

and hence

$$\begin{aligned}\nabla a &= \nabla e_{[\mu p]}a' = e_{[\mu p-e]}\nabla a' \\ &\equiv \frac{\mu p}{e} \cdot \pi_K^{\mu p-e} + \cdots - \left(\frac{\mu p+e(p-1)}{e} - \frac{\mu p}{e}x\right) \cdot \pi_K^{\mu p-e+e(p-1)} + \cdots \pmod{p^2},\end{aligned}$$

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 $v_\varpi(\nabla a) = l(\nabla a)$ write

$$\nabla a = \sum_i b_i \cdot \pi_K^i.$$

The terms for $i = \mu p-e$ and $i = \mu p-e+e(p-1)$ contribute the leading term in the variable ϖ

$$\begin{aligned}&\frac{\mu p}{e} \cdot \varpi^{\mu p-e} - \left(\frac{\mu p+e(p-1)}{e} - \frac{\mu p}{e}x\right) \cdot \varpi^{\mu p-e+e(p-1)} + \cdots \\ &\qquad = \left(-\frac{\mu}{e} - \frac{\mu p+e(p-1)}{e} + \frac{\mu p}{e}x\right) \cdot \varpi^{\mu p-e+e(p-1)} + \cdots\end{aligned}$$

since, similarly to (44), we have $p \nmid -\frac{\mu}{e}+1$ as $e < p$. For the terms with $i \neq \mu p-e+e(p-1), \mu p-e$ we must again verify that

$$v_\varpi(b_i\varpi^i) > \mu p-e+e(p-1).$$

This is clear for $i > \mu p-e+e(p-1)$ and for

$$\mu p-e < i < \mu p-e+e(p-1)$$

since then $p \mid b_i$. For $i < \mu p - e$ it suffices to show by (51) that we have instead

$$v_\varpi(a_i \varpi^i) > \mu p + e(p-1)$$

for $i < \mu p$. Since $l_2(a) = \mu p$ we have $v_\varpi(a_i) \geq 2e(p-1)$ for

$$\mu p - e(p-1) < i < \mu p$$

and hence $v_\varpi(a_i \varpi^i) > \mu p + e(p-1)$. For

$$\mu p - 2e(p-1) < i \leq \mu p - e(p-1)$$

we have by $v_\varpi(a_i) \geq 3e(p-1)$ by Proposition 34(a) since

$$i \leq \mu p - e(p-1) < -\left(3 - \frac{2}{p}\right) \cdot e.$$

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 μ. Finally for

$$i \leq \mu p - 2e(p-1) < -e - e(p-1) = -ep$$

we have by (41)

$$\begin{aligned} v_\varpi(a_i \varpi^i) &\geq -(p-1)i - pe > (p-1)pe - pe = (p-2)pe \\ &\geq (2p-3)e = -e + 2e(p-1) > \mu p + e(p-1). \end{aligned}$$

□

Proposition 51. *If $e < p/4$ the identity (47) holds for $r = 2$.*

Proof. By Lemmas 49 and 50 we can choose α_i such that

$$v_\varpi(\nabla \alpha_i) = l(\nabla \alpha_i) = \begin{cases} n_i - 2e & \text{if } p \nmid n_i \text{ and } p \nmid n_i - e, \\ n_i - 2e + e(p-1) & \text{if } p \mid n_i \text{ or } p \mid n_i - e. \end{cases}$$

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

$$q_i := \nabla \alpha_i^{\sigma^{-1}}(\sqrt[e]{\zeta_p - 1}) \in K$$

be the basis of L_2. We again find that

$$\begin{aligned} p\cdot q_1,\, p^{-1}\cdot q_2,\dots,\, p\cdot q_h,\, p^{-1}\cdot q_{h+1},\dots,q_{f_\eta} &\quad \text{if } n_1<e \text{ and } e<n_h<2e,\\ p\cdot q_1,\, p^{-1}\cdot q_2,\dots,\, q_h,\, q_{h+1},\dots,q_{f_\eta} &\quad \text{if } n_1<e \text{ and } \nexists\, e<n_h<2e,\\ q_1,q_2,\dots,\, p\cdot q_h,\, p^{-1}\cdot q_{h+1},\dots,q_{f_\eta} &\quad \text{if } \nexists\, n_1<e \text{ and } e<n_h<2e,\\ q_1,q_2,\dots,q_h,q_{h+1},\dots,q_{f_\eta} &\quad \text{if } \nexists\, n_1<e \text{ nor } e<n_h<2e, \end{aligned}$$

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

[Benois and Berger 2008] D. Benois and L. Berger, "Théorie d'Iwasawa des représentations cristallines, II", *Comment. Math. Helv.* **83**:3 (2008), 603–677. MR 2410782 Zbl 1157.11041

[Bley and Cobbe 2016] W. Bley and A. Cobbe, "The equivariant local ϵ-constant conjecture for unramified twists of $\mathbb{Z}_p(1)$", preprint, 2016. arXiv 1602.07858

[Bloch and Kato 1990] S. Bloch and K. Kato, "L-functions and Tamagawa numbers of motives", pp. 333–400 in *The Grothendieck Festschrift, I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR 1086888 Zbl 0768.14001

[Breuning 2004] M. Breuning, "Equivariant local epsilon constants and étale cohomology", *J. London Math. Soc.* (2) **70**:2 (2004), 289–306. MR 2078894 Zbl 1068.11075

[Burns and Flach 2006] D. Burns and M. Flach, "On the equivariant Tamagawa number conjecture for Tate motives, II", *Doc. Math.* Extra Vol. (2006), 133–163. MR 2290586 Zbl 1156.11042

[Cherbonnier and Colmez 1999] F. Cherbonnier and P. Colmez, "Théorie d'Iwasawa des représentations p-adiques d'un corps local", *J. Amer. Math. Soc.* **12**:1 (1999), 241–268. MR 1626273 Zbl 0933.11056

[Coleman 1979] R. F. Coleman, "Division values in local fields", *Invent. Math.* **53**:2 (1979), 91–116. MR 560409 Zbl 0429.12010

[Colmez 1999] P. Colmez, "Représentations cristallines et représentations de hauteur finie", *J. Reine Angew. Math.* **514** (1999), 119–143. MR 1711279 Zbl 1191.11032

[Daigle 2014] J. Daigle, *On the local Tamagawa Number Conjecture for Tate motives*, Ph.D. thesis, Caltech, 2014, available at http://thesis.library.caltech.edu/8427/.

[Deligne 1973] P. Deligne, "Les constantes des équations fonctionnelles des fonctions L", pp. 501–597 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuijk, Lecture Notes in Math **349**, Springer, Berlin, 1973. MR 0349635 Zbl 0271.14011

[Deligne 1987] P. Deligne, "Le déterminant de la cohomologie", pp. 93–177 in *Current trends in arithmetical algebraic geometry* (Arcata, CA, 1985), edited by K. A. Ribet, Contemp. Math. **67**, Amer. Math. Soc., Providence, RI, 1987. MR 902592 Zbl 0629.14008

[Fontaine and Perrin-Riou 1994] J.-M. Fontaine and B. Perrin-Riou, "Autour des conjectures de Bloch et Kato: Cohomologie galoisienne et valeurs de fonctions L", pp. 599–706 in *Motives* (Seattle, WA, 1991), edited by U. Jannsen et al., Proc. Sympos. Pure Math. **55**, Amer. Math. Soc., Providence, RI, 1994. MR 1265546 Zbl 0821.14013

[Fröhlich 1976] A. Fröhlich, "Arithmetic and Galois module structure for tame extensions", *J. Reine Angew. Math.* **286/287** (1976), 380–440. MR 0432595 Zbl 0385.12004

[Fukaya and Kato 2006] T. Fukaya and K. Kato, "A formulation of conjectures on p-adic zeta functions in noncommutative Iwasawa theory", pp. 1–85 in *Proceedings of the St. Petersburg Mathematical Society, XII*, edited by N. N. Uraltseva, Amer. Math. Soc. Transl. (2) **219**, Amer. Math. Soc., Providence, RI, 2006. MR 2276851 Zbl 1238.11105

[Kato 1993] K. Kato, "Lectures on the approach to Iwasawa theory for Hasse–Weil L-functions via B_{dR}, II", unpublished preprint, 1993.

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, New York, 2002. MR 1878556 Zbl 0984.00001

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften **322**, Springer, Berlin, 1999. MR 1697859 Zbl 0956.11021

[Neukirch et al. 2000] J. Neukirch, A. Schmidt, and K. Wingberg, *Cohomology of number fields*, Grundlehren der Mathematischen Wissenschaften **323**, Springer, Berlin, 2000. MR 1737196 Zbl 0948.11001

[Perrin-Riou 1990] B. Perrin-Riou, "Théorie d'Iwasawa p-adique locale et globale", *Invent. Math.* **99**:2 (1990), 247–292. MR 1031902 Zbl 0715.11030

[Perrin-Riou 1994] B. Perrin-Riou, "Théorie d'Iwasawa des représentations p-adiques sur un corps local", *Invent. Math.* **115**:1 (1994), 81–149. MR 1248080 Zbl 0838.11071

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 554237 Zbl 0423.12016

[Venjakob 2013] O. Venjakob, "On Kato's local ϵ-isomorphism conjecture for rank-one Iwasawa modules", *Algebra Number Theory* **7**:10 (2013), 2369–2416. MR 3194646 Zbl 1305.11095

[Washington 1997] L. C. Washington, *Introduction to cyclotomic fields*, 2nd ed., Graduate Texts in Mathematics **83**, Springer, New York, 1997. MR 1421575 Zbl 0966.11047

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1277

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.
MSC2010: primary 14G35; secondary 14H40, 11F27, 11F30.
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 $J_{\mathfrak{m}}(X)$ of a modular curve X of arbitrary level with cuspidal modulus $\mathfrak{m}$. Then we prove that the generating series of these classes is a weakly holomorphic modular form of weight $\frac{3}{2}$ with values in $J_{\mathfrak{m}}(X)$. Our argument is a generalization of Borhcerds' 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 $J_{\mathfrak{m}}(X)$ 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 tr_F on $J_{\mathfrak{m}}(X)$. The value of 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 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 $\Gamma_0^*(p)$ of $\Gamma_0(p)$ in $\mathrm{PSL}_2(\mathbb{Z})$ 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 ∞ be the cusp of $X_0^*(p)$ and let m be a nonnegative integer. Then $\mathfrak{m} = m\cdot(\infty)$ is an effective divisor. Recall that the generalized Jacobian $J_{\mathfrak{m}}(X_0^*(p))$ of $X_0^*(p)$ associated with the modulus $\mathfrak{m}$ is a commutative algebraic group whose rational points correspond to classes of divisors of degree zero modulo $\mathfrak{m}$-equivalence; see Section 2 and [Serre 1988]. If $m = 0$, then $J_{\mathfrak{m}}(X_0^*(p))$ is simply the usual Jacobian. For any integer d, let $\mathcal{Q}_{p,d}$ 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 $\mathcal{Q}_{p,d}$ with finitely many orbits.

If d is positive, then any $Q \in \mathcal{Q}_{p,d}$ defines a point α_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

$$Y(d) = \sum_{Q \in \mathcal{Q}_{p,d}/\Gamma_0^*(p)} \frac{1}{|\Gamma_0^*(p)_Q|} \cdot (\alpha_Q),$$

where $\Gamma_0^*(p)_Q$ is the (finite) stabilizer of Q (see Equation (1.5) in [Bruinier and

Funke 2006]). The divisor

$$Z(d) = Y(d) - \deg(Y(d)) \cdot (\infty)$$

has degree zero and is defined over $\mathbb{Q}$. We denote by $[Z(d)]_{\mathfrak{m}}$ its class in the generalized Jacobian $J_{\mathfrak{m}}(X_0^*(p))$.

If d is negative, any $Q \in \mathcal{Q}_{p,d}$ defines an oriented geodesic cycle on $\mathbb{H} \cup P^1(\mathbb{R})$, given by the equation $a|z|^2 + b\Re(z) + c = 0$. 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 ∞ 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 ∞. We let $h_Q \in \mathbb{Q}(X_0^*(p))^\times$ be a function satisfying

$$h_Q = 1 - q_\infty^b + O(q_\infty^m)$$

at the cusp ∞, where q_∞ is the uniformizing parameter of the completed local ring at ∞ given by the Tate curve over $\mathbb{Z}[\![q_\infty]\!]$. Then we define

$$[Z(d)]_{\mathfrak{m}} = [\operatorname{div}(h_{[0,b,0]})]_{\mathfrak{m}} = \sum_{Q \in \mathcal{Q}_{p,d}/\Gamma_0^*(p)} \frac{1}{b} \cdot [\operatorname{div}(h_Q)]_{\mathfrak{m}}.$$

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 $[Z(d)]_{\mathfrak{m}} = 0$. Finally, for $d = 0$, we define $[Z(0)]_{\mathfrak{m}}$ 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 $[Z(d)]_{\mathfrak{m}}$, we consider the generating series

$$A_{\mathfrak{m}}(\tau) = \sum_{\substack{d \in \mathbb{Z} \\ d > -m^2}} [Z(d)]_{\mathfrak{m}} \cdot q^d \in \mathbb{C}((q)) \otimes J_{\mathfrak{m}}(X_0^*(p)).$$

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 $A_{\mathfrak{m}}(\tau)$ is a weakly holomorphic modular form of weight $\frac{3}{2}$ for the group $\Gamma_0(4p)$, that is, $A_{\mathfrak{m}}(\tau) \in M^!_{3/2}(\Gamma_0(4p)) \otimes J_{\mathfrak{m}}(X_0^*(p))$.*

Under the natural map

$$J_{\mathfrak{m}}(X_0^*(p)) \to J(X_0^*(p))$$

the classes $[Z(d)]_{\mathfrak{m}}$ with $d \leq 0$ are mapped to zero. Applying it to $A_{\mathfrak{m}}(\tau)$, 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 $J_{\mathfrak{m}}(X_0^*(p))$. 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

$$F^+(\tau) = \sum_{n \gg -\infty} c_F^+(n) \cdot q_\infty^n.$$

Proposition 1.3. *Assume that $c_F^+(n) = 0$ for $n \leq -m$ and $c_F^+(0) = 0$. Then there is a linear map $\mathrm{tr}_F : J_{\mathfrak{m}}(X_0^*(p)) \to \mathbb{C}$ defined by*

$$[D]_{\mathfrak{m}} \mapsto \mathrm{tr}_F(D) := \sum_{a \in \mathrm{supp}(D) \setminus \{\infty\}} n_a \cdot F(a),$$

for divisors $D = \sum_a n_a \cdot (a)$ in $\mathrm{Div}^0(X_0^*(p))$.

The images under tr_F of the classes $[Z(d)]_{\mathfrak{m}}$ 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 $\mathrm{tr}_F(A_{\mathfrak{m}})$ is a weakly holomorphic modular form in the space $M^!_{3/2}(\Gamma_0(4p))$. It is explicitly given by*

$$\mathrm{tr}_F(A_{\mathfrak{m}}) = \sum_{d>0} F(Y(d)) \cdot q^d + \sum_{n \geq 1} c_F^+(-n)(\sigma_1(n) + p\sigma_1(n/p)) - \sum_{b>0} \sum_{n>0} c_F^+(-bn) \cdot b \cdot q^{-b^2}.$$

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 SL_2 and $\mathrm{O}(1, 2)$. 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 $\mathrm{SL}_2(\mathbb{Z})$ 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 $\mathrm{Div}^0(X)$ 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 $\mathrm{Div}^0(X)/P(X)$.

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

$$\mathfrak{m} = \sum_{s\in S} m_s \cdot (s)$$

is an effective divisor defined over k. Let $\mathcal{O}_s$ be the ring of integers in the completion $k(X)_s$ of $k(X)$ at s, and let $\pi_s \in \mathcal{O}_s$ be a uniformizer. If $f, g \in k(X)_s$ and $n \in \mathbb{Z}$, we write

$$f = g + O(\pi_s^n)$$

if $f - g \in \pi_s^n \mathcal{O}_s$. We consider the subgroup

$$P_{\mathfrak{m}}(X) = \{\mathrm{div}(f) : f \in k(X)^\times \text{ with } \pi_s^{-\mathrm{ord}_s(f)} f = 1 + O(\pi_s^{m_s}) \text{ for all } s \in S\}$$

of $P(X)$. The generalized Jacobian $J_{\mathfrak{m}}(X)$ associated with the modulus $\mathfrak{m}$ is a commutative algebraic group over k, whose k-rational points satisfy

$$J_{\mathfrak{m}}(X)(k) \cong \mathrm{Div}^0(X)/P_{\mathfrak{m}}(X). \tag{2-1}$$

The quotient on the right hand side is also canonically isomorphic to the subgroup of divisors in $\mathrm{Div}^0(X)$ coprime to S modulo $\mathfrak{m}$-equivalence. For a divisor $D \in \mathrm{Div}^0(X)$ we denote by $[D]_{\mathfrak{m}}$ the corresponding class in $J_{\mathfrak{m}}(X)(k)$.

There is a canonical rational map $\varphi_{\mathfrak{m}} : X \to J_{\mathfrak{m}}(X)$ defined over k which is regular outside S, see [Serre 1988, Chapter 5, Theorem 1]. If $\mathfrak{m}'$ is another effective divisor on X satisfying $\mathfrak{m} \geq \mathfrak{m}' \geq 0$, there exists a unique homomorphism $J_{\mathfrak{m}} \to J_{\mathfrak{m}'}$

which is compatible with $\varphi_{\mathfrak{m}}$ and $\varphi_{\mathfrak{m}'}$. It is surjective and separable [Serre 1988, Chapter 5, Proposition 6]. In particular, there exists a surjective homomorphism

$$J_{\mathfrak{m}}(X) \to J(X). \tag{2-2}$$

Its kernel is isomorphic to

$$H_{\mathfrak{m}} = \Big(\prod_{\substack{s\in S\\ m_s>0}} \mathbb{G}_m \times \mathbb{G}_a^{m_s-1} \Big) / \mathbb{G}_m, \tag{2-3}$$

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

$$h_{s,n} = \begin{cases} 1 - \pi_s^n + O(\pi_s^{m_s}) & \text{at } s, \\ 1 + O(\pi_t^{m_t}) & \text{at all } t \in S \setminus \{s\}. \end{cases} \tag{2-4}$$

An argument as in [Serre 1988, Chapter 5, Proposition 8] shows that the “additive part” of $H_{\mathfrak{m}}$ is generated by the classes

$$[\mathrm{div}(h_{s,n})]_{\mathfrak{m}}, \tag{2-5}$$

for $s \in S$ and $0 < n < m_s$. Note that for $n \geq m_s$ the class $[\mathrm{div}(h_{s,n})]_{\mathfrak{m}}$ 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 $J_{\mathfrak{m}}(X)$ as follows. It is easily seen that there exists a rational section f of $\mathcal{L}$ such that

$$\phi_s^{-1} f = \pi_s^{a_s} \cdot (1 + O(\pi_s^{m_s})), \tag{2-6}$$

for some $a_s \in \mathbb{Z}$ at every $s \in S$. Then we define

$$[(\mathcal{L}, (\phi_s))]_{\mathfrak{m}} = [\mathrm{div}(f) - \deg(\mathcal{L}) \cdot (s_0)]_{\mathfrak{m}} \in J_{\mathfrak{m}}(X)(k). \tag{2-7}$$

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 $\mathrm{disc}(L) = |L'/L|$ the discriminant of L. We let $\mathrm{SO}(L)$ be the special orthogonal group of L and write $\mathrm{SO}^+(L)$ for the intersection of $\mathrm{SO}(L)$ with the connected component of the identity of $\mathrm{SO}(L)(\mathbb{R})$. The even Clifford algebra of $L_\mathbb{Q}$ is isomorphic to the matrix algebra $\mathrm{Mat}_2(\mathbb{Q})$, which induces an isomorphism $\mathrm{PGL}_2(\mathbb{Q}) \cong \mathrm{SO}(L)(\mathbb{Q})$. We realize the hermitian symmetric space corresponding to $\mathrm{SO}(L)$ as the domain

$$\mathcal{D} = \{z \in L_\mathbb{C} : (z, z) = 0, (z, \bar{z}) < 0\}/\mathbb{C}^\times.$$

It decomposes into 2 connected components. We fix one of these components and denote it by $\mathcal{D}^+$.

Let $\Gamma = \Gamma_L$ be the discriminant kernel subgroup of $\mathrm{SO}^+(L)$, that is, the kernel of the natural homomorphism

$$\mathrm{SO}^+(L) \to \mathrm{Aut}(L'/L).$$

Recall that rescaling the quadratic form by a factor of n does not change $\mathrm{SO}^+(L)$ while it replaces the discriminant kernel by the full congruence subgroup of level n. We denote by

$$Y_\Gamma = \Gamma\backslash\mathcal{D}^+ \tag{3-1}$$

the noncompact modular curve associated with Γ.

Let $\mathrm{Iso}(L)$ be the set of isotropic lines in L (i.e., primitive isotropic rank-1 sublattices $I \subset L$). The group Γ acts with finitely many orbits on $\mathrm{Iso}(L)$. We denote by X_Γ the compact modular curve obtained by adding to Y_Γ the cusps corresponding to the Γ-classes of isotropic lines $I \in \mathrm{Iso}(L)$; see, e.g., [Bruinier and Funke 2006]. It is well known that X_Γ 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 $I_0 \in \mathrm{Iso}(L)$ together with an orientation on I_0 given by a basis vector $x_0 \in I_{0,\mathbb{R}}$. For any other $I \in \mathrm{Iso}(L)$ we choose a $g \in \mathrm{SO}^+(L)(\mathbb{R})$ such that $gI_{0,\mathbb{R}} = I_\mathbb{R}$. Then $gx_0 \in I_\mathbb{R}$ 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

$$(\ell, \ell') = 1. \tag{3-2}$$

We let K be the even negative definite lattice

$$K = L \cap \ell^\perp \cap \ell'^\perp. \tag{3-3}$$

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

$$\mathbb{H} \to \mathcal{D}, \quad w \mapsto \mathbb{C}^\times\big(w \otimes \ell_K + \ell' - Q(w \otimes \ell_K)\ell - Q(\ell')\ell\big) \tag{3-4}$$

is injective and has one of the two connected components of $\mathcal{D}$ as its image. Possibly replacing ℓ_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 $\mathrm{PGL}_2^+(\mathbb{Q})$ on $\mathbb{H}$ by fractional linear transformations and on $\mathcal{D}^+$ via the isomorphism with $\mathrm{SO}^+(L)(\mathbb{Q})$. For $\mu \in L_{\mathbb{Q}} \cap I^\perp$ we consider the Eichler transformation

$$E_{\ell,\mu}(x) = x + (x,\ell)\mu - (x,\mu)\ell - (x,\ell)Q(\mu)\ell \tag{3-5}$$

in $\mathrm{SO}^+(L)(\mathbb{Q})$. It belongs to Γ if $\mu \in K$.

Lemma 3.1. *The stabilizer in Γ 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 ℓ and ℓ' fixed. Consequently, it maps the orthogonal complement K to itself, and therefore ℓ_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 σ taking ℓ to $-\ell$, and ℓ' to $-\ell'$, and ℓ_K to itself belongs to $\mathrm{SO}(L)(\mathbb{Q})$. The element $\sigma\gamma \in \mathrm{SO}(L)(\mathbb{Q})$ fixes ℓ. Arguing as above, we see that it is equal to an Eichler transformation $E_{\ell,u} \in \mathrm{SO}^+(L)(\mathbb{Q})$. This implies that σ belongs to the connected component of the identity of $\mathrm{SO}(L)(\mathbb{R})$. But this leads to a contradiction, since the spinor norm of σ is negative, showing that the case $\gamma\ell = -\ell$ cannot occur. □

The action of $\mathbb{Z}$ on $\mathbb{H}$ by translations corresponds to the action of Γ_I on $\mathcal{D}^+$. The induced map

$$\mathbb{Z}\backslash\mathbb{H} \to \Gamma_I\backslash\mathcal{D}^+ \tag{3-6}$$

is an isomorphism. Hence, $q_I = e^{2\pi i w}$ defines a local parameter at the cusp c_I of X_Γ.

Example 3.2. In the special case when $N_I = 1$, then $4N = -Q(\ell_K)$ and the discriminant kernel subgroup Γ is isomorphic to $\Gamma_0(N/4)$. The curve X_Γ is isomorphic to $X_0(N/4)$, with c_I corresponding to the cusp at ∞; see, e.g., [Bruinier and Ono 2010, Section 2.4].

The Weil representation. Let $\mathrm{Mp}_2(\mathbb{Z})$ be the metaplectic extension of $\mathrm{SL}_2(\mathbb{Z})$ by $\{\pm 1\}$, realized by the two possible choices of a holomorphic square root of the automorphy factor $c\tau + d$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$; see, e.g., [Borcherds 1998; Kudla 2003].

Recall that there is a Weil representation ω_L of $\mathrm{Mp}_2(\mathbb{Z})$ 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 ω_L can be viewed as the restriction of the usual Weil representation of $\mathrm{Mp}_2(\hat{\mathbb{Q}})$ on $L \otimes \hat{\mathbb{Q}}$ with respect to the standard additive character of $\hat{\mathbb{Q}}$; see [Kudla 2003]. The representation ω_L is the complex conjugate of the representation ρ_L in [Borcherds 1998; Bruinier 2002; Bruinier and Funke 2006]. The action of $\mathrm{Mp}_2(\mathbb{Z})$ on S_L commutes with the natural action of $\mathrm{Aut}(L'/L)$ 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 $\mathrm{Mp}_2(\mathbb{Z})$ of weight k with representation ω_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 Γ 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 λ is finite, and there is a unique point $z_\lambda \in \mathcal{D}^+$ which is orthogonal to λ. For $d \in \mathbb{Q}_{>0}$ and $\varphi \in S_L$ we consider the Heegner divisor

$$Y(d, \varphi) = \sum_{\lambda \in L'_d/\Gamma} \frac{1}{2|\Gamma_\lambda|} \varphi(\lambda) \cdot (z_\lambda) \tag{3-7}$$

on X_Γ. It is defined over the field of definition of X_Γ and has coefficients in the field of definition of φ. Let $I_0 \in \mathrm{Iso}(L)$ be a fixed isotropic line. We define a divisor of degree 0 on X_Γ by putting

$$Z(d, \varphi) = Y(d, \varphi) - \deg(Y(d, \varphi)) \cdot (c_{I_0}). \tag{3-8}$$

4. A generalized Gross–Kohnen–Zagier theorem

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 $S = \{c_I : I \in \mathrm{Iso}(L)/\Gamma\}$ be the set of cusps of X and let

$$\mathfrak{m} = \sum_{I \in \mathrm{Iso}(L)/\Gamma} m_I \cdot (c_I)$$

be a fixed effective divisor supported on S. We consider the generalized Jacobian of X associated with the modulus $\mathfrak{m}$. For $I \in \operatorname{Iso}(L)$, 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 $X_\Gamma \cong X_0(N/4)$).

Since, throughout this section, we are only interested in the k-valued points of the generalized Jacobian, we briefly write $J_{\mathfrak{m}}(X)$ instead of $J_{\mathfrak{m}}(X)(k)$. For every degree-zero divisor $D = \sum a_I \cdot (c_I) \in \operatorname{Div}^0(X)$ 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

$$u_{D,r} = r_I q_I^{a_I} \cdot (1 + O(q_I^{m_I})) \tag{4-1}$$

at c_I for $I \in \operatorname{Iso}(L)$. We write $H_{\mathbb{G}_m,\mathfrak{m}}$ for the subgroup of $J_{\mathfrak{m}}(X)$ generated by the classes $[\operatorname{div}(u_{D,r})]_{\mathfrak{m}}$ of all these functions and let

$$J_{\mathfrak{m}}^{\mathrm{add}}(X) = J_{\mathfrak{m}}(X)/H_{\mathbb{G}_m,\mathfrak{m}}. \tag{4-2}$$

By definition we have $J_{\mathfrak{m}}^{\mathrm{add}}(X) = J_{\mathfrak{m}}(X)$ when $|S| = 1$. By the Manin–Drinfeld theorem we have $J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{Q}} = J_{\mathfrak{m}}(X)_{\mathbb{Q}}$ when $\mathfrak{m} = 0$. For general $\mathfrak{m}$ the kernel of the induced homomorphism

$$J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{Q}} \to J(X)_{\mathbb{Q}} \tag{4-3}$$

is isomorphic to the product of the groups $\mathbb{G}_a^{m_I - 1}$ for $I \in \operatorname{Iso}(L)/\Gamma$ with $m_I > 0$.

For $d \in \mathbb{Q}_{>0}$ and $\varphi \in S_L$ we consider the class

$$[Z(d,\varphi)]_{\mathfrak{m}} \in J_{\mathfrak{m}}(X)_{\mathbb{C}} \tag{4-4}$$

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 $\mathrm{SO}(L)(\mathbb{Q}) \cong \mathrm{PGL}_2(\mathbb{Q})$.) Recall that $\mathcal{T}$ is canonically trivial in small neighborhoods of the cusps. Hence, taking the induced trivializations and putting $s_0 = c_{I_0}$ in (2-7), we obtain a class $[\mathcal{M}_k]_{\mathfrak{m}} \in J_{\mathfrak{m}}(X)_{\mathbb{Q}}$. For $d = 0$ we define

$$[Z(0,\varphi)]_{\mathfrak{m}} = \varphi(0) \cdot [\mathcal{M}_{-1}]_{\mathfrak{m}}. \tag{4-5}$$

We also define classes for $d \in \mathbb{Q}_{<0}$ as follows. For a vector $\lambda \in L'_d$, the orthogonal complement $\lambda^\perp \subset L_{\mathbb{Q}}$ is isotropic if and only if $d \in -2\operatorname{disc}(L)(\mathbb{Q}^\times)^2$. In this case there is a unique pair of isotropic lines $I, \tilde{I} \in \operatorname{Iso}(L)$ such that $\lambda^\perp = I_{\mathbb{Q}} \oplus \tilde{I}_{\mathbb{Q}}$ and such that the triple $(\lambda, x, \tilde{x})$ is a positively oriented basis of $L_{\mathbb{Q}}$ 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* λ 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

$$(\mu, L \cap I^\perp) = n_I(\mu) \cdot \mathbb{Z} \tag{4-6}$$

and $\operatorname{sgn}(n_I(\mu)) \cdot \mu \sim I$.

Now, if $d \in -2\operatorname{disc}(L)(\mathbb{Q}^\times)^2$ and $\lambda \in L'_d$, we let $I \in \operatorname{Iso}(L)$ be the isotropic line associated to λ and let $\ell' \in L'$ be as in (3-2) such that $(\ell', I) = \mathbb{Z}$. We choose a function $h_\lambda \in k(X)^\times$ such that

$$h_\lambda = \begin{cases} 1 - e^{2\pi i(\lambda,\ell')} q_I^{n_I(\lambda)} + O(q_I^{m_I}) & \text{at the cusp } c_I, \\ 1 + O(q_J^{m_J}), & \text{at all other cusps } c_J. \end{cases} \tag{4-7}$$

The existence of h_λ follows for instance from the approximation theorem for valuations, see page 29 in [Serre 1988]. If $(\lambda, \ell') \in \mathbb{Z}$, then h_λ agrees with the function $h_{c_I, n_I(\lambda)} \in k(X)^\times$ defined in (2-4). For $\varphi \in S_L$ we define

$$[Z(d,\varphi)]_{\mathfrak{m}} = \sum_{\lambda \in L'_d/\Gamma} \frac{1}{2n_I(\lambda)} (\varphi(\lambda) + \varphi(-\lambda)) \cdot [\operatorname{div}(h_\lambda^{-1})]_{\mathfrak{m}}. \tag{4-8}$$

It is easily checked that this class is independent of the choices of the functions h_λ. If $d < 0$ and $d \notin -2\operatorname{disc}(L)(\mathbb{Q}^\times)^2$, we put $[Z(d,\varphi)]_{\mathfrak{m}} = 0$.

Finally, for all $d \in \mathbb{Q}$ we write $[Z(d)]_{\mathfrak{m}}$ for the element of

$$\operatorname{Hom}(S_L, J_{\mathfrak{m}}(X)_{\mathbb{C}}) \cong J_{\mathfrak{m}}(X)_{\mathbb{C}} \otimes S_L^\vee$$

given by $\varphi \mapsto [Z(d,\varphi)]_{\mathfrak{m}}$.

The classes $[Z(d,\varphi)]_{\mathfrak{m}}$ 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*

$$[Z(d,\varphi)]_{\mathfrak{m}} = \sum_{I \in \operatorname{Iso}(L)/\Gamma} \sum_{\lambda \in L'_{d,I}/I} \tfrac{1}{2}(\varphi(\lambda) + \varphi(-\lambda)) \cdot [\operatorname{div}(h_\lambda^{-1})]_{\mathfrak{m}}.$$

Proof. If $L_{d,I}$ is nonempty and if we fix $\lambda_0 \in L'_{d,I}$, we have

$$\begin{aligned} L'_{d,I} &= \{\lambda_0 + a\ell/N_I : a \in \mathbb{Z}\}, \\ L'_{d,I}/\Gamma_I &= \{\lambda_0 + a\ell/N_I : a \in \mathbb{Z}/N_I n_I(\lambda_0)\mathbb{Z}\}, \\ L'_{d,I}/I &= \{\lambda_0 + a\ell/N_I : a \in \mathbb{Z}/N_I\mathbb{Z}\}. \end{aligned}$$

This implies the assertion. ◻

An abstract modularity theorem. To describe the relations in the generalized Jacobian among the classes $[Z(d)]_{\mathfrak{m}}$ we form the generating series

$$A_{\mathfrak{m}}(\tau) = \sum_{d\in\frac{1}{N}\mathbb{Z}} [Z(d)]_{\mathfrak{m}} \cdot q^d \in S_L^\vee((q)) \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}. \tag{4-9}$$

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 $S_L^\vee \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}$.

Theorem 4.2. *The generating series $A_{\mathfrak{m}}(\tau)$ is the q-expansion of a weakly holomorphic modular form in $M^!_{3/2}(\omega_L^\vee) \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}$.*

To prove this result, we use the following variant of Borcherds' modularity criterion [Borcherds 1999, Theorem 3.1]. Let ρ be a finite dimensional representation of $\mathrm{Mp}_2(\mathbb{Z})$ on a complex vector space V which is trivial on some congruence subgroup. The stabilizer in $\mathrm{Mp}_2(\mathbb{Z})$ of the cusp ∞ is generated by the elements $T = \left(\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right), 1\right)$ and $Z = \left(\left(\begin{smallmatrix}-1&0\\0&-1\end{smallmatrix}\right), i\right)$. The hypothesis on ρ 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 $\mathrm{Mp}_2(\mathbb{Z})$ of weight $k \in \frac{1}{2}\mathbb{Z}$ with representation ρ, 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

$$\rho(T)a(n) = e^{2\pi i n} a(n), \tag{4-10}$$

$$\rho(Z)a(n) = e^{-\pi i k} a(n). \tag{4-11}$$

We write $\rho^\vee$ for the representation dual to ρ, 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

$$f(\tau) = \sum_{n\in\mathbb{Q}} c(n) \cdot q^n \in M^!_{2-k}(\rho^\vee).$$

[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 ρ by a twist with a power of the line bundle $\mathcal{L}(\infty)$ corresponding to the cusp at ∞.

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 Δ 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 Δ 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

$$\sum_{d\in\mathbb{Q}} (c(-d), [Z(d)]_{\mathfrak{m}}) = 0 \in J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}, \tag{4-12}$$

for every

$$f(\tau) = \sum_{d\in\mathbb{Q}} c(d)\cdot q^d \in M^!_{1/2}(\omega_L). \tag{4-13}$$

Since the space $M^!_{1/2}(\omega_L)$ 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 Γ of weight $c(0,0)$ with some multiplier system of finite order. Its divisor on X is given by

$$\operatorname{div}(\Psi(z, f)) = \sum_{d>0} (c(-d), Z(d)) + B(f),$$

where $B(f)$ is a divisor of degree $\frac{1}{12}c(0,0)$ supported at the cusps of X. Let $I \in \operatorname{Iso}(L)$. 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

$$\Psi(w, f) = R_I \cdot q_I^{\rho_I} \prod_{\substack{\lambda\in(L'\cap I^\perp)/I\\ n_I(\lambda)>0}} \bigl(1 - e^{2\pi i(\lambda,\ell')} q_I^{n_I(\lambda)}\bigr)^{c(-Q(\lambda),\lambda)}, \tag{4-14}$$

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 λ 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

$$\prod_{\substack{a\in\mathbb{Z}/N_I\mathbb{Z}\\ a\neq 0}} \bigl(1 - e^{2\pi i a/N_I}\bigr)^{c(0,a\ell/N_I)/2}.$$

Hence, the (finite) product

$$\Psi(w,f)\times R_I^{-1}\prod_{\substack{\lambda\in(L'\cap I^\perp)/I\\ m_I>n_I(\lambda)>0}} h_\lambda^{-c(-Q(\lambda),\lambda)}$$

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

$$\Psi(w,f)\times u_{D,(R_I)}^{-1}\times\prod_{I\in\mathrm{Iso}(L)/\Gamma}\prod_{\substack{\lambda\in(L'\cap I^\perp)/I\\ m_I>n_I(\lambda)>0}} h_\lambda^{-c(-Q(\lambda),\lambda)}$$

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 $u_{D,r}\in H_{\mathbb{G}_m,\mathfrak{m}}$ 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 $J_\mathfrak{m}(X)$ is given by

$$\begin{aligned}[\mathcal{M}_{c(0,0)}]_\mathfrak{m}&=[\mathrm{div}(\Psi(f))-\deg(\mathcal{M}_{c(0,0)})(c_{I_0})]_\mathfrak{m}-[\mathrm{div}(u_{D,(R_I)})]_\mathfrak{m}\\&\quad-\sum_{I\in\mathrm{Iso}(L)/\Gamma}\sum_{\substack{\lambda\in(L'\cap I^\perp)/I\\ m_I>n_I(\lambda)>0}} c(-Q(\lambda),\lambda)\cdot[\mathrm{div}(h_\lambda)]_\mathfrak{m}.\end{aligned}\tag{4-15}$$

Using Lemma 4.1, we see that

$$\sum_{I\in\mathrm{Iso}(L)/\Gamma}\sum_{\substack{\lambda\in(L'\cap I^\perp)/I\\ m_I>n_I(\lambda)>0}} c(-Q(\lambda),\lambda)\cdot[\mathrm{div}(h_\lambda)]_\mathfrak{m}=-\sum_{d<0}\bigl(c(-d),[Z(d)]_\mathfrak{m}\bigr).$$

Inserting this into (4-15), we obtain the relation

$$-c(0,0)[\mathcal{M}_{-1}]_\mathfrak{m}=\sum_{d>0}\bigl(c(-d),[Z(d)]_\mathfrak{m}\bigr)+\sum_{d<0}\bigl(c(-d),[Z(d)]_\mathfrak{m}\bigr)-[\mathrm{div}(u_{D,(R_I)})]_\mathfrak{m}$$

in $J_\mathfrak{m}(X)_\mathbb{C}$. This implies (4-12) in $J_\mathfrak{m}^{\mathrm{add}}(X)_\mathbb{C}$, concluding the proof. □

Remark 4.4. To be able to describe the generating series in $J_\mathfrak{m}(X)_\mathbb{C}$ instead of in the quotient $J_\mathfrak{m}^{\mathrm{add}}(X)_\mathbb{C}$, 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–Drinfed theorem, the natural homomorphism $J_\mathfrak{m}(X)\to J(X)$ induces a linear map

$$J_\mathfrak{m}^{\mathrm{add}}(X)_\mathbb{C}\to J(X)_\mathbb{C}.$$

The classes $[Z(d)]_\mathfrak{m}$ with $d\le 0$ are in the kernel. Applying this map coefficientwise to the generating series $A_\mathfrak{m}$ in Theorem 4.2, we obtain the Gross–Kohnen–Zagier theorem.

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 $S_{3/2}(\omega_L^\vee) \otimes J(X)_{\mathbb{C}}$.

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 $J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}$. Applying it to the generating series $A_{\mathfrak{m}}$, 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 Γ 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 $I \in \mathrm{Iso}(L)$ by

$$F_I^+ = \sum_{j\in\mathbb{Z}} c_{F,I}^+(j) \cdot q_I^j. \tag{5-1}$$

We define the order of F at the cusp c_I by

$$\mathrm{ord}_{c_I}(F) = \min\{j \in \mathbb{Z} : c_{F,I}^+(j) \neq 0\}.$$

Proposition 5.1. *Assume that for all $I \in \mathrm{Iso}(L)$ we have $\mathrm{ord}_{c_I}(F) > -m_I$ and $c_{F,I}^+(0) = 0$.*

(i) *There is a linear map,*

$$\mathrm{tr}_F : J_{\mathfrak{m}}(X) \to \mathbb{C},$$

defined by

$$[D]_{\mathfrak{m}} \mapsto \mathrm{tr}_F(D) := \sum_{a\in \mathrm{supp}(D)\setminus S} n_a \cdot F(a),$$

for divisors $D = \sum_a n_a \cdot (a)$ in $\mathrm{Div}^0(X)$.

(ii) *The map tr_F vanishes on $H_{\mathbb{G}_m,\mathfrak{m}}$ and factors through $J_{\mathfrak{m}}^{\mathrm{add}}(X)$.*

Proof. (i) We have to show that $\mathrm{tr}_F(D) = 0$, for every divisor $D = \mathrm{div}(g) \in P_{\mathfrak{m}}(X)$ given by a rational function $g \in k(X)^\times$ satisfying

$$q_I^{-\mathrm{ord}_{c_I}(g)} g = 1 + O(q_I^{m_I})$$

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

$$\frac{dg}{g} = \mathrm{ord}_{c_I}(g) q_I^{-1} + O(q_I^{m_I - 1}).$$

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 η vanishes, and we have

$$\sum_{a \in X \setminus S} \mathrm{res}_a(\eta) = -\sum_{a \in S} \mathrm{res}_a(\eta).$$

The left-hand side of this equality is given by $\mathrm{tr}_F(D)$, while the right-hand side satisfies

$$\begin{aligned}\sum_{a \in S} \mathrm{res}_a(\eta) &= \sum_{I \in \mathrm{Iso}(L)/\Gamma} \mathrm{res}_{c_I}(\eta) \\ &= \sum_{I \in \mathrm{Iso}(L)/\Gamma} \mathrm{res}_{q_I = 0}\biggl(\bigl(\mathrm{ord}_{c_I}(g) q_I^{-1} + O(q_I^{m_I - 1})\bigr) \sum_{j > -m_I} c_{F,I}^+(j) \cdot q_I^j\biggr) = 0.\end{aligned}$$

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_ε be the manifold with boundary obtained from X by cutting out small oriented discs of radius ε around the points in $\mathrm{supp}(\mathrm{div}(g)) \cup S$. 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_ε, we find that

$$\int_{\partial X_\varepsilon} \eta = \int_{X_\varepsilon} (\bar\partial F) \wedge (\partial \log |g|^2) = -\int_{X_\varepsilon} \partial\bigl((\bar\partial F) \log |g|^2\bigr) = -\int_{\partial X_\varepsilon} (\bar\partial F) \log |g|^2.$$

In the latter integral, the differential $\bar\partial F = \overline{\xi_0(F)} d\bar z$ 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

$$\lim_{\varepsilon \to 0} \int_{\partial X_\varepsilon} \eta = \mathrm{tr}_F(D) + \sum_{I \in \mathrm{Iso}(L)/\Gamma} \mathrm{res}_{c_I}\left(F_I^+ \cdot \frac{dg}{g}\right). \tag{5-2}$$

The vanishing of the second summand on the right-hand side follows as before, proving that $\mathrm{tr}_F(D) = 0$ again.

(ii) Let $u_{D,r}$ be as in (4-1). The same argument shows that $\mathrm{tr}_F(\mathrm{div}(u_{D,r}))$ vanishes and tr_F factors through $J_{\mathfrak{m}}^{\mathrm{add}}(X)$. $\square$

Theorem 5.2. *Assume that* $\mathrm{ord}_{c_I}(F) > -m_I$ *and* $c_{F,I}^{+}(0) = 0$ *for all* $I \in \mathrm{Iso}(L)$. *Then* $\mathrm{tr}_F(A_{\mathfrak{m}})$ *is a weakly holomorphic modular form in* $M^!_{3/2}(\omega_L^\vee)$, *and*

$$\mathrm{tr}_F(A_{\mathfrak{m}})(\varphi) = \sum_{d<0} \mathrm{tr}_F([Z(d,\varphi)]_{\mathfrak{m}}) \cdot q^d + \mathrm{tr}_F([\mathcal{M}_{-1}]_{\mathfrak{m}})\varphi(0) + \sum_{d>0} F(Y(d,\varphi)) \cdot q^d.$$

Moreover, for $d<0$ *the quantity* $\mathrm{tr}_F([Z(d,\varphi)]_{\mathfrak{m}})$ *is given by the finite sum*

$$\mathrm{tr}_F([Z(d,\varphi)]_{\mathfrak{m}}) \\ = -\frac{1}{2} \sum_{I\in\mathrm{Iso}(L)/\Gamma} \; \sum_{\lambda \in L'_{d,I}/\Gamma_I} (\varphi(\lambda)+\varphi(-\lambda)) \cdot \sum_{j\geq 1} e^{2\pi i(\lambda,\ell_I')j} c_{F,I}^{+}(-n_I(\lambda)j).$$

Proof. The modularity of $\mathrm{tr}_F(A_{\mathfrak{m}})$ 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 tr_F that

$$\mathrm{tr}_F([Z(d,\varphi)]_{\mathfrak{m}}) = F(Y(d,\varphi)).$$

If $d<0$ and $d \in -2\,\mathrm{disc}(L)(\mathbb{Q}^\times)^2$, we obtain by the definition of the class $[Z(d)]_{\mathfrak{m}}$ that

$$\begin{aligned} \mathrm{tr}_F([Z(d,\varphi)]_{\mathfrak{m}}) &= \sum_{\lambda\in L'_d/\Gamma} \frac{1}{2n_I(\lambda)} (\varphi(\lambda)+\varphi(-\lambda)) \cdot \mathrm{tr}_F\big(\mathrm{div}(h_\lambda^{-1})\big) \\ &= -\sum_{I\in\mathrm{Iso}(L)/\Gamma} \; \sum_{\lambda\in L'_{d,I}/\Gamma_I} \frac{1}{2n_I(\lambda)} (\varphi(\lambda)+\varphi(-\lambda)) \cdot F(\mathrm{div}(h_\lambda)). \end{aligned}$$

Arguing as in the proof of Proposition 5.1, in particular (5-2), we find for $\lambda \in L'_{d,I}$ that

$$\begin{aligned} F(\mathrm{div}(h_\lambda)) &= -\sum_{J\in\mathrm{Iso}(L)/\Gamma} \mathrm{res}_{c_J}\Big(F_J^{+} \cdot \frac{dh_\lambda}{h_\lambda}\Big) \\ &= \mathrm{res}_{c_I}\left(F_I^{+} \cdot \frac{n_I(\lambda)\cdot e^{2\pi i(\lambda,\ell_I')} q_I^{n_I(\lambda)-1} + O(q_I^{m_I-1})}{1-e^{2\pi i(\lambda,\ell_I')} q_I^{n_I(\lambda)}}\right) \\ &= n_I(\lambda) \sum_{j\geq 1} e^{2\pi i(\lambda,\ell_I')j} c_{F,I}^{+}(-n_I(\lambda)j). \end{aligned} \tag{5-3}$$

Inserting this into the previous equation, we obtain the assertion. $\square$

Remark 5.3. The constant term $\mathrm{tr}_F([Z(0,\varphi)]_{\mathfrak{m}})$ 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

$$L = \left\{ \begin{pmatrix} b & a/M \\ c & -b \end{pmatrix} : a, b, c \in \mathbb{Z} \right\}, \tag{5-4}$$

with the quadratic form $Q(X) = M \det(X)$. Then $L'/L \cong \mathbb{Z}/2M\mathbb{Z}$ and $\mathrm{SO}^+(L)$ is isomorphic to the extension $\Gamma_0^*(M)$ of $\Gamma_0(M)$ by the Atkin–Lehner involutions. The discriminant kernel subgroup Γ is isomorphic to $\Gamma_0(M)$, and the modular curve X_Γ 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 ∞; see, e.g., [Bruinier and Ono 2010, Section 2.4].

The group $\Gamma_0^*(M)$ acts transitively on $\mathrm{Iso}(L)$, and the orbits are represented by the lines $I_D = W_D.I_0$ for the positive divisors $D \mid M$. Here $W_D \in \mathrm{PGL}_2^+(\mathbb{Q})$ denotes the Atkin–Lehner involution with index D. In particular, the set S of cusps of $X_0(M)$ is in bijection with the set of positive divisors of M. If $I \in \mathrm{Iso}(L)$, we write D_I for the unique positive divisor of M such that I is equivalent to $W_{D_I}.I_0$ under Γ. Let $F \in H_0^+(\Gamma)$ 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 $F \mid W_D$.

Proposition 5.4. *Assume that for all $I \in \mathrm{Iso}(L)$ we have $\mathrm{ord}_{c_I}(F) > -m_I$ and $c_{F,I}^+(0) = 0$. The constant term of the generating series $\mathrm{tr}_F(A_\mathfrak{m})$ is given by*

$$\mathrm{tr}_F([\mathcal{M}_{-1}]_\mathfrak{m}) = 2 \sum_{D \mid M} \sum_{j \geq 1} c_{F,I_D}^+(-j) \cdot D \cdot \sigma_1(j/D).$$

Remark 5.5. As shown in [Bruinier and Funke 2006, Remark 4.9], the right-hand side above is also equal to $-\frac{1}{4\pi} \int_{\Gamma_0(M)\backslash\mathbb{H}}^{\mathrm{reg}} F \, d\mu$. 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

$$s = \Delta \cdot \prod_{I \in \mathrm{Iso}(L)/\Gamma} \prod_{\substack{\lambda \in (L' \cap I^\perp)/I \\ m_I > n_I(\lambda) > 0}} h_{D_I \lambda}^{-24}$$

has the expansion

$$s = D^{-6} q_{I_D}^{D} \cdot \left(1 + O\left(q_{I_D}^{Dm_I}\right)\right)$$

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 $\mathcal{M}_{12}$ satisfies the local conditions (2-6) at all cusps. Therefore, in view of (2-7) and Proposition 5.1 (ii), we have

$$\begin{aligned} \mathrm{tr}_F([\mathcal{M}_{12}]_{\mathfrak{m}}) &= \mathrm{tr}_F([\mathrm{div}(s \cdot u_{0,r}) - \deg(\mathcal{M}_{12}) \cdot (c_{I_0})]_{\mathfrak{m}}) \\ &= -24 \sum_{I \in \mathrm{Iso}(L)/\Gamma} \sum_{\substack{\lambda \in (L' \cap I^\perp)/I \\ m_I > n_I(\lambda) > 0}} F(\mathrm{div}(h_{D_I \lambda})). \end{aligned}$$

Using formula (5-3), we get

$$\begin{aligned} \mathrm{tr}_F([\mathcal{M}_{12}]_{\mathfrak{m}}) &= -24 \sum_{I \in \mathrm{Iso}(L)/\Gamma} \sum_{n=1}^{m_I - 1} D_I n \sum_{j \geq 1} c^+_{F,I}(-D_I n j) \\ &= -24 \sum_{D\,|\,M} \sum_{j \geq 1} c^+_{F,I_D}(-j) \cdot D \cdot \sigma_1(j/D). \end{aligned}$$

This concludes the proof of the proposition. □

We now explain how to obtain a scalar-valued generating series from $\mathrm{tr}_F(A_{\mathfrak{m}})$. By means of the canonical pairing $(S_L, S_L^\vee) \to \mathbb{C}$, we define a map

$$S_L^\vee \to \mathbb{C}, \quad u \mapsto (\chi_1, u),$$

given by the pairing with the constant function χ_1 with value 1. It induces a map from $S_L^\vee$-valued to scalar-valued modular forms,

$$M^!_{3/2}(\omega_L^\vee) \to M^!_{3/2}(\Gamma_0(4M)), \quad f(\tau) \mapsto f^{\mathrm{scal}}(\tau) := f(\chi_1)(4M\tau),$$

see [Eichler and Zagier 1985, §5]. The image lies in the Kohnen plus-space. Applying this map to the generating series $A_{\mathfrak{m}}$ 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 $I \in \mathrm{Iso}(L)$ we have $\mathrm{ord}_{c_I}(F) > -m_I$ and $c^+_{F,I}(0) = 0$. Then $\mathrm{tr}_F(A^{\mathrm{scal}}_{\mathfrak{m}}) \in M^!_{3/2}(\Gamma_0(4M))$, and*

$$\begin{aligned} \mathrm{tr}_F(A^{\mathrm{scal}}_{\mathfrak{m}}) = &-\sum_{D\,|\,M} \sum_{b \geq 1} \sum_{n \geq 1} c^+_{F,I_D}(-bn) \cdot b \cdot q^{-b^2} \\ &+ 2 \sum_{D\,|\,M} \sum_{n \geq 1} c^+_{F,I_D}(-n) \cdot D \cdot \sigma_1(n/D) + \sum_{d \in \mathbb{Z}_{>0}} F(Y(d/4M, \chi_1)) \cdot q^d. \end{aligned}$$

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 $\mathrm{PSL}_2(\mathbb{Z})$ 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* $\frac{3}{2}$ *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 (c_{I_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 $\mathrm{Cl}_{\mathfrak{m}}(X)$ be the generalized class group of X with respect to the modulus $\mathfrak{m}$, which we define as the quotient of the group of divisors on X defined over k modulo the subgroup $P_{\mathfrak{m}}(X)$. Moreover, in analogy with (4-2) we put

$$\mathrm{Cl}_{\mathfrak{m}}^{\mathrm{add}}(X) = \mathrm{Cl}_{\mathfrak{m}}(X)/H_{\mathbb{G}_m,\mathfrak{m}}. \tag{6-1}$$

If $d > 0$, we write $[Y(d, \varphi)]_{\mathfrak{m}}$ for the class of the divisor $Y(d, \varphi)$ in $\mathrm{Cl}_{\mathfrak{m}}(X)$. For $d = 0$ we put $[Y(0, \varphi)]_{\mathfrak{m}} = \varphi(0)[\mathcal{M}_{-1}]_{\mathfrak{m}}$, where the class in $\mathrm{Cl}_{\mathfrak{m}}(X)$ of a line bundle $\mathcal{L}$ is defined as in (2-7) but without the summand $\deg(\mathcal{L}) \cdot (s_0)$. Finally, for $d < 0$ we let $[Y(d, \varphi)]_{\mathfrak{m}} = [Z(d, \varphi)]_{\mathfrak{m}}$.

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 $\mathrm{Mp}_2(\mathbb{Z})$ with representation $\omega_L^\vee$ and generalizes Zagier's nonholomorphic Eisenstein series [1975]. Its Fourier expansion decomposes as

$$E_{\frac{3}{2},L}(\tau) = E^+_{\frac{3}{2},L}(\tau) + E^-_{\frac{3}{2},L}(\tau),$$

where the holomorphic part is the generating series of the degrees of Heegner divisors,

$$E^+_{\frac{3}{2},L}(\tau) = \sum_{d \geq 0} \deg(Y(d)) \cdot q^d,$$

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*

$$\tilde{A}_{\mathfrak{m}}(\tau) = \sum_{d \in \frac{1}{N}\mathbb{Z}} [Y(d)]_{\mathfrak{m}} \cdot q^d + E^-_{\frac{3}{2},L} \cdot (c_{I_0})$$

is a nonholomorphic modular form of weight $\frac{3}{2}$ for $\mathrm{Mp}_2(\mathbb{Z})$ with representation $\omega_L^\vee$ with values in $\mathrm{Cl}^{\mathrm{add}}_{\mathfrak{m}}(X)$. Moreover, we have

$$A_{\mathfrak{m}} = \tilde{A}_{\mathfrak{m}} - E_{\frac{3}{2},L} \cdot (c_{I_0}).$$

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 $\Delta \equiv r^2 \bmod 4M$, we can define a generalized genus character χ_Δ on L' as in [Gross et al. 1987, Section I.2] and [Bruinier and Ono 2010, Section 4] let

$$\chi_\Delta(\lambda) = \begin{cases} \left(\frac{\Delta}{n}\right) & \text{if } \Delta \mid b^2 - 4Mac \text{ and } (b^2-4Mac)/\Delta \text{ is a square modulo } 4M \\ & \qquad \text{and } \gcd(a,b,c,\Delta) = 1, \\ 0 & \text{otherwise,} \end{cases}$$

with $\lambda = \left(\begin{smallmatrix} b/2M & -a/M \\ c & -b/2M \end{smallmatrix}\right) \in L'$ and $n \in \mathbb{Z}$ any integer prime to Δ represented by one of the quadratic forms $[M_1 a, b, M_2 c]$ with $M_1 M_2 = M$. Note that χ_Δ is invariant under $\mathrm{SO}^+(L)$.

If $\lambda \in L'$ with $Q(\lambda) \in -4M(\mathbb{Q}^\times)^2$, let I be the isotropic line associated with λ, and let $h_{\Delta,\lambda} \in \mathbb{Q}(\sqrt{\Delta})(X)^\times$ be a rational function with the following expansions:

$$h_{\Delta,\lambda} = \begin{cases} \prod_{b \in \mathbb{Z}/\Delta\mathbb{Z}} \left(1 - e^{2\pi i b/\Delta} q_I^{n_I(\lambda)}\right)^{\left(\frac{\Delta}{b}\right)} + O(q_I^{m_I}) & \text{at the cusp } c_I, \\ 1 + O(q_J^{m_J}) & \text{at all other cusps } c_J. \end{cases}$$

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 $Z_{\Delta,r}(d,\varphi) \in \mathrm{Div}^0(X)_{\mathbb{C}}$ by

$$Z_{\Delta,r}(d,\varphi) := \begin{cases} \displaystyle\sum_{\lambda \in L'_{d|\Delta|}/\Gamma} \frac{\chi_\Delta(\lambda)\varphi(r^{-1}\lambda)}{2|\Gamma_\lambda|} \cdot (z_\lambda) & \text{if } d > 0, \\ \displaystyle\sum_{\lambda \in L'_{d/|\Delta|}/\Gamma} \frac{\varphi(r\lambda) + \mathrm{sgn}(\Delta)\varphi(-r\lambda)}{2n_I(\lambda)} \operatorname{div}(h^{-1}_{\Delta,\lambda}) & \text{if } d \in -\dfrac{|\Delta|}{4M}(\mathbb{Z}_{>0})^2, \\ 0 & \text{otherwise.} \end{cases}$$

All these divisors are defined over $\mathbb{Q}(\sqrt{\Delta})$ and have coefficients in the field of definition of φ. We write $[Z_{\Delta,r}(d)]_{\mathfrak{m}} \in S_L^\vee \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)$ for the element that sends φ to $[Z_{\Delta,r}(d,\varphi)]_{\mathfrak{m}} \in J_{\mathfrak{m}}^{\mathrm{add}}(X)$. Define the representation $\tilde{\omega}_L$ to be ω_L if $\Delta > 0$ and $\bar{\omega}_L$ if $\Delta < 0$. Then we have the following abstract modularity result.

Theorem 6.2. *The generating series*

$$A_{\Delta,r,\mathfrak{m}}(\tau) := \sum_{d \in \frac{1}{4M}\mathbb{Z}} [Z_{\Delta,r}(d)]_{\mathfrak{m}} \cdot q^d \in S_L^\vee((q)) \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}$$

is the q-expansion of a weakly holomorphic modular form in $M^!_{3/2}(\tilde{\omega}_L) \otimes J_{\mathfrak{m}}^{\mathrm{add}}(X)_{\mathbb{C}}$.

This result comes out of calculating the effect of the intertwining operator in [Alfes and Ehlen 2013, Section 3] applied to the generating series $A_{\mathfrak{m}}(\tau)$ 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 $\operatorname{sgn}(\Delta)$, 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 Γ be the discriminant kernel subgroup of $\mathrm{SO}^+(L)$. Denote by X_Γ a (suitable) toroidal compactification of the connected Shimura variety Y_Γ associated to Γ. It would be interesting to define a generalized divisor class group as the group of divisors on X_Γ modulo divisors of rational functions that satisfy certain growth conditions along the boundary of X_Γ. 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_Γ 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_Γ 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 $\mathfrak{m} = m \cdot (s)$. 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 (∞, ∞) we write $f = O(q^m)$ if in the Taylor expansion of f at (∞, ∞) only terms of total degree at least m occur.

Let $\mathrm{Div}_{\mathfrak{m}}(X_\Gamma)$ be the free abelian group generated by pairs (D, g_D), where D is a prime Weil divisor on X_Γ 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 $P_{\mathfrak{m}}(X_\Gamma)$ be the subgroup of pairs $(D, g_D) \in \mathrm{Div}_{\mathfrak{m}}(X_\Gamma)$ for which $D = \mathrm{div}(f)$ 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

$$\mathrm{Cl}_{\mathfrak{m}}(X_\Gamma) = \mathrm{Div}_{\mathfrak{m}}(X_\Gamma) / P_{\mathfrak{m}}(X_\Gamma).$$

It would be interesting to define suitable classes of special divisors in $\mathrm{Cl}_{\mathfrak{m}}(X_\Gamma)$ 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

[Alfes and Ehlen 2013] C. Alfes and S. Ehlen, "Twisted traces of CM values of weak Maass forms", *J. Number Theory* **133**:6 (2013), 1827–1845. MR 3027941

[Borcherds 1998] R. E. Borcherds, "Automorphic forms with singularities on Grassmannians", *Invent. Math.* **132**:3 (1998), 491–562. MR 1625724

[Borcherds 1999] R. E. Borcherds, "The Gross–Kohnen–Zagier theorem in higher dimensions", *Duke Math. J.* **97**:2 (1999), 219–233. MR 1682249

[Bringmann et al. 2005] K. Bringmann, K. Ono, and J. Rouse, "Traces of singular moduli on Hilbert modular surfaces", *Int. Math. Res. Not.* **2005**:47 (2005), 2891–2912. MR 2192218

[Bruinier 2002] J. H. Bruinier, *Borcherds products on $O(2, l)$ and Chern classes of Heegner divisors*, Lecture Notes in Mathematics **1780**, Springer, Berlin, 2002. MR 1903920

[Bruinier and Funke 2004] J. H. Bruinier and J. Funke, "On two geometric theta lifts", *Duke Math. J.* **125**:1 (2004), 45–90. MR 2097357

[Bruinier and Funke 2006] J. H. Bruinier and J. Funke, "Traces of CM values of modular functions", *J. Reine Angew. Math.* **594** (2006), 1–33. MR 2248151

[Bruinier and Ono 2010] J. Bruinier and K. Ono, "Heegner divisors, L-functions and harmonic weak Maass forms", *Ann. of Math.* (2) **172**:3 (2010), 2135–2181. MR 2726107

[Duke and Jenkins 2008] W. Duke and P. Jenkins, "Integral traces of singular values of weak Maass forms", *Algebra Number Theory* **2**:5 (2008), 573–593. MR 2429454

[Eichler and Zagier 1985] M. Eichler and D. Zagier, *The theory of Jacobi forms*, Progress in Mathematics **55**, Birkhäuser, Boston, 1985. MR 781735

[Funke 2002] J. Funke, "Heegner divisors and nonholomorphic modular forms", *Compositio Math.* **133**:3 (2002), 289–321. MR 1930980

[Gross 2012] B. H. Gross, "The classes of singular moduli in the generalized Jacobian", pp. 137–141 in *Geometry and arithmetic*, edited by C. Faber et al., EMS Ser. Congr. Rep., Eur. Math. Soc., Zürich, 2012. MR 2987658

[Gross et al. 1987] B. Gross, W. Kohnen, and D. Zagier, "Heegner points and derivatives of L-series, II", *Math. Ann.* **278**:1-4 (1987), 497–562. MR 909238

[Kim 2004] C. H. Kim, "Borcherds products associated with certain Thompson series", *Compos. Math.* **140**:3 (2004), 541–551. MR 2041767

[Kudla 1997] S. S. Kudla, "Algebraic cycles on Shimura varieties of orthogonal type", *Duke Math. J.* **86**:1 (1997), 39–78. MR 1427845

[Kudla 2003] S. S. Kudla, "Integrals of Borcherds forms", *Compositio Math.* **137**:3 (2003), 293–349. MR 1988501

[Kudla 2004] S. S. Kudla, "Special cycles and derivatives of Eisenstein series", pp. 243–270 in *Heegner points and Rankin L-series*, edited by Cambridge, Math. Sci. Res. Inst. Publ. **49**, Cambridge Univ. Press, 2004. MR 2083214

[Kudla 2014] S. Kudla, "Another product for a Borcherds form", preprint, 2014. arXiv 1402.0443

[Liu 2011] Y. Liu, "Arithmetic theta lifting and L-derivatives for unitary groups, II", *Algebra Number Theory* **5**:7 (2011), 923–1000. MR 2928564

[McGraw 2003] W. J. McGraw, "The rationality of vector valued modular forms associated with the Weil representation", *Math. Ann.* **326**:1 (2003), 105–122. MR 1981614

[Serre 1988] J.-P. Serre, *Algebraic groups and class fields*, Graduate Texts in Mathematics **117**, Springer, New York, 1988. MR 918564

[Zagier 1975] D. Zagier, "Nombres de classes et formes modulaires de poids $3/2$", *C. R. Acad. Sci. Paris Sér. A-B* **281**:21 (1975), Ai, A883–A886. MR 0429750

[Zagier 2002] D. Zagier, "Traces of singular moduli", pp. 211–244 in *Motives, polylogarithms and Hodge theory, I* (Irvine, CA, 1998), edited by F. Bogomolov and L. Katzarkov, Int. Press Lect. Ser. **3**, Int. Press, Somerville, MA, 2002. MR 1977587

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1301

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 $\mathrm{GL}_2(\mathbb{Q}_2)$. 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 ϖ. 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*

$$\rho : G_{F,S} \to \mathrm{GL}_2(\mathcal{O})$$

be a continuous representation of the Galois group of the maximal extension of F unramified outside S. Suppose:

(i) $\bar{\rho} : G_{F,S} \xrightarrow{\rho} \mathrm{GL}_2(\mathcal{O}) \to \mathrm{GL}_2(k)$ *is modular with nonsolvable image.*

(ii) *If $v \mid 2$ then $\rho|_{G_{F_v}}$ is potentially semistable with distinct Hodge–Tate weights.*

(iii) $\det \rho$ *is totally odd.*

(iv) *If $v \mid 2$ then $\bar{\rho}|_{G_{F_v}} \not\cong \left(\begin{smallmatrix} \chi & * \\ 0 & \chi \end{smallmatrix}\right)$ for any character $\chi : G_{F_v} \to k^\times$.*

Then ρ 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 $G := \mathrm{GL}_2(\mathbb{Q}_p)$ 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 $\mathrm{Mod}^{\mathrm{sm}}_G(\mathcal{O})$ 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 $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}^{\mathrm{sm}}_G(\mathcal{O})$, consisting of representations on which the center of G acts by the character ψ and which are equal to the union of their admissible subrepresentations. The categories $\mathrm{Mod}^{\mathrm{sm}}_G(\mathcal{O})$ and $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ 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 $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ 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 $\mathrm{Irr}^{\mathrm{adm}}_G$ be the set of irreducible representations in $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$. We define an equivalence relation $\sim$ on $\mathrm{Irr}^{\mathrm{adm}}_G$ by writing $\pi \sim \tau$ if there exists a sequence $\pi = \pi_1, \pi_2, \ldots, \pi_n = \tau$ in $\mathrm{Irr}^{\mathrm{adm}}_G$ such that for each i one of the following holds:

(1) $\pi_i \cong \pi_{i+1}$; (2) $\mathrm{Ext}^1_G(\pi_i, \pi_{i+1}) \neq 0$; (3) $\mathrm{Ext}^1_G(\pi_{i+1}, \pi_i) \neq 0$. We have a canonical decomposition

$$\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O}) \cong \prod_{\mathfrak{B} \in \mathrm{Irr}^{\mathrm{adm}}_G/\sim} \mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}], \tag{1}$$

where $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ is the full subcategory of $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ consisting of representations with all irreducible subquotients in $\mathfrak{B}$. A block is an equivalence class of $\sim$.

For a block $\mathfrak{B}$ let $\pi_{\mathfrak{B}} = \bigoplus_{\pi \in \mathfrak{B}} \pi$, let $\pi_{\mathfrak{B}} \hookrightarrow J_{\mathfrak{B}}$ be an injective envelope of $\pi_{\mathfrak{B}}$ and let $E_{\mathfrak{B}} := \mathrm{End}_G(J_{\mathfrak{B}})$. Then $J_{\mathfrak{B}}$ is an injective generator for $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$, $E_{\mathfrak{B}}$ is a pseudocompact ring and the functor $\kappa \mapsto \mathrm{Hom}_G(\kappa, J_{\mathfrak{B}})$ induces an antiequivalence of categories between $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ and the category of right pseudocompact $E_{\mathfrak{B}}$-modules. The inverse functor is given by $\mathrm{m} \mapsto \left(\mathrm{m} \widehat{\otimes}_{E_{\mathfrak{B}}} J_{\mathfrak{B}}^{\vee}\right)^{\vee}$, where $\vee$ denotes the Pontryagin dual; see [Gabriel 1962, Chapitre IV, §4]. The main result of [Paškūnas 2013] computes the rings $E_{\mathfrak{B}}$ for each block $\mathfrak{B}$ and describes the Galois representation of $G_{\mathbb{Q}_p}$ obtained by applying the Colmez's functor to $J_{\mathfrak{B}}$ under the assumption $p \geq 5$ or $p \geq 3$, depending on the block $\mathfrak{B}$.

If $\pi \in \mathrm{Irr}^{\mathrm{adm}}_G$ then one may show that, after extending scalars, π 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) $\mathfrak{B} = \{\pi\}$ with π supersingular;

(ii) $\mathfrak{B} = \left\{\left(\mathrm{Ind}^G_B \chi_1 \otimes \chi_2\omega^{-1}\right)_{\mathrm{sm}}, \left(\mathrm{Ind}^G_B \chi_2 \otimes \chi_1\omega^{-1}\right)_{\mathrm{sm}}\right\}$ with $\chi_2\chi_1^{-1} \neq \omega^{\pm 1}, \mathbf{1}$;

(iii) $p > 2$ and $\mathfrak{B} = \left\{\left(\mathrm{Ind}^G_B \chi \otimes \chi\omega^{-1}\right)_{\mathrm{sm}}\right\}$;

(iv) $p = 2$ and $\mathfrak{B} = \{\mathbf{1}, \mathrm{Sp}\} \otimes \chi \circ \det$;

(v) $p \geq 5$ and $\mathfrak{B} = \left\{\mathbf{1}, \mathrm{Sp}, \left(\mathrm{Ind}^G_B \omega \otimes \omega^{-1}\right)_{\mathrm{sm}}\right\} \otimes \chi \circ \det$;

(vi) $p = 3$ and $\mathfrak{B} = \{\mathbf{1}, \mathrm{Sp}, \omega \circ \det, \mathrm{Sp} \otimes \omega \circ \det\} \otimes \chi \circ \det$;

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 $\omega(x) = x|x| \pmod{\varpi}$ and we view $\chi_1 \otimes \chi_2$ as a character of the subgroup of upper-triangular matrices B in G which sends $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ to $\chi_1(a)\chi_2(d)$. An absolutely irreducible representation π is supersingular if it is not a subquotient of a principal series representation (they have been classified by Breuil [2003a]) and Sp denotes the Steinberg representation.

To each block above one may attach a semisimple 2-dimensional k-representation $\bar{\rho}^{\mathrm{ss}}$ of $G_{\mathbb{Q}_p}$: in case (i) $\bar{\rho}^{\mathrm{ss}}$ is absolutely irreducible, and such that Colmez's functor $\boldsymbol{V}$ (see Section 2B1) maps π to $\bar{\rho}^{\mathrm{ss}}$; in case (ii) $\bar{\rho}^{\mathrm{ss}} = \chi_1 \oplus \chi_2$; in cases (iii) and (iv) $\bar{\rho}^{\mathrm{ss}} = \chi \oplus \chi$; in cases (v) and (vi) $\bar{\rho}^{\mathrm{ss}} = \chi \oplus \chi\omega$, 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 $\bar{\rho}^{\mathrm{ss}}$ is equal to $\psi\varepsilon$ modulo ϖ, where ε is the p-adic cyclotomic character and ω is its reduction modulo ϖ.

Theorem 1.2. *If $\mathfrak{B} = \{\pi\}$ with π supersingular (so that $\bar{\rho}^{\mathrm{ss}}$ is irreducible) then $E_{\mathfrak{B}}$ is naturally isomorphic to the quotient of the universal deformation ring of $\bar{\rho}^{\mathrm{ss}}$ parametrizing deformations with determinant $\psi\varepsilon$, and $\boldsymbol{V}(J_{\mathfrak{B}})^{\vee}(\psi\varepsilon)$ is a tautological deformation of $\bar{\rho}^{\mathrm{ss}}$ to $E_{\mathfrak{B}}$.*

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 $(\operatorname{tr}\bar{\rho}^{\mathrm{ss}}, \det\bar{\rho}^{\mathrm{ss}})$, and let $R^{\mathrm{ps},\psi}$ be the quotient of R^{ps} parametrizing those which have determinant $\psi\varepsilon$.

Theorem 1.3. *Assume that the block $\mathfrak{B}$ is given by* (i) *or* (ii) *above. Then the center of the category $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ is naturally isomorphic to $R^{\mathrm{ps},\psi}$.*

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 $\mathfrak{B}=\{\pi\}$ with π 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 $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ of admissible unitary L-Banach space representations of G on which the center of G acts by ψ; see Section 2B4. An admissible unitary L-Banach space representation Π lies in $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$ if and only if all the irreducible subquotients of the reduction modulo ϖ of a unit ball in Π modulo ϖ lie in $\mathfrak{B}$. An irreducible Π 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 $\mathfrak{B}$ is given by* (i) *or* (ii) *above. Colmez's Montreal functor $\Pi \mapsto \check{\boldsymbol{V}}(\Pi)$ induces a bijection between the isomorphism classes of*

- *absolutely irreducible nonordinary $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$;*
- *absolutely irreducible $\tilde{\rho} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(L)$ such that $\det\tilde{\rho} = \psi\varepsilon$ and the semisimplification of the reduction modulo ϖ of a $G_{\mathbb{Q}_p}$-invariant $\mathcal{O}$-lattice in $\tilde{\rho}$ is isomorphic to $\bar{\rho}^{\mathrm{ss}}$.*

A stronger result, avoiding the assumption on $\mathfrak{B}$, 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 $\mathrm{GL}_2(\mathbb{Q}_p)$ out of 2-dimensional representations of $G_{\mathbb{Q}_p}$ via the theory of (φ, Γ)-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 $\mathfrak{B} = \{\pi\}$ with π supersingular. Let $\bar{\rho} = V(\pi)$, let $R_{\bar{\rho}}$ be the universal deformation ring of $\bar{\rho}$ and let $R^{\psi}_{\bar{\rho}}$ be the quotient of $R_{\bar{\rho}}$ parametrizing deformations with determinant $\psi\varepsilon$. We follow the strategy outlined in [Paškūnas 2013, §5.8]. We show that $J^{\vee}_{\mathfrak{B}}$ is the universal deformation of $\pi^{\vee}$ and $E_{\mathfrak{B}}$ 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 $E_{\mathfrak{B}}$ is commutative. To apply this criterion, one needs the ring $R^{\psi}_{\bar{\rho}}$ to be formally smooth and to control the image of some Ext^1-group in some 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 $\bar{\rho} \ncong \bar{\rho} \otimes \omega$, 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 π occurs in the reduction modulo ϖ with multiplicity one. We deduce from [Colmez et al. 2014, Corollary 2.22] that the ring $E_{\mathfrak{B}}$ is commutative. The argument of Kisin [2010] shows that $V(J_{\mathfrak{B}})^{\vee}(\psi\varepsilon)$ is a deformation of $\bar{\rho}$ to $E_{\mathfrak{B}}$ and we have surjections $R_{\bar{\rho}} \twoheadrightarrow E_{\mathfrak{B}} \twoheadrightarrow R^{\psi}_{\bar{\rho}}$.

To prove Theorem 1.2 we have to show that the surjection $\varphi : E_{\mathfrak{B}} \twoheadrightarrow R^{\psi}_{\bar{\rho}}$ 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 $R^{\psi}_{\bar{\rho}}$ 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 φ 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 $E_{\mathfrak{B}}$ and $R^{\psi}_{\bar{\rho}}$, and it is not at all clear that $E_{\mathfrak{B}}$ is reduced. However, by using techniques of [Paškūnas 2015b] we can show that certain quotients $E_{\mathfrak{B}}/\mathfrak{a}$ are reduced and identify them with crystabeline deformation rings of $\bar{\rho}$ via φ. 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 $E_{\mathfrak{B}}$ 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 $\det \check{V}(\Pi) = \psi\varepsilon$ for all $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$. 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 $\bar{\rho}^{\mathrm{ss}} = \chi_1 \oplus \chi_2$, with $\chi_1\chi_2^{-1} \neq \mathbf{1}, \omega^{\pm 1}$, then we prove the Breuil–Mézard conjecture for both nonsplit extensions $\left(\begin{smallmatrix} \chi_1 & * \\ 0 & \chi_2 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \chi_1 & 0 \\ * & \chi_2 \end{smallmatrix}\right)$ 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 $\rho|_{G_{F_v}}$ 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 ρ 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 ρ_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

1974; Saito 2009] that the Weil–Deligne representation associated to $\rho|_{G_{F_v}}$ is pure. Kisin shows that this implies that the point on the generic fiber of the potentially semistable deformation ring, defined by $\rho_f|_{G_{F_v}}$, 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 ρ_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 $\rho|_{G_{F_v}}$ for $v \mid 2$. Kisin assumes that $\rho|_{G_{F_v}}$ 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 $\rho|_{G_{F_v}}$ is ordinary. Thorne removes this last assumption, but requires instead that $\bar\rho|_{G_{F_v}}$ 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 $\mathrm{GL}_2(\mathbb{Q}_p)$.

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 $\mathrm{Mod}^{\mathrm{pro}}_K(\mathcal{O})$ 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 $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}^{\mathrm{pro}}_K(\mathcal{O})$ such that $M \in \mathrm{Mod}^{\mathrm{pro}}_K(\mathcal{O})$ lies in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$ if and only if $Z(K)$ acts on M via ψ^{-1}. Let $\{V_i\}_{i\in I}$ be a family of continuous representations of K on finite-dimensional L-vector spaces, and let $M \in \mathrm{Mod}^{\mathrm{pro}}_K(\mathcal{O})$.

Definition 2.1. We say that $\{V_i\}_{i\in I}$ *captures* M if the smallest quotient $M \twoheadrightarrow Q$ such that $\mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(Q, V_i^*) \cong \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(M, V_i^*)$ for all $i \in I$ is equal to M.

We let $c := \left(\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right)$ and note that the center of $\mathrm{SL}_2(\mathbb{Z}_p)$ is equal to $\{1, c\}$.

Lemma 2.2. *If $K = \mathrm{SL}_2(\mathbb{Z}_p)$ 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 $\mathcal{O}[K/K_n]$ is a free $\mathcal{O}[Z]$-module, where Z is the center of K. This implies that $\mathcal{O}[K/K_n]/(c \pm 1)$ is a free $\mathcal{O}$-module and by passing to the limit we obtain the assertion. □

Lemma 2.3. *Let $K = \mathrm{SL}_2(\mathbb{Z}_p)$, let Z be the center of K and let $\{V_i\}_{i\in I}$ be a family which captures $\mathcal{O}[\![K]\!]$ 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 $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$. If $\psi(c) = -1$ then I^- captures every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$.*

Proof. If $M \in \mathrm{Mod}^{\mathrm{pro}}_K(\mathcal{O})$ is $\mathcal{O}$-torsion-free then I captures M if and only if the image of the evaluation map $\bigoplus_{i\in I} V_i \otimes \mathrm{Hom}_K(V_i, \Pi) \to \Pi$ is dense, where $\Pi = \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}}(M, L)$ is the Banach space representation of K with the topology induced by the supremum norm [Colmez et al. 2014, Lemma 2.10]. Let $\Pi = \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}}(\mathcal{O}[\![K]\!], L)$ and $\Pi^\pm := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}}(\mathcal{O}[\![K]\!]/(c \pm 1), L)$. Since $\Pi = \Pi^+ \oplus \Pi^-$, and $\{V_i\}$ captures $\mathcal{O}[\![K]\!]$, we deduce that the image of the evaluation map $\bigoplus_{i\in I} V_i \otimes \mathrm{Hom}_K(V_i, \Pi^\pm) \to \Pi^\pm$ is dense. If $i \in I^+$ then c acts trivially on V_i and so $\mathrm{Hom}_K(V_i, \Pi^-) = 0$. This implies the image of $\bigoplus_{i\in I^+} V_i \otimes \mathrm{Hom}_K(V_i, \Pi^+) \to \Pi^+$ is dense. Using Lemma 2.2 we deduce that I^+ captures $\mathcal{O}[\![K]\!]/(c-1)$. A similar argument shows that I^- captures $\mathcal{O}[\![K]\!]/(c+1)$. Every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$ can be realized as a direct summand of a product of some copies of $\mathcal{O}[\![K]\!]/(c - \psi(c))$, which implies the assertion; see the proof of [Colmez et al. 2014, Lemma 2.11]. □

Lemma 2.4. *Let $K = \mathrm{SL}_2(\mathbb{Z}_p)$, 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 ψ_V. If $\psi(c) = \psi_V(c)$ then $\{V \otimes \mathrm{Sym}^{2a} L^2\}_{a\in\mathbb{N}}$ captures every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$; if $\psi(c) = -\psi_V(c)$ then $\{V \otimes \mathrm{Sym}^{2a+1} L^2\}_{a\in\mathbb{N}}$ captures every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$.*

Proof. Proposition 2.12 in [Colmez et al. 2014] implies that $\{\mathrm{Sym}^a L^2\}_{a\in\mathbb{N}}$ captures $\mathcal{O}[\![K]\!]$. We leave it as an exercise for the reader to check that this implies that $\{V \otimes \mathrm{Sym}^a L^2\}_{a\in\mathbb{N}}$ also captures $\mathcal{O}[\![K]\!]$. The assertion follows from Lemma 2.3. □

Lemma 2.5. *Let $M \in \mathrm{Mod}^{\mathrm{pro}}_{\mathrm{GL}_2(\mathbb{Z}_p),\psi}(\mathcal{O})$ and let V be a continuous representation of K on a finite-dimensional L-vector space with a central character ψ. Then*

$$\bigcap_{\phi} \mathrm{Ker}\,\phi = \bigcap_{\xi,\eta} \mathrm{Ker}\,\xi,$$

where the first intersection is taken over all $\phi \in \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![\mathrm{SL}_2(\mathbb{Z}_p)]\!]}(M, V^*)$ and the second intersection is taken over all characters $\eta : \mathbb{Z}_p^\times \to L^\times$ with $\eta^2 = 1$ and all $\xi \in \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![\mathrm{GL}_2(\mathbb{Z}_p)]\!]}(M, (V \otimes \eta \circ \det)^*)$.

Proof. Let Z be the center of $\mathrm{GL}_2(\mathbb{Z}_p)$. The determinant induces the isomorphism $\mathrm{GL}_2(\mathbb{Z}_p)/Z\,\mathrm{SL}_2(\mathbb{Z}_p) \cong \mathbb{Z}_p^\times/(\mathbb{Z}_p^\times)^2$, 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, $\mathrm{Ind}_{Z\,\mathrm{SL}_2(\mathbb{Z}_p)}^{\mathrm{GL}_2(\mathbb{Z}_p)} \mathbf{1} \cong \bigoplus \eta \circ \det$, where the sum is taken over all characters η with $\eta^2 = 1$. The isomorphism

$$\begin{aligned}\mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![\mathrm{SL}_2(\mathbb{Z}_p)]\!]}(M, V^*) &\cong \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![Z\,\mathrm{SL}_2(\mathbb{Z}_p)]\!]}(M, V^*)\\ &\cong \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![\mathrm{GL}_2(\mathbb{Z}_p)]\!]}(M, V^* \otimes \mathrm{Ind}_{Z\,\mathrm{SL}_2(\mathbb{Z}_p)}^{\mathrm{GL}_2(\mathbb{Z}_p)} \mathbf{1})\\ &\cong \bigoplus_{\eta} \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![\mathrm{GL}_2(\mathbb{Z}_p)]\!]}(M, V^* \otimes \eta \circ \det)\end{aligned}$$

implies the assertion. $\square$

Lemma 2.6. *Let $M \in \mathrm{Mod}^{\mathrm{pro}}_{\mathrm{GL}_2(\mathbb{Z}_p),\psi}(\mathcal{O})$ 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 ψ. If $\{V_i|_{\mathrm{SL}_2(\mathbb{Z}_p)}\}_{i\in I}$ captures $M|_{\mathrm{SL}_2(\mathbb{Z}_p)}$ then $\{V_i \otimes \eta \circ \det\}_{i\in I,\eta}$ captures M, where η runs over all characters $\eta : \mathbb{Z}_p^\times \to L^\times$ with $\eta^2 = 1$.*

Proof. The assertion follows from Lemma 2.5 and [Colmez et al. 2014, Lemma 2.7]. $\square$

Proposition 2.7. *Let $K = \mathrm{GL}_2(\mathbb{Z}_p)$, and let Z be the center of K and $\psi : Z \to L^\times$ a continuous character. There is a smooth irreducible representation τ of K which is a type for a Bernstein component containing a principal series representation, but not containing a special series representation, such that*

$$\{\tau \otimes \mathrm{Sym}^a L^2 \otimes \eta' \circ \det\}_{a\in\mathbb{N},\eta'}$$

captures every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$. Here, for each $a \in \mathbb{N}$, η' runs over all continuous characters $\eta' : \mathbb{Z}_p^\times \to L^\times$ such that $\tau \otimes \mathrm{Sym}^a L^2 \otimes \eta' \circ \det$ has central character ψ.

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 $\chi : \mathbb{Z}_p^\times \to L^\times$ and a continuous character $\eta_0 : \mathbb{Z}_p^\times \to L^\times$ such that $\psi = \chi\eta_0^2$. Let e be the smallest integer such that χ is trivial on $1 + p^e\mathbb{Z}_p$. Let

$$J = \begin{pmatrix} \mathbb{Z}_p^\times & \mathbb{Z}_p \\ p^e\mathbb{Z}_p & \mathbb{Z}_p^\times \end{pmatrix},$$

and let $\chi \otimes \mathbf{1} : J \to L^\times$ be the character which sends $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto \chi(a)$. The representation $\tau := \mathrm{Ind}_J^K(\chi \otimes \mathbf{1})$ is irreducible and is a type. More precisely, for an irreducible smooth $\bar{L}$-representation π of $G = \mathrm{GL}_2(\mathbb{Q}_p)$, we have $\mathrm{Hom}_K(\tau, \pi) \neq 0$ if and only if $\pi \cong (\mathrm{Ind}_B^G \psi_1 \otimes \psi_2)_{\mathrm{sm}}$, where B is a Borel subgroup and $\psi_1|_{\mathbb{Z}_p^\times} = \chi$ and $\psi_2|_{\mathbb{Z}_p^\times} = \mathbf{1}$; see [Henniart 2002, §A.2.2]. The central character of τ is equal to χ. We claim that the family $\left\{\tau \otimes \mathrm{Sym}^{2a} L^2 \otimes (\det)^{-a} \otimes \eta\eta_0 \circ \det\right\}_{a\in\mathbb{N},\eta}$, where η runs over all the characters with $\eta^2 = 1$, captures every projective object in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$. If $M \in \mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$ is projective then $M|_{\mathrm{SL}_2(\mathbb{Z}_p)}$ is projective in $\mathrm{Mod}^{\mathrm{pro}}_{\mathrm{SL}_2(\mathbb{Z}_p),\psi}(\mathcal{O})$ [Emerton 2010b, Proposition 2.1.11]. Lemma 2.4 implies that the family captures $M|_{\mathrm{SL}_2(\mathbb{Z}_p)}$. 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. $\square$

2B. ***The image of Colmez's Montreal functor.*** Let $G = \mathrm{GL}_2(\mathbb{Q}_p)$, $K = \mathrm{GL}_2(\mathbb{Z}_p)$. 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 $\psi : Z \to \mathcal{O}^\times$.

Let $\mathrm{Mod}^{\mathrm{pro}}_G(\mathcal{O})$ be the category of profinite augmented representations of G [Emerton 2010a, Definition 2.1.6]. The Pontryagin duality

$$\pi \mapsto \pi^\vee := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}}(\pi, L/\mathcal{O})$$

induces an antiequivalence of categories between $\mathrm{Mod}^{\mathrm{sm}}_G(\mathcal{O})$ and $\mathrm{Mod}^{\mathrm{pro}}_G(\mathcal{O})$ [Emerton 2010a, (2.2.8)]. Let $\mathrm{Mod}^{\mathrm{l.adm}}_G(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}^{\mathrm{sm}}_G(\mathcal{O})$ consisting of locally admissible [Emerton 2010a, Definition 2.2.17] representations of G and let $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}^{\mathrm{l.adm}}_G(\mathcal{O})$ consisting of those representations on which Z acts by the character ψ. Let $\mathfrak{C}(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}^{\mathrm{pro}}_G(\mathcal{O})$ antiequivalent to $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ via the Pontryagin duality. For $\pi_1, \pi_2 \in \mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ we let $\mathrm{Ext}^i_{G,\psi}(\pi_1, \pi_2)$ be the Yoneda Ext group computed in $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$.

Let $\pi \in \mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ be absolutely irreducible and either supersingular [Barthel and Livné 1994; Breuil 2003a] or a principal series representation isomorphic to $\left(\mathrm{Ind}^G_B\, \chi_1 \otimes \chi_2\omega^{-1}\right)_{\mathrm{sm}}$, for some smooth characters $\chi_1, \chi_2 : \mathbb{Q}_p^\times \to k^\times$ such that $\chi_1\chi_2^{-1} \neq \omega^{\pm 1}, \mathbf{1}$. This hypothesis ensures that $\pi' := \left(\mathrm{Ind}^G_B\, \chi_2 \otimes \chi_1\omega^{-1}\right)_{\mathrm{sm}}$ is also absolutely irreducible and $\pi \not\cong \pi'$. Let $P \twoheadrightarrow \pi^\vee$ be a projective envelope of $\pi^\vee$ in $\mathfrak{C}(\mathcal{O})$ and let $E = \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(P)$. Then E is naturally a topological ring with a unique maximal ideal and residue field $k = \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(\pi^\vee)$; see [Paškūnas 2013, §2].

Proposition 2.8. *If π is supersingular then $k \widehat{\otimes}_E P \cong \pi^\vee$. If π is a principal series then $k \widehat{\otimes}_E P \cong \kappa^\vee$, where κ 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 $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ which is maximal with respect to the following conditions: (1) $\mathrm{soc}_G(k \widehat{\otimes}_E P)^\vee \cong \pi$; (2) π occurs in $(k \widehat{\otimes}_E P)^\vee$ with multiplicity one; see [Paškūnas 2013, Remark 1.13]. For, if $\tau \in \mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ satisfies both conditions, then (1) and [Paškūnas 2013, Lemma 2.10] imply that the natural map $\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \tau^\vee) \widehat{\otimes}_E P \to \tau^\vee$ is surjective, and (2) and the exactness of $\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P, *)$ imply that $\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \tau^\vee) \cong \mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \pi^\vee) \cong k$. Hence, dually we obtain an injection $\tau \hookrightarrow (k \widehat{\otimes}_E P)^\vee$.

Let π_1 be an irreducible representation in $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ such that $\mathrm{Ext}^1_{G,\psi}(\pi_1,\pi)$ is nonzero. It follows from Corollary 1.2 in [Paškūnas 2014] that if π is supersingular then $\pi_1 \cong \pi$ and hence $(k \widehat{\otimes}_E P)^\vee \cong \pi$, and if π 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 $\mathrm{Ext}^1_{G,\psi}(\pi',\pi)$ is one-dimensional. Let κ be the unique nonsplit extension $0 \to \pi \to \kappa \to \pi' \to 0$. We claim that $\mathrm{Ext}^n_{G,\psi}(\pi',\kappa)=0$ for all $n \geq 0$. The claim for $n = 1$ implies that $(k \widehat{\otimes}_E P)^\vee \cong \kappa$. It is proved in [Emerton and Paškūnas 2010, Corollary 3.12] that the δ-functor $H^\bullet \mathrm{Ord}_B$, defined in [Emerton 2010b, Definition 3.3.1], is effaceable in $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$. Hence it coincides with the derived functor $\mathbb{R}^\bullet \mathrm{Ord}_B$. An open compact subgroup N_0 of the unipotent radical of B is isomorphic to $\mathbb{Z}_p$, and hence $H^i(N_0,*)$ vanishes for $i \geq 2$. This implies that $\mathbb{R}^i \mathrm{Ord}_B = H^i \mathrm{Ord}_B = 0$ for $i \geq 2$. The proof of [Paškūnas 2013, Lemma 8.1] does not use the assumption $p > 2$ and gives that

$$\mathrm{Ord}_B \kappa \cong \mathrm{Ord}_B \pi \cong \mathbb{R}^1 \mathrm{Ord}_B \pi' \cong \mathbb{R}^1 \mathrm{Ord}_B \kappa \cong \chi_2\omega^{-1} \otimes \chi_1. \tag{2}$$

Our assumption on χ_1 and χ_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 $\pi' \cong \left(\mathrm{Ind}_{\overline{B}}^G \chi_1\omega^{-1} \otimes \chi_2\right)_{\mathrm{sm}}$, 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 $\mathrm{Ext}^n_{G,\psi}(\pi',\kappa)$ are zero. Hence, $\mathrm{Ext}^n_{G,\psi}(\pi_2,\kappa)=0$ for all $n \geq 0$. Let us also note that the 5-term exact sequence associated to the spectral sequence implies that $\mathrm{Ext}^1_{G,\psi}(\pi,\kappa)$ is finite-dimensional. □

Proposition 2.9. *If π is supersingular then let $S = Q = \pi^\vee$. If π is a principal series then let $S = \pi^\vee$ and $Q = \kappa^\vee$. Then S and Q satisfy the hypotheses* (H0)–(H5) *of* [Paškūnas 2013, §3].

Proof. If π is supersingular then there are no other irreducible representations in the block of π and hence the only hypothesis to check is (H4), which is equivalent to the finite-dimensionality of $\mathrm{Ext}^1_{G,\psi}(\pi,\pi)$. This follows from Proposition 9.1 in [Paškūnas 2010b]. If π 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 $\widehat{\otimes}_E P$ is an exact functor from the category of pseudo-compact right E-modules to $\mathfrak{C}(\mathcal{O})$.*

If m is a pseudocompact right E-module then $\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \mathrm{m}\,\widehat{\otimes}_E\, P) \cong \mathrm{m}$ by [Paškūnas 2013, Lemma 2.9]. This implies that the functor is fully faithful, so that

$$\mathrm{Hom}_E^{\mathrm{cont}}(\mathrm{m}_1, \mathrm{m}_2) \cong \mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(\mathrm{m}_1\,\widehat{\otimes}_E\, P, \mathrm{m}_2\,\widehat{\otimes}_E\, P). \tag{3}$$

Proposition 2.11. *E is commutative.*

Proof. Let $\widetilde{\mathfrak{C}}(\mathcal{O})$ be the full subcategory of $\mathrm{Mod}_G^{\mathrm{pro}}(\mathcal{O})$ which is antiequivalent to $\mathrm{Mod}_G^{\mathrm{l.adm}}(\mathcal{O})$ via the Pontryagin duality. Let $\widetilde{P}$ be a projective envelope of $\pi^\vee$ in $\widetilde{\mathfrak{C}}(\mathcal{O})$, let $\widetilde{E} := \mathrm{End}_{\widetilde{\mathfrak{C}}(\mathcal{O})}(\widetilde{P})$ and let $\mathfrak{a}$ be the closed two-sided ideal of $\widetilde{E}$ generated by the elements $z - \psi^{-1}(z)$, for all z in the center of G. We may consider $\mathfrak{C}(\mathcal{O})$ as a full subcategory of $\widetilde{\mathfrak{C}}(\mathcal{O})$. Since the center of G acts on $\widetilde{P}/\mathfrak{a}\widetilde{P}$ by ψ^{-1}, we have $\widetilde{P}/\mathfrak{a}\widetilde{P} \in \mathfrak{C}(\mathcal{O})$. The functor $\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(\widetilde{P}/\mathfrak{a}\widetilde{P}, *)$ is exact, since

$$\mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(\widetilde{P}/\mathfrak{a}\widetilde{P}, M) = \mathrm{Hom}_{\widetilde{\mathfrak{C}}(\mathcal{O})}(\widetilde{P}, M) \tag{4}$$

for all $M \in \mathfrak{C}(\mathcal{O})$, and $\widetilde{P}$ is projective. Hence, $\widetilde{P}/\mathfrak{a}\widetilde{P}$ is projective in $\mathfrak{C}(\mathcal{O})$. Its G-cosocle is isomorphic to $\pi^\vee$, since the same is true of $\widetilde{P}$. Hence, $\widetilde{P}/\mathfrak{a}\widetilde{P}$ is a projective envelope of $\pi^\vee$ in $\mathfrak{C}(\mathcal{O})$. Since projective envelopes are unique up to isomorphism, $\widetilde{P}/\mathfrak{a}\widetilde{P}$ is isomorphic to P. Since $\mathfrak{a}$ is generated by central elements, any $\phi \in \widetilde{E}$ maps $\mathfrak{a}\widetilde{P}$ to itself. This yields a ring homomorphism $\widetilde{E} \to \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(\widetilde{P}/\mathfrak{a}\widetilde{P}) \cong E$. Projectivity of $\widetilde{P}$ and (4) applied with $M = \widetilde{P}/\mathfrak{a}\widetilde{P}$ implies that the homomorphism is surjective and induces an isomorphism $\widetilde{E}/\mathfrak{a} \cong \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(\widetilde{P}/\mathfrak{a}\widetilde{P})$. Since $\widetilde{E}$ 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 $\mathfrak{m}$, which implies that E is complete for the $\mathfrak{m}$-adic topology. It follows from Lemma 3.7, Proposition 3.8(iii) of [Paškūnas 2013] that $\mathfrak{m}/(\mathfrak{m}^2 + (\varpi))$ 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 π is supersingular and let $Q = \kappa^\vee$ if π is a principal series. The ring E represents the universal deformation problem of Q in $\mathfrak{C}(\mathcal{O})$, 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 ψ as a character of $G_{\mathbb{Q}_p}$ via local class field theory, normalized so that the uniformizers correspond to geometric Frobenii. Let $\varepsilon : G_{\mathbb{Q}_p} \to \mathcal{O}^\times$ be the p-adic cyclotomic character. Similarly, we will identify ε with the character of $\mathbb{Q}_p^\times$, which maps x to $x\,|x|$.

Colmez [2010] has defined an exact and covariant functor $\boldsymbol{V}$ 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 $\mathrm{GL}_2(\mathbb{Q}_p)$ and $G_{\mathbb{Q}_p}$ worlds. We modify Colmez's functor to obtain an exact covariant functor

$$\check{\boldsymbol{V}} : \mathfrak{C}(\mathcal{O}) \to \mathrm{Mod}^{\mathrm{pro}}_{G_{\mathbb{Q}_p}}(\mathcal{O})$$

as follows. Let M be in $\mathfrak{C}(\mathcal{O})$. If it is of finite length then $\check{\boldsymbol{V}}(M) := \boldsymbol{V}(M^\vee)^\vee(\varepsilon\psi)$, where $\vee$ denotes the Pontryagin dual and ε 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 $\mathfrak{C}(\mathcal{O})$, and we define $\check{\boldsymbol{V}}(M) := \varprojlim \check{\boldsymbol{V}}(M_i)$. If $\pi \in \mathrm{Mod}^{\mathrm{l.fin}}_{G,\psi}(k)$ is absolutely irreducible, then $\pi^\vee$ is an object of $\mathfrak{C}(\mathcal{O})$, and if π is supersingular in the sense of [Barthel and Livné 1994], then $\check{\boldsymbol{V}}(\pi^\vee) \cong \boldsymbol{V}(\pi)$ is an absolutely irreducible continuous representation of $G_{\mathbb{Q}_p}$ associated to π by Breuil [2003a]. If $\pi \cong \left(\mathrm{Ind}_B^G \chi_1 \otimes \chi_2\omega^{-1}\right)_{\mathrm{sm}}$ then $\check{\boldsymbol{V}}(\pi^\vee) \cong \chi_1$. If $\pi \cong \chi \circ \det$ then $\check{\boldsymbol{V}}(\pi^\vee) = 0$ and if $\pi \cong \mathrm{Sp} \otimes \chi \circ \det$, where Sp is the Steinberg representation, then $\check{\boldsymbol{V}}(\pi^\vee) \cong \chi$. Since $\check{\boldsymbol{V}}$ is exact we obtain an exact sequence of $G_{\mathbb{Q}_p}$-representations

$$0 \to \chi_2 \to \check{\boldsymbol{V}}(\kappa^\vee) \to \chi_1 \to 0. \tag{5}$$

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

$$\check{\boldsymbol{V}}(\mathrm{m} \mathbin{\widehat{\otimes}}_E P) \cong \mathrm{m} \mathbin{\widehat{\otimes}}_E \check{\boldsymbol{V}}(P), \tag{6}$$

by [Paškūnas 2013, Lemma 5.53]. It follows from (6) and Proposition 2.10 that $\check{\boldsymbol{V}}(P)$ is a deformation of $\rho := \check{\boldsymbol{V}}(k \mathbin{\widehat{\otimes}}_E P)$ to E. If π is supersingular then ρ is an absolutely irreducible 2-dimensional representation of $G_{\mathbb{Q}_p}$, and if π is a principal series then ρ is a nonsplit extension of distinct characters; see (5). In both cases, $\mathrm{End}_{G_{\mathbb{Q}_p}}(\rho) = k$ and so the universal deformation problem of ρ is represented by a complete local noetherian $\mathcal{O}$-algebra R. Let R^ψ be the quotient of R parametrizing deformations of ρ with determinant equal to $\psi\varepsilon$.

Proposition 2.14. *The functor $\check{\boldsymbol{V}}$ induces surjective homomorphisms $R \twoheadrightarrow E$ and $\varphi : E \twoheadrightarrow R^\psi$.*

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 $\check{V}$ induces an injection

$$\mathrm{Ext}^1_{\mathfrak{C}(\mathcal{O})}(Q, Q) \hookrightarrow \mathrm{Ext}^1_{G_{\mathbb{Q}_p}}(\rho, \rho).$$

This follows from [Colmez 2010, Théorème VII.5.2]. To prove the second surjection, we observe that R^{ψ} is reduced and $\mathcal{O}$-torsion-free: if $p \geq 5$ then R^{ψ} 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 $\operatorname{Spec} R^{\psi}[1/p]$ is contained in $\operatorname{Spec} E$. This is equivalent to showing that for every deformation $\tilde{\rho}$ of ρ with determinant $\psi\varepsilon$ there is a Banach space representation Π lifting $Q^{\vee}$ with central character ψ such that $\check{V}(\Pi) \cong \tilde{\rho}$. This follows from [Colmez et al. 2015, Theorem 10.1]. □

2B2. *Banach space representations.* Let $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ be the category of admissible unitary L-Banach space representations [Schneider and Teitelbaum 2002, §3] on which Z acts by the character ψ. If $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ then we let

$$\check{V}(\Pi) := \check{V}(\Theta^d) \otimes_{\mathcal{O}} L, \tag{7}$$

where Θ is any open bounded G-invariant lattice in Π. Therefore, $\check{V}$ is exact and contravariant on $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$.

Remark 2.15. One of the reasons we use $\check{V}$ instead of V is that this allows us to define $\check{V}(\Pi)$ without making the assumption that the reduction of Π modulo ϖ has finite length as a G-representation.

If m is an $E[1/p]$-module of finite length then we let

$$\Pi(\mathrm{m}) := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}}(\mathrm{m}^0 \widehat{\otimes}_E P, L), \tag{8}$$

where m^0 is any E-stable $\mathcal{O}$-lattice in m. Then $\Pi(\mathrm{m})$ 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 $\widehat{\otimes}_E P$ is exact by Proposition 2.10, the functor $\mathrm{m} \mapsto \Pi(\mathrm{m})$ is exact and contravariant. Moreover, it is fully faithful, as

$$\begin{aligned}\mathrm{Hom}_G(\Pi(\mathrm{m}_1), \Pi(\mathrm{m}_2)) &\cong \mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}\big(\mathrm{m}_2^0 \widehat{\otimes}_E P, \mathrm{m}_1^0 \widehat{\otimes}_E P\big)_L \\ &\cong \mathrm{Hom}_{E[1/p]}(\mathrm{m}_2, \mathrm{m}_1),\end{aligned} \tag{9}$$

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* $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ *be such that* π *does not occur as a subquotient in the reduction of an open bounded*

G-invariant lattice in Π modulo ϖ. Then $\operatorname{Ext}^1_G(\Pi, \Pi(\mathrm{m}))$ computed in $\operatorname{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ is zero.

Proof. If Θ is an open bounded G-invariant lattice in $\mathrm{B} \in \operatorname{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ then we define $\mathrm{m}(\mathrm{B}) := \operatorname{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \Theta^d)_L$. Proposition 4.17 in [Paškūnas 2013] implies that $\mathrm{m}(\mathrm{B})$ is a finitely generated $E[1/p]$-module. The functor $\mathrm{B} \mapsto \mathrm{m}(\mathrm{B})$ is exact by [Paškūnas 2013, Lemma 4.9]. The evaluation map $\operatorname{Hom}_{\mathfrak{C}(\mathcal{O})}(P, \Theta^d) \widehat{\otimes}_E P \to \Theta^d$ induces a continuous G-equivariant map $\mathrm{B} \to \Pi(\mathrm{m}(\mathrm{B}))$. If m is an $E[1/p]$-module of finite length and $\mathrm{B} \cong \Pi(\mathrm{m})$ then $\mathrm{m}(\mathrm{B}) \cong \mathrm{m}$ and the map $\mathrm{B} \to \Pi(\mathrm{m}(\mathrm{B}))$ is an isomorphism by [Paškūnas 2013, Lemma 4.28]. Moreover, $\mathrm{m}(\mathrm{B}) = 0$ if and only if π 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 $0 \to \Pi(\mathrm{m}) \to \mathrm{B} \to \Pi \to 0$ then by applying the functor m to it, we obtain an isomorphism $\mathrm{m} \cong \mathrm{m}(\Pi(\mathrm{m})) \cong \mathrm{m}(\mathrm{B})$ and hence an isomorphism $\Pi(\mathrm{m}) \cong \Pi(\mathrm{m}(\mathrm{B}))$. The map $\mathrm{B} \to \Pi(\mathrm{m}(\mathrm{B}))$ splits the exact sequence. $\square$

The proof of [Paškūnas 2015b, Lemma 4.3] shows that we have a natural isomorphism of $G_{\mathbb{Q}_p}$-representations

$$\check{V}(\Pi(\mathrm{m})) \cong \mathrm{m} \otimes_E \check{V}(P). \tag{10}$$

Let us point out a special case of this isomorphism. If $\mathfrak{n}$ is a maximal ideal of $E[1/p]$ then its residue field $\kappa(\mathfrak{n})$ is a finite extension of L. Let $\mathcal{O}_{\kappa(\mathfrak{n})}$ be the ring of integers in $\kappa(\mathfrak{n})$ and let $\varpi_{\kappa(\mathfrak{n})}$ be the uniformizer. Then $\Theta := \operatorname{Hom}^{\mathrm{cont}}_{\mathcal{O}}(\mathcal{O}_{\kappa(\mathfrak{n})} \widehat{\otimes}_E P, \mathcal{O})$ is an open bounded G-invariant lattice in $\Pi(\kappa(\mathfrak{n}))$. The evaluation map induces an isomorphism $\Theta^d \cong \mathcal{O}_{\kappa(\mathfrak{n})} \widehat{\otimes}_E P$. Since E is noetherian, $\mathcal{O}_{\kappa(\mathfrak{n})}$ is a finitely presented E-module and thus the usual and completed tensor products coincide. We obtain

$$\check{V}(\Theta^d) \cong \mathcal{O}_{\kappa(\mathfrak{n})} \otimes_E \check{V}(P), \quad \check{V}(\Pi(\kappa(\mathfrak{n}))) \cong \kappa(\mathfrak{n}) \otimes_E \check{V}(P). \tag{11}$$

Since the residue field of $\mathcal{O}_{\kappa(\mathfrak{n})}$ is k, we have

$$\Theta/(\varpi_{\kappa(\mathfrak{n})}) \cong \operatorname{Hom}^{\mathrm{cont}}_k(k \widehat{\otimes}_E P, k) \cong (k \widehat{\otimes}_E P)^\vee. \tag{12}$$

Recall from [Paškūnas 2013, §4] that $\Pi \in \operatorname{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ 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 $\operatorname{Ban}^{\mathrm{adm}}_{G,\psi}(L')$ for every finite field extension L'/L. An irreducible Π 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 $\mathfrak{n}$ is a maximal ideal of $E[1/p]$ then either the $\kappa(\mathfrak{n})$-Banach space representation $\Pi(\kappa(\mathfrak{n}))$ is absolutely irreducible nonordinary or*

$$\pi \cong \left(\operatorname{Ind}^G_B \chi_1 \otimes \chi_2 \omega^{-1}\right)_{\mathrm{sm}}$$

and (after possibly replacing $\kappa(\mathfrak{n})$ by a finite extension) there exists a nonsplit extension

$$0 \to \left(\operatorname{Ind}_B^G \delta_1 \otimes \delta_2 \varepsilon^{-1}\right)_{\mathrm{cont}} \to \Pi(\kappa(\mathfrak{n})) \to \left(\operatorname{Ind}_B^G \delta_2 \otimes \delta_1 \varepsilon^{-1}\right)_{\mathrm{cont}} \to 0, \tag{13}$$

where $\delta_1, \delta_2 : \mathbb{Q}_p^\times \to \kappa(\mathfrak{n})^\times$ are unitary characters congruent to χ_1 and χ_2, respectively, such that $\delta_1 \delta_2 = \psi \varepsilon$.

Proof. It follows from (11) that $\dim_{\kappa(\mathfrak{n})} \check{V}(\Pi(\kappa(\mathfrak{n}))) = 2$. Since $\check{V}$ applied to a parabolic induction of a unitary character is a one-dimensional representation of $G_{\mathbb{Q}_p}$, we deduce that if $\Pi(\kappa(\mathfrak{n}))$ is absolutely irreducible then it cannot be ordinary.

If π is supersingular then (12) implies that $\Theta/(\varpi_{\kappa(\mathfrak{n})}) \cong \pi$, which is absolutely irreducible. This implies that $\Pi(\kappa(\mathfrak{n}))$ is absolutely irreducible. If π is a principal series then $\Theta/(\varpi_{\kappa(\mathfrak{n})})$ is of length 2 and both irreducible subquotients are absolutely irreducible. Hence, $\Pi(\kappa(\mathfrak{n}))$ is either irreducible or of length 2. Let us assume that $\Pi(\kappa(\mathfrak{n}))$ is not absolutely irreducible. Then after possibly replacing $\kappa(\mathfrak{n})$ by a finite extension we have an exact sequence of admissible $\kappa(\mathfrak{n})$-Banach space representations $0 \to \Pi_1 \to \Pi(\kappa(\mathfrak{n})) \to \Pi_2 \to 0$. This sequence is nonsplit, since otherwise $\check{V}(\Pi(\kappa(\mathfrak{n})))$ 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 Θ_2 be the image of Θ in Π_2. Since we are dealing with admissible representations, Θ_2 is a bounded $\mathcal{O}$-lattice in Π_2. Lemma 5.5 of [Paškūnas 2010a] says that we have the exact sequences of $\mathcal{O}_{\kappa(\mathfrak{n})}$-modules

$$0 \to \Theta_1 \to \Theta \to \Theta_2 \to 0, \tag{14}$$

$$0 \to \Theta_1/(\varpi_{\kappa(\mathfrak{n})}) \to \Theta/(\varpi_{\kappa(\mathfrak{n})}) \to \Theta_2/(\varpi_{\kappa(\mathfrak{n})}) \to 0. \tag{15}$$

It follows from (12) that the exact sequence of G-representations in (15) is the unique nonsplit extension $0 \to \pi \to \kappa \to \pi' \to 0$. Proposition 4.2.14 of [Emerton 2010b] applied with $A = \mathcal{O}_{\kappa(\mathfrak{n})}/(\varpi^n_{\kappa(\mathfrak{n})})$ for all $n \geq 1$ implies that

$$\Pi_1 \cong \left(\operatorname{Ind}_B^G \delta_1 \otimes \delta_2 \varepsilon^{-1}\right)_{\mathrm{cont}}, \quad \Pi_2 \cong \left(\operatorname{Ind}_B^G \delta_2' \otimes \delta_1' \varepsilon^{-1}\right)_{\mathrm{cont}},$$

where $\delta_1, \delta_2, \delta_1', \delta_2' : \mathbb{Q}_p^\times \to \kappa(\mathfrak{n})^\times$ are unitary characters with δ_1, δ_1' congruent to χ_1 and δ_2, δ_2' congruent to χ_2 modulo $\varpi_{\kappa(\mathfrak{n})}$. We reduce (14) modulo $\varpi^n_{\kappa(\mathfrak{n})}$ to obtain an exact sequence to which we apply Ord_B. This gives us an injection $\operatorname{Ord}_B\left(\Theta_2/(\varpi^n_{\kappa(\mathfrak{n})})\right) \hookrightarrow \mathbb{R}^1 \operatorname{Ord}_B\left(\Theta_2/(\varpi^n_{\kappa(\mathfrak{n})})\right)$. Since both are free $\mathcal{O}_{\kappa(\mathfrak{n})}/(\varpi^n_{\kappa(\mathfrak{n})})$-modules of rank 1, the injection is an isomorphism. This implies that δ_1 is congruent to δ_1' and δ_2 is congruent δ_2' modulo $\varpi^n_{\kappa(\mathfrak{n})}$ for all $n \geq 1$. Hence, $\delta_1 = \delta_1'$ and $\delta_2 = \delta_2'$. □

2B3. *Main local result.* We will prove that the surjection $\varphi : E \twoheadrightarrow R^\psi$ 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^ψ 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 ψ of the form $\tau \otimes \mathrm{Sym}^a L^2 \otimes \eta \circ \det$, where $\eta : \mathbb{Z}_p^\times \to L^\times$ is a continuous character, and τ is a type for a Bernstein component containing a principal series representation, but not containing a special series representation.

Proposition 2.18. *If* $\mathfrak{n}$ *is a maximal ideal of* $E[1/p]$ *then the following hold*:

(i) $\dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))) \leq 1$.

(ii) $\dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Pi(E_\mathfrak{n}/\mathfrak{n}^2)) \leq 2$.

Moreover, *if* $\mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))) \neq 0$ *then* $\det \check{V}(\Pi(\kappa(\mathfrak{n}))) = \psi\varepsilon$.

Proof. If m is an $E[1/p]$-module of finite length and L' is a finite extension of L, then $\Pi(\mathrm{m} \otimes_L L') \cong \Pi(\mathrm{m}) \otimes_L L'$ and $\mathrm{Hom}_K(V, \Pi(\mathrm{m})) \otimes_L L' \cong \mathrm{Hom}_K(V, \Pi(\mathrm{m}) \otimes_L L')$. This implies that it is enough to prove the assertions after replacing $\kappa(\mathfrak{n})$ by a finite extension. In particular, we may assume that $\Pi(\kappa(\mathfrak{n}))$ is either absolutely irreducible or a nonsplit extension as in Proposition 2.17. Since $\check{V}$ is compatible with twisting by characters, to prove the proposition it is enough to assume that η is trivial, so that V is a locally algebraic representation of K.

Since τ is a type and $\Pi(\kappa(\mathfrak{n}))$ is admissible, $\mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))) \neq 0$ if and only if (after possibly replacing $\kappa(\mathfrak{n})$ by a finite extension) $\Pi(\kappa(\mathfrak{n}))$ contains a subrepresentation of the form $\Psi \otimes \mathrm{Sym}^a L^2$, where Ψ is an absolutely irreducible smooth principal series representation in the Bernstein component described by τ; see the proof of [Paškūnas 2010a, Theorem 7.2]. Let Π be the universal unitary completion of $\Psi \otimes \mathrm{Sym}^a L^2$. Then Π is absolutely irreducible, by [Berger and Breuil 2010, Corollaire 5.3.4] and [Breuil and Emerton 2010, Proposition 2.2.1].

If $\Pi(\kappa(\mathfrak{n}))$ is absolutely irreducible, we deduce that $\Pi(\kappa(\mathfrak{n})) \cong \Pi$. Since Π in [Berger and Breuil 2010] is constructed out of a (φ, Γ)-module of a 2-dimensional crystabeline representation of $G_{\mathbb{Q}_p}$ with determinant $\psi\varepsilon$, applying $\check{V}$ undoes this construction to obtain the Galois representation we started with. In particular, $\det \check{V}(\Pi(\kappa(\mathfrak{n}))) = \psi\varepsilon$. Moreover, it follows from [Colmez 2010, Théorème VI.6.50] that the locally algebraic vectors in $\Pi(\kappa(\mathfrak{n}))$ are isomorphic to $\Psi \otimes \mathrm{Sym}^a L^2$, which implies that

$$\dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))) = \dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Psi \otimes \mathrm{Sym}^a L^2) = 1, \qquad (16)$$

giving part (i).

If $\Pi(\kappa(\mathfrak{n}))$ is reducible, then using the fact that (13) is nonsplit we deduce that Π is the unique irreducible subrepresentation of $\Pi(\kappa(\mathfrak{n}))$. It follows from [Paškūnas

2013, Lemma 12.5][1] that the locally algebraic vectors in Π are isomorphic to $\Psi \otimes \mathrm{Sym}^a L^2$ and the locally algebraic vectors in $\Pi(\kappa(\mathfrak{n}))/\Pi$ are zero. Thus locally algebraic vectors in $\Pi(\kappa(\mathfrak{n}))$ are isomorphic to $\Psi \otimes \mathrm{Sym}^a L^2$ and so part (i) holds. Moreover, applying $\check{V}$ to (13) we obtain an exact sequence $0 \to \delta_2 \to \check{V}(\Pi(\kappa(\mathfrak{n}))) \to \delta_1 \to 0$. Hence, $\det \check{V}(\Pi(\kappa(\mathfrak{n}))) = \delta_1\delta_2 = \psi\varepsilon$.

The exact sequence $0 \to \mathfrak{n}/\mathfrak{n}^2 \to E_{\mathfrak{n}}/\mathfrak{n}^2 \to \kappa(\mathfrak{n}) \to 0$ of $E[1/p]$-modules gives rise to an exact sequence of admissible Banach space representations of G

$$0 \to \Pi(\kappa(\mathfrak{n})) \to \Pi(E_{\mathfrak{n}}/\mathfrak{n}^2) \to \Pi(\kappa(\mathfrak{n}))^{\oplus d} \to 0,$$

where $d = \dim_{\kappa(\mathfrak{n})} \mathfrak{n}/\mathfrak{n}^2$. We claim that $\mathrm{Hom}_G(\Pi, \Pi(E_{\mathfrak{n}}/\mathfrak{n}^2))$ is one-dimensional as a $\kappa(\mathfrak{n})$-vector space. Given the claim we can deduce part (ii) by the same argument as in [Paškūnas 2015b, Corollary 4.21]. To show the claim let $\Pi' := \Pi(\kappa(\mathfrak{n}))/\Pi$. If Π' is zero then the assertion follows from (9). If Π' is nonzero then the reduction of the unit ball modulo $\varpi_{\kappa(\mathfrak{n})}$ is isomorphic to π'. Since (13) is nonsplit we obtain $\mathrm{Hom}_G(\Pi', \Pi(\kappa(\mathfrak{n}))) = 0$, and Lemma 2.16 implies that $\mathrm{Ext}^1_G(\Pi', \Pi(\kappa(\mathfrak{n}))) = 0$. Hence, $\mathrm{Hom}_G(\Pi(\kappa(\mathfrak{n})), \Pi(E_{\mathfrak{n}}/\mathfrak{n}^2)) \cong \mathrm{Hom}_G(\Pi, \Pi(E_{\mathfrak{n}}/\mathfrak{n}^2))$ and the claim follows from (9). □

Let Θ be a K-invariant $\mathcal{O}$-lattice in V and let $M(\Theta) := \mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(P, \Theta^d)^d$, where $(*)^d := \mathrm{Hom}_{\mathcal{O}}(*, \mathcal{O})$. It follows from Proposition 2.8 that $(k \widehat{\otimes}_E P)^\vee$ is an admissible representation of G; dually, this implies that $k \widehat{\otimes}_E P$ is a finitely generated $\mathcal{O}[\![K]\!]$-module. Hence, [Paškūnas 2015b, Proposition 2.15] implies that $M(\Theta)$ is a finitely generated E-module. We will denote by m-Spec the set of maximal ideals of a commutative ring.

Proposition 2.19. *Let $\mathfrak{a}$ be the E-annihilator of $M(\Theta)$. Then $E/\mathfrak{a}$ is reduced and $\mathcal{O}$-torsion-free. Moreover,* m-Spec$(E/\mathfrak{a})[1/p]$ *is contained in the image of* m-Spec $R^\psi[1/p]$ *under* $\varphi^\sharp : \mathrm{Spec}\, R^\psi \to \mathrm{Spec}\, E$.

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 $\mathcal{O}[\![K]\!]$-module which is projective in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$. 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

$$\dim_L \mathrm{Hom}_K(V, \Pi(\mathrm{m})) = \dim_L \mathrm{m} \otimes_E M(\Theta), \tag{17}$$

by [Paškūnas 2015b, Proposition 2.22]. Proposition 2.18 together with [Paškūnas 2015b, Proposition 2.32] imply that $E/\mathfrak{a}$ is reduced. It is $\mathcal{O}$-torsion-free, since $M(\Theta)$ is $\mathcal{O}$-torsion-free. Let $\mathfrak{n}$ be a maximal ideal of $E[1/p]$. Since E is a quotient of R, $\mathfrak{n}$ lies in the image of m-Spec $R^\psi[1/p]$ if and only if $\det \kappa(\mathfrak{n}) \otimes_E \check{V}(P) = \psi\varepsilon$.

[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 $(E/\mathfrak{a})[1/p]$. $\square$

Corollary 2.20. *The surjection $\varphi : E \twoheadrightarrow R^{\psi}$, 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 φ induces an isomorphism between $E/\mathfrak{a}$ and the image of R^{ψ} in the maximal reduced quotient of $(R^{\psi}/\varphi(\mathfrak{a}))[1/p]$. This implies that the surjection $E/\mathfrak{a} \twoheadrightarrow R^{\psi}/\varphi(\mathfrak{a})$ is injective, and hence an isomorphism. $\square$

Lemma 2.21. *The E-annihilators of* $\mathrm{Hom}_K^{\mathrm{cont}}(P, V^*)$ *and* $M(\Theta)$ *are equal.*

Proof. One inclusion is trivial; the other follows from [Paškūnas 2015b, (11)], which says that $\mathrm{Hom}_K^{\mathrm{cont}}(P, V^*)$ is naturally isomorphic to $\mathrm{Hom}_{\mathcal{O}}^{\mathrm{cont}}(M(\Theta), L)$. $\square$

Theorem 2.22. *The functor $\check{V}$ induces an isomorphism $\varphi : E \xrightarrow{\cong} R^{\psi}$. Moreover, $\check{V}(P)$ is the universal deformation of ρ with determinant $\psi\varepsilon$.*

Proof. It follows from Corollary 2.20 and Lemma 2.21 that the kernel of φ is contained in the E-annihilator of $\mathrm{Hom}_K^{\mathrm{cont}}(P, V^*)$. It follows from Proposition 2.7 that the intersection of the annihilators as V varies is zero. Hence, φ is injective, and hence an isomorphism by Proposition 2.14. The second part is a formal consequence of the first part. $\square$

2B4. *Blocks.* As explained in the introduction the category $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ decomposes into a product of subcategories

$$\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O}) \cong \prod_{\mathfrak{B} \in \mathrm{Irr}^{\mathrm{adm}}_G/\sim} \mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}], \tag{18}$$

where $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ is the full subcategory of $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})$ consisting of representations with all irreducible subquotients in $\mathfrak{B}$. Dually we obtain a decomposition

$$\mathfrak{C}(\mathcal{O}) \cong \prod_{\mathfrak{B} \in \mathrm{Irr}^{\mathrm{adm}}_G/\sim} \mathfrak{C}(\mathcal{O})[\mathfrak{B}], \tag{19}$$

where $M \in \mathfrak{C}(\mathcal{O})$ lies in $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$ if and only if $M^{\vee}$ lies in $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$.

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 $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ and $P_{\mathfrak{B}}$ is a projective generator of $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$. The ring $E_{\mathfrak{B}} := \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(P_{\mathfrak{B}})$ carries a natural topology with respect to which it is a pseudocompact ring; see [Gabriel 1962, Chapitre IV, Proposition 13]. In addition, the functor

$$M \mapsto \mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P_{\mathfrak{B}}, M)$$

induces an equivalence of categories between $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$ and the category of right pseudocompact $E_{\mathfrak{B}}$-modules; see Corollaire 1 after [Gabriel 1962, Chapitre IV, Théorème 4]. The inverse functor is given by $\mathrm{m} \mapsto \mathrm{m}\,\widehat{\otimes}_{E_{\mathfrak{B}}} P_{\mathfrak{B}}$, as follows from Lemmas 2.9 and 2.10 in [Paškūnas 2013]. Moreover, the center of the category of $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$, which by definition is the ring of the natural transformations of the identity functor, is naturally isomorphic to the center of the ring $E_{\mathfrak{B}}$; see Corollaire 5 after [Gabriel 1962, Chapitre IV, Théorème 4].

Let us prove Theorem 1.2, stated in the introduction. If $\mathfrak{B}$ is a block containing a supersingular representation π then $\mathfrak{B} = \{\pi\}$ and so $\pi_{\mathfrak{B}} = \pi$, $P_{\mathfrak{B}}$ is a projective envelope of $\pi^{\vee}$ and $E_{\mathfrak{B}}$ coincides with the ring denoted by E in the previous section. Theorem 2.22 implies that $E_{\mathfrak{B}}$ is naturally isomorphic to R_{ρ}^{ψ}, the quotient of the universal deformation ring of $\rho := \check{\mathbf{V}}(\pi^{\vee})$ parametrizing deformations with determinant $\psi\varepsilon$. Since this ring is commutative, we deduce that the center of $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$ is naturally isomorphic to R_{ρ}^{ψ}. Moreover, $\check{\mathbf{V}}(P_{\mathfrak{B}})$ is the tautological deformation of ρ to R_{ρ}^{ψ}; see Theorem 2.22.

If $\mathfrak{B}$ contains a generic principal series representation then $\mathfrak{B} = \{\pi_1, \pi_2\}$, where

$$\pi_1 \cong \left(\mathrm{Ind}_B^G \chi_1 \otimes \chi_2\omega^{-1}\right)_{\mathrm{sm}}, \quad \pi_2 \cong \left(\mathrm{Ind}_B^G \chi_2 \otimes \chi_1\omega^{-1}\right)_{\mathrm{sm}}, \tag{20}$$

and $\chi_1, \chi_2 : \mathbb{Q}_p^{\times} \to k^{\times}$ are continuous characters such that $\chi_1\chi_2^{-1} \neq \mathbf{1}, \omega^{\pm 1}$. Then $\pi_{\mathfrak{B}} = \pi_1 \oplus \pi_2$ and so $P_{\mathfrak{B}} \cong P_1 \oplus P_2$, where P_1 is a projective envelope of $\pi_1^{\vee}$ and P_2 is a projective envelope of $\pi_2^{\vee}$ in $\mathfrak{C}(\mathcal{O})$. Thus

$$E_{\mathfrak{B}} \cong \mathrm{End}_{\mathfrak{C}(\mathcal{O})}(P_1 \oplus P_2) \cong \mathrm{End}^{\mathrm{cont}}_{G_{\mathbb{Q}_p}}(\check{\mathbf{V}}(P_1) \oplus \check{\mathbf{V}}(P_2)), \tag{21}$$

where the last isomorphism follows from [Paškūnas 2013, Lemma 8.10]. The assumption on the characters χ_1, χ_2 implies that if we consider them as representations of $G_{\mathbb{Q}_p}$ via the local class field theory, 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 ρ_1, let R_1^{ψ} be the quotient of R_1 parametrizing deformations of ρ_1 with determinant $\psi\varepsilon$, and let ρ_1^{univ} be the tautological deformation of ρ_1 to R_1^{ψ}. We define R_2, R_2^{ψ} and ρ_2^{univ} in the same way with ρ_2 instead of ρ_1. It follows from Theorem 2.22 and (21) that

$$E_{\mathfrak{B}} \cong \mathrm{End}^{\mathrm{cont}}_{G_{\mathbb{Q}_p}}(\rho_1^{\mathrm{univ}} \oplus \rho_2^{\mathrm{univ}}). \tag{22}$$

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 $\rho : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(k)$ be a continuous representation. Let $\mathfrak{A}$ be the category of local artinian augmented $\mathcal{O}$-algebras with residue field k. Let $D^{\mathrm{ps}} : \mathfrak{A} \to \mathrm{Sets}$ be the functor which maps $(A, \mathfrak{m}_A) \in \mathfrak{A}$ 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 $\mathfrak{m}_A$.
- $t : G_{\mathbb{Q}_p} \to A$ is a continuous function with $t(1) = 2$.
- For all $g, h \in G_{\mathbb{Q}_p}$, the following are satisfied:
  (i) $t(g) \equiv \operatorname{tr} \rho(g) \pmod{\mathfrak{m}_A}$.
  (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 $R^{\mathrm{ps},\psi}$ be the quotient of R^{ps} parametrizing those pairs (t, d) where $d = \psi\varepsilon$. Combining (22) with [Paškūnas 2015a, Propositions 3.12 and 4.3, Corollary 4.4] we obtain the following:

Theorem 2.23. *Let $\mathfrak{B} = \{\pi_1, \pi_2\}$ as above and let $\rho = \chi_1 \oplus \chi_2$. The center of $E_{\mathfrak{B}}$, and hence the center of the category $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$, is naturally isomorphic to $R^{\mathrm{ps},\psi}$. Moreover, $E_{\mathfrak{B}}$ is a free $R^{\mathrm{ps},\psi}$-module of rank* 4:

$$E_{\mathfrak{B}} \cong \begin{pmatrix} R^{\mathrm{ps},\psi} e_{\chi_1} & R^{\mathrm{ps},\psi} \tilde{\Phi}_{12} \\ R^{\mathrm{ps},\psi} \tilde{\Phi}_{21} & R^{\mathrm{ps},\psi} e_{\chi_2} \end{pmatrix}.$$

The generators satisfy the following relations:

$$e_{\chi_1}^2 = e_{\chi_1}, \quad e_{\chi_2}^2 = e_{\chi_2}, \quad e_{\chi_1} e_{\chi_2} = e_{\chi_2} e_{\chi_1} = 0, \tag{23}$$

$$e_{\chi_1} \tilde{\Phi}_{12} = \tilde{\Phi}_{12} e_{\chi_2} = \tilde{\Phi}_{12}, \quad e_{\chi_2} \tilde{\Phi}_{21} = \tilde{\Phi}_{21} e_{\chi_1} = \tilde{\Phi}_{21}, \tag{24}$$

$$e_{\chi_2} \tilde{\Phi}_{12} = \tilde{\Phi}_{12} e_{\chi_1} = e_{\chi_1} \tilde{\Phi}_{21} = \tilde{\Phi}_{21} e_{\chi_2} = \tilde{\Phi}_{12}^2 = \tilde{\Phi}_{21}^2 = 0, \tag{25}$$

$$\tilde{\Phi}_{12} \tilde{\Phi}_{21} = c e_{\chi_1}, \quad \tilde{\Phi}_{21} \tilde{\Phi}_{12} = c e_{\chi_2}. \tag{26}$$

The element c is regular in $R^{\mathrm{ps},\psi}$ and generates the reducibility ideal.

In order to state the result about the center of $\mathfrak{C}(\mathcal{O})[\mathfrak{B}]$ in a uniform way, as in Theorem 1.3, we note that if ρ is an irreducible representation then mapping a deformation ρ_A to $(\operatorname{tr} \rho_A, \det \rho_A)$ induces a homomorphism of $\mathcal{O}$-algebras $R^{\mathrm{ps}} \to R_\rho$, which is an isomorphism by [Chenevier 2014, Theorem 2.22, Example 3.4].

For a block $\mathfrak{B}$, let $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$ be the full subcategory of $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ consisting of those Π for which, for some (equivalently any) open bounded G-invariant lattice Θ, all the irreducible subquotients of $\Theta \otimes_{\mathcal{O}} k$ lie in $\mathfrak{B}$. It is shown in

[Paškūnas 2013, Proposition 5.36] that $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)$ decomposes into a direct sum of subcategories

$$\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L) \cong \bigoplus_{\mathfrak{B}\in \mathrm{Irr}^{\mathrm{adm}}_G/\sim} \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}].$$

Corollary 2.24. *If $\mathfrak{B} = \{\pi\}$ with π supersingular then let $\rho = \check{V}(\pi^\vee)$. If $\mathfrak{B} = \{\pi_1, \pi_2\}$ with π_1, π_2 given by (20) then let $\rho = \check{V}(\pi_1^\vee) \oplus \check{V}(\pi_2^\vee) = \chi_1 \oplus \chi_2$. The map $\Pi \mapsto \check{V}(\Pi)$ induces a bijection between the isomorphism classes of*

- *absolutely irreducible nonordinary $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$;*
- *absolutely irreducible $\tilde{\rho} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(L)$ such that $\det \tilde{\rho} = \psi\varepsilon$ and the semisimplification of the reduction modulo ϖ of a $G_{\mathbb{Q}_p}$-invariant $\mathcal{O}$-lattice in $\tilde{\rho}$ is isomorphic to ρ.*

Proof. Given Theorems 1.2 and 2.23, this is proved in the same way as [Paškūnas 2013, Theorem 11.4]. □

If $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$ and Θ is an open bounded G-invariant lattice in Π, then Θ/ϖ^n is an object of $\mathrm{Mod}^{\mathrm{l.adm}}_{G,\psi}(\mathcal{O})[\mathfrak{B}]$ for all $n \geq 1$. Theorem 1.3 gives a natural action of $R^{\mathrm{ps},\psi}$ on Θ/ϖ^n for all $n \geq 1$. Passing to the limit and inverting p, we obtain a natural homomorphism $R^{\mathrm{ps},\psi}[1/p] \to \mathrm{End}^{\mathrm{cont}}_G(\Pi)$.

Corollary 2.25. *Let $\mathfrak{B}$ be as in Corollary 2.24 and let $\Pi \in \mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$ be absolutely irreducible. Then $\mathrm{tr}\, \check{V}(\Pi)$ is equal to the specialization of the universal pseudocharacter $t^{\mathrm{univ}} : G_{\mathbb{Q}_p} \to R^{\mathrm{ps},\psi}$ at $x : R^{\mathrm{ps},\psi} \to \mathrm{End}^{\mathrm{cont}}_G(\Pi) \cong L$.*

Proof. This is proved in the same way as [Paškūnas 2013, Proposition 11.3]. To carry out that proof we need to verify that $\check{V}(P_{\mathfrak{B}})$ is annihilated by $g^2 - t^{\mathrm{univ}}(g)g + \psi\varepsilon(g)$ for all $g \in G_{\mathbb{Q}_p}$. If $\mathfrak{B}$ contains a supersingular representation this follows from Cayley–Hamilton since $\check{V}(P_{\mathfrak{B}})$ is the universal deformation of ρ with determinant $\psi\varepsilon$, and $\mathrm{tr}\, \check{V}(P_{\mathfrak{B}}) = t^{\mathrm{univ}}$ by [Chenevier 2014, Theorem 2.22, Example 3.4]. If $\mathfrak{B}$ contains a generic principal series then $\check{V}(P_{\mathfrak{B}}) \cong \rho_1^{\mathrm{univ}} \oplus \rho_2^{\mathrm{univ}}$ and the assertion follows from [Paškūnas 2015a, Proposition 3.9]. □

Corollary 2.26. *For any Π as in Corollary 2.24, we have $\dim_L \mathrm{Ext}^1_{G,\psi}(\Pi, \Pi) = 3$.*

Proof. Let $\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]$ be the full subcategory of $\mathrm{Ban}^{\mathrm{adm}}_{G,\psi}(L)[\mathfrak{B}]$ consisting of objects of finite length. It follows from [Paškūnas 2013, Theorem 4.36] that this category decomposes into a direct sum of subcategories

$$\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}] \cong \bigoplus_{\mathfrak{n}\in \text{m-Spec}\, R^{\mathrm{ps},\psi}[1/p]} \mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]_{\mathfrak{n}},$$

where, for a maximal ideal $\mathfrak{n}$ of $R^{\mathrm{ps},\psi}[1/p]$, the direct summand $\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]_{\mathfrak{n}}$ consists of those finite-length representations which are killed by a power of $\mathfrak{n}$. Moreover, the last part of [Paškūnas 2013, Theorem 4.36] implies that the functor

$\Pi \mapsto \mathrm{Hom}_{\mathfrak{C}(\mathcal{O})}(P_{\mathfrak{B}}, \Theta^d)[1/p]$, where Θ is any open bounded G-invariant lattice in Π, induces an antiequivalence of categories between $\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]_{\mathfrak{n}}$ and the category of modules of finite length over the $\mathfrak{n}$-adic completion of $E_{\mathfrak{B}}[1/p]$, which we denote by $\widehat{E}_{\mathfrak{B},\mathfrak{n}}$.

Let $\tilde{\rho} = \check{V}(\Pi)$. Corollary 2.24 tells us that $\tilde{\rho}$ is an absolutely irreducible representation with determinant $\psi\varepsilon$. Let $\mathfrak{n}$ be the maximal ideal of $R^{\mathrm{ps},\psi}[1/p]$ corresponding to the pair $(\operatorname{tr}\tilde{\rho}, \det\tilde{\rho})$. It follows from Corollary 2.25 that Π is annihilated by $\mathfrak{n}$ and hence lies in $\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]_{\mathfrak{n}}$. Let A be the completion of $R^{\mathrm{ps},\psi}[1/p]$ at $\mathfrak{n}$. In the supersingular case, $E_{\mathfrak{B}} = R^{\mathrm{ps},\psi} = R^{\psi}$, and so $\widehat{E}_{\mathfrak{B},\mathfrak{n}} = A$. In the generic principal series case, since $\tilde{\rho}$ is absolutely irreducible, the image of the generator of the reducible locus in $R^{\mathrm{ps},\psi}$ in $\kappa(\mathfrak{n})$ is nonzero. It follows from the description of $E_{\mathfrak{B}}$ in Theorem 2.23 that $\widehat{E}_{\mathfrak{B},\mathfrak{n}}$ is isomorphic to the algebra of 2×2 matrices with entries in A. Thus in both cases we get that $\mathrm{Ban}^{\mathrm{adm.fl}}_{G,\psi}(L)[\mathfrak{B}]_{\mathfrak{n}}$ is antiequivalent to the category of A-modules of finite length, and Π is identified with the residue field $\kappa(\mathfrak{n})$ of A. Hence,

$$\mathrm{Ext}^1_{G,\psi}(\Pi, \Pi) \cong \mathrm{Ext}^1_A(\kappa(\mathfrak{n}), \kappa(\mathfrak{n})).$$

Arguing as in [Kisin 2009c, Lemma 2.3.3] we may identify A with the universal deformation ring parametrizing pseudocharacters with determinant $\psi\varepsilon$ and values in local artinian L-algebras which lift $\operatorname{tr}\tilde{\rho}$. 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 parametrizing deformations with determinant $\psi\varepsilon$. This ring is formally smooth over L of dimension 3, as $H^2(G_{\mathbb{Q}_p}, \mathrm{ad}^0(\tilde{\rho})) \cong H^0(G_{\mathbb{Q}_p}, \mathrm{ad}^0(\tilde{\rho})(1)) = 0$ and so the deformation problem of $\tilde{\rho}$ is unobstructed. In particular, $\dim_L \mathrm{Ext}^1_A(\kappa(\mathfrak{n}), \kappa(\mathfrak{n})) = \dim_L \mathfrak{n}A/\mathfrak{n}^2A = 3$. $\square$

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 $\rho : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(k)$ be a continuous representation which is either absolutely irreducible, in which case we let π be a supersingular representation of G such that $V(\pi) \cong \rho$, or which is isomorphic to $\left(\begin{smallmatrix}\chi_1 & * \\ 0 & \chi_2\end{smallmatrix}\right)$, a nonsplit extension with $\chi_1\chi_2^{-1} \neq \mathbf{1}, \omega^{\pm 1}$, in which case we let $\pi = \left(\mathrm{Ind}_B^G \chi_1 \otimes \chi_2\omega^{-1}\right)_{\mathrm{sm}}$. As before we let R^{ψ} be the quotient of the universal deformation ring of ρ parametrizing deformations with determinant $\psi\varepsilon$ and let ρ^{univ} be the tautological deformation of ρ to R^{ψ}.

Proposition 2.27. *P satisfies the hypotheses* (N0)–(N2) *of* [Paškūnas 2015b, §4].

Proof. (N0) says that $k \widehat{\otimes}_{R^{\psi}} P$ is of finite length and finitely generated over $\mathcal{O}[[K]]$. This follows from Proposition 2.8. To verify (N1) we need to show that

$$\mathrm{Hom}_{\mathrm{SL}_2(\mathbb{Q}_p)}(\mathbf{1}, P^{\vee}) = 0.$$

The $\mathrm{SL}_2(\mathbb{Q}_p)$-invariants in $P^\vee$ are stable under the action of G. Since $P^\vee$ is an injective envelope of π, if the subspace is nonzero then it must intersect π nontrivially. However, $\pi^{\mathrm{SL}_2(\mathbb{Q}_p)} = 0$, which concludes the proof. (N2) requires $\check{V}(P)$ and ρ^{univ} to be isomorphic as $R^\psi[\![G_{\mathbb{Q}_p}]\!]$-modules and this is proved in Theorem 2.22. □

Recall from [Serre 2000, §V.A] that the group of d-dimensional cycles $\mathcal{Z}_d(A)$ of a noetherian ring A is a free abelian group generated by $\mathfrak{p} \in \operatorname{Spec} A$ with $\dim A/\mathfrak{p} = d$. For d-dimensional cycles $\sum_{\mathfrak{p}} n_{\mathfrak{p}}\mathfrak{p}$ and $\sum_{\mathfrak{p}} m_{\mathfrak{p}}\mathfrak{p}$, we write $\sum_{\mathfrak{p}} n_{\mathfrak{p}}\mathfrak{p} \le \sum_{\mathfrak{p}} m_{\mathfrak{p}}\mathfrak{p}$, if $n_{\mathfrak{p}} \le m_{\mathfrak{p}}$ for all $\mathfrak{p} \in \operatorname{Spec} A$ with $\dim A/\mathfrak{p} = d$.

If M is a finitely generated A-module of dimension at most d then $M_{\mathfrak{p}}$ is an $A_{\mathfrak{p}}$-module of finite length, which we denote by $\ell_{A_{\mathfrak{p}}}(M_{\mathfrak{p}})$, for all $\mathfrak{p}$ with $\dim A/\mathfrak{p} = d$. We note that $\ell_{A_{\mathfrak{p}}}(M_{\mathfrak{p}})$ is nonzero only for finitely many $\mathfrak{p}$. Thus $z_d(M) := \sum_{\mathfrak{p}} \ell_{A_{\mathfrak{p}}}(M_{\mathfrak{p}})\mathfrak{p}$, where the sum is taken over all $\mathfrak{p} \in \operatorname{Spec} A$ such that $\dim A/\mathfrak{p} = d$, is a well defined element of $\mathcal{Z}_d(A)$.

If $(A, \mathfrak{m})$ is a local ring then we define a Hilbert–Samuel multiplicity $e(z)$ of a cycle $z = \sum_{\mathfrak{p}} n_{\mathfrak{p}}\mathfrak{p} \in \mathcal{Z}_d(A)$ to equal $\sum_{\mathfrak{p}} n_{\mathfrak{p}} e(A/\mathfrak{p})$, where $e(A/\mathfrak{p})$ is the Hilbert–Samuel multiplicity of the ring $A/\mathfrak{p}$. 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 Θ is a continuous representation of K on a free $\mathcal{O}$-module of finite rank, we let

$$M(\Theta) := \bigl(\operatorname{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(P, \Theta^d)\bigr)^d,$$

where $(*)^d := \operatorname{Hom}_{\mathcal{O}}(*, \mathcal{O})$. If λ is a smooth representation of K on an $\mathcal{O}$-torsion module of finite length then we let

$$M(\lambda) := \bigl(\operatorname{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(P, \lambda^\vee)\bigr)^\vee,$$

where the superscript $\vee$ denotes the Pontryagin dual.

Proposition 2.28. *Let Θ be a continuous representation of K on a free $\mathcal{O}$-module of finite rank with central character ψ. Then $M(\Theta)$ is a finitely generated R^ψ-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*

$$z_1(M(\Theta)/\varpi) = \sum_{\sigma} m_\sigma z_1(M(\sigma)), \tag{27}$$

where the sum is taken over all the irreducible smooth k-representations of K, and m_σ denotes the multiplicity with which σ appears as a subquotient of $\Theta \otimes_{\mathcal{O}} k$.

Moreover, $M(\sigma) \neq 0$ if and only if $\operatorname{Hom}_K(\sigma, \pi) \neq 0$, *in which case the Hilbert–Samuel multiplicity of $z_1(M(\sigma))$ is equal to* 1.

Proof. We showed in Proposition 2.27 that $k \widehat{\otimes}_{R^\psi} P$ 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^{ψ}-module. The restriction of P to K is projective in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$ 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^{ψ} such that we have an exact sequence $0 \to P \xrightarrow{x} P \to P/xP \to 0$, where the restriction of P/xP to K is a projective envelope of $(\mathrm{soc}_K \pi)^{\vee}$ in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$. Lemma 2.33 in [Paškūnas 2015b] implies that $M(\Theta)$ is a Cohen–Macaulay module of dimension 2 and that ϖ, x is a regular sequence of parameters. If σ is an irreducible smooth k-representation of K with central character ψ then the proof of [Paškūnas 2015b, Lemma 2.33] yields an exact sequence

$$0 \longrightarrow M(\sigma) \xrightarrow{\;x\;} M(\sigma) \longrightarrow \big(\mathrm{Hom}^{\mathrm{cont}}_{\mathcal{O}[\![K]\!]}(P/xP, \sigma^{\vee})\big)^{\vee} \longrightarrow 0.$$

Since P/xP is a projective envelope of $(\mathrm{soc}_K \pi)^{\vee}$ in $\mathrm{Mod}^{\mathrm{pro}}_{K,\psi}(\mathcal{O})$, we deduce that $\dim_k M(\sigma)/xM(\sigma)$ is equal to $\dim_k \mathrm{Hom}_K(\sigma, \pi)$. If $\mathrm{Hom}_K(\sigma, \pi)$ is zero then Nakayama's lemma implies that $M(\sigma) = 0$. If $\mathrm{Hom}_K(\sigma, \pi)$ is nonzero then it is a one-dimensional k-vector space, since the K-socle of π is multiplicity free. The exact sequence $0 \to M(\sigma) \xrightarrow{x} M(\sigma) \to k \to 0$ implies that $M(\sigma)$ is a cyclic module, and if $\mathfrak{a}$ denotes its annihilator then $R^{\psi}/\mathfrak{a} \cong k[\![x]\!]$. $\square$

Remark 2.29. If ρ is absolutely irreducible and $\rho|_{I_{\mathbb{Q}_p}} \cong \big(\omega_2^{r+1} \oplus \omega_2^{p(r+1)}\big) \otimes \omega^m$ then

$$\mathrm{soc}_K \pi \cong \big(\mathrm{Sym}^r k^2 \oplus \mathrm{Sym}^{p-1-r} k^2 \otimes \det{}^r\big) \otimes \det{}^m,$$

where $0 \le r \le p-1$, $0 \le m \le p-2$ and ω_2 is the fundamental character of Serre of niveau 2; see [Breuil 2003a; 2003b]. If $\rho \cong \big(\begin{smallmatrix}\chi_1 & * \\ 0 & \chi_2\omega^{r+1}\end{smallmatrix}\big) \otimes \omega^m$, where χ_1, χ_2 are unramified and $\chi_1 \ne \chi_2\omega^{r+1}$ then

$$\pi \cong \big(\mathrm{Ind}_B^G \chi_1 \otimes \chi_2\omega^r\big)_{\mathrm{sm}} \otimes \omega^m \circ \det.$$

Hence, $\mathrm{soc}_K \pi \cong \mathrm{Sym}^r k^2 \otimes \det^m$ if $0 < r < p-1$ and $\det^m \oplus \mathrm{Sym}^{p-1} k^2 \otimes \det^m$ otherwise. In particular, $\mathrm{soc}_K \pi$ is multiplicity free.

If $\mathfrak{n} \in \text{m-Spec}\, R^{\psi}[1/p]$ then the residue field $\kappa(\mathfrak{n})$ is a finite extension of L. Let $\mathcal{O}_{\kappa(\mathfrak{n})}$ be the ring of integers in $\kappa(\mathfrak{n})$. By specializing the universal deformation at $\mathfrak{n}$, we obtain a continuous representation $\rho^{\mathrm{univ}}_{\mathfrak{n}} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\mathcal{O}_{\kappa(\mathfrak{n})})$, which reduces to ρ modulo the maximal ideal of $\mathcal{O}_{\kappa(\mathfrak{n})}$. A p-adic Hodge type $(\boldsymbol{w}, \tau, \psi)$ consists of the following data: $\boldsymbol{w} = (a, b)$ is a pair of integers with $b > a$, $\tau : I_{\mathbb{Q}_p} \to \mathrm{GL}_2(L)$ is a representation of the inertia subgroup with an open kernel and $\psi : G_{\mathbb{Q}_p} \to \mathcal{O}^{\times}$ is a continuous character such that $\psi\varepsilon \equiv \det\rho \pmod{\varpi}$, $\psi|_{I_{\mathbb{Q}_p}} = \varepsilon^{a+b-1}\det\tau$, where ε is the p-adic cyclotomic character. If $\rho^{\mathrm{univ}}_{\mathfrak{n}}$ is potentially semistable then we say that it is of type $(\boldsymbol{w}, \tau, \psi)$ if its Hodge–Tate weights are equal to $\boldsymbol{w}$, the determinant is equal to ψ and the restriction of the Weil–Deligne representation, associated to $\rho_{\mathfrak{n}}^{\mathrm{un}}$ by Fontaine [1994], to $I_{\mathbb{Q}_p}$ is isomorphic to τ.

Henniart [2002] has shown the existence of a smooth irreducible representation $\sigma(\tau)$ (resp. $\sigma^{\mathrm{cr}}(\tau)$) of K on an L-vector space such that if π is a smooth absolutely irreducible infinite-dimensional representation of G and $\mathrm{LL}(\pi)$ is the Weil–Deligne representation attached to π by the classical local Langlands correspondence then $\mathrm{Hom}_K(\sigma(\tau), \pi) \neq 0$ (resp. $\mathrm{Hom}_K(\sigma^{\mathrm{cr}}(\tau), \pi) \neq 0$) if and only if $\mathrm{LL}(\pi)|_{I_{\mathbb{Q}_p}} \cong \tau$ (resp. $\mathrm{LL}(\pi)|_{I_{\mathbb{Q}_p}} \cong \tau$ and the monodromy operator N is 0). The representations $\sigma(\tau)$ and $\sigma^{\mathrm{cr}}(\tau)$ are uniquely determined if $p > 2$. If $p = 2$ there might be different choices; we choose one.

We let $\sigma(\boldsymbol{w}, \tau) := \sigma(\tau) \otimes \mathrm{Sym}^{b-a-1} L^2 \otimes \det^a$. Then $\sigma(\boldsymbol{w}, \tau)$ is a finite-dimensional L-vector space. Since K is compact and the action of K on $\sigma(\boldsymbol{w}, \tau)$ is continuous, there is a K-invariant $\mathcal{O}$-lattice Θ in $\sigma(\boldsymbol{w}, \tau)$. Then $\Theta/(\varpi)$ is a smooth finite-length k-representation of K, and we let $\overline{\sigma(\boldsymbol{w}, \tau)}$ be its semisimplification. One may show that $\overline{\sigma(\boldsymbol{w}, \tau)}$ does not depend on the choice of a lattice. For each smooth irreducible k-representation σ of K we let $m_\sigma(\boldsymbol{w}, \tau)$ be the multiplicity with which σ occurs in $\overline{\sigma(\boldsymbol{w}, \tau)}$. We let $\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau) := \sigma^{\mathrm{cr}}(\tau) \otimes \mathrm{Sym}^{b-a-1} L^2 \otimes \det^a$ and let $m_\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)$ be the multiplicity of σ in $\overline{\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)}$. If $p = 2$ then one may show that $\overline{\sigma(\boldsymbol{w}, \tau)}$ and $\overline{\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)}$ do not depend on the choice of $\sigma(\tau)$ and $\sigma^{\mathrm{cr}}(\tau)$.

Proposition 2.30. *Let $V = \sigma(\boldsymbol{w}, \tau)$ (resp. $V = \sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)$) and let Θ be a K-invariant lattice in V. Then $\mathfrak{n} \in$ m-Spec $R^\psi[1/p]$ lies in the support of $M(\Theta)$ if and only if $\rho_{\mathfrak{n}}^{\mathrm{univ}}$ is potentially semistable (resp. potentially crystalline) of type $(\boldsymbol{w}, \tau, \psi)$. Moreover, for such $\mathfrak{n}$, we have $\dim_{\kappa(\mathfrak{n})} M(\Theta) \otimes_{R^\psi} \kappa(\mathfrak{n}) = 1$.*

Proof. Proposition 2.22 of [Paškūnas 2015b] implies that

$$\dim_{\kappa(\mathfrak{n})} M(\Theta) \otimes_{R^\psi} \kappa(\mathfrak{n}) = \dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))).$$

Since V is a locally algebraic representation,

$$\mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))) \cong \mathrm{Hom}_K(V, \Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}}),$$

where the superscript alg denotes the subspace of locally algebraic vectors. This last subspace is nonzero if and only if $\rho_{\mathfrak{n}}^{\mathrm{univ}}$ is potentially semistable (resp. potentially crystalline) of type $(\boldsymbol{w}, \tau, \psi)$, 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 ρ is generic, we don't have to consider the nasty cases here. $\square$

Corollary 2.31. *There exists a reduced, $\mathcal{O}$-torsion-free quotient $R^\psi(\boldsymbol{w}, \tau)$ of R^ψ such that a map of $\mathcal{O}$-algebras $x : R^\psi \to L'$ into a finite field extension of L factors through $R^\psi(\boldsymbol{w}, \tau)$ if and only if ρ_x^{univ} is potentially semistable of type $(\boldsymbol{w}, \tau, \psi)$.*

Moreover, if Θ is a K-invariant $\mathcal{O}$-lattice in $\sigma(\boldsymbol{w}, \tau)$ and $\mathfrak{a}$ is the R^ψ-annihilator of $M(\Theta)$ then $R^\psi(\boldsymbol{w}, \tau) = R^\psi/\sqrt{\mathfrak{a}}$.

The same result holds if we consider potentially crystalline instead of potentially semistable representations with $\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)$ instead of $\sigma(\boldsymbol{w}, \tau)$.

Proof. Since the support of $M(\Theta)$ is closed in Spec R^{ψ}, the assertion follows from Proposition 2.30. □

Corollary 2.32. *Let Θ be a K-invariant lattice in either $\sigma(\boldsymbol{w}, \tau)$ or $\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)$ and let $\mathfrak{a}$ be the R^{ψ}-annihilator of $M(\Theta)$. Then we have equalities of cycles*

$$z_2(R^{\psi}/\mathfrak{a}) = z_2(M(\Theta)), \quad z_1(R^{\psi}/(\mathfrak{a}, \varpi)) = z_1(M(\Theta)/\varpi).$$

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 ϖ is both $R^{\psi}/\mathfrak{a}$- and $M(\Theta)$-regular; see Proposition 2.2.13 in [Emerton and Gee 2014]. □

Proposition 2.33. *Let $\mathfrak{a}$ be the R^{ψ}-annihilator of $M(\Theta)$, where Θ is a K-invariant $\mathcal{O}$-lattice in $\sigma(\boldsymbol{w}, \tau)$ (resp. $\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)$). Then $R^{\psi}/\mathfrak{a}$ is reduced. In particular, it is equal to $R^{\psi}(\boldsymbol{w}, \tau)$ (resp. $R^{\psi,\mathrm{cr}}(\boldsymbol{w}, \tau)$).*

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 $\mathfrak{n}$ in m-Spec $R^{\psi}[1/p]$ lying in the support of $M(\Theta)$,

$$\dim_{\kappa(\mathfrak{n})} \mathrm{Hom}_K(V, \Pi(R^{\psi}_{\mathfrak{n}}/\mathfrak{n}^2 R^{\psi}_{\mathfrak{n}})) \le 2.$$

This amounts to checking that the subspace $\mathcal{E}$ of $\mathrm{Ext}^1_G(\Pi(\kappa(\mathfrak{n})), \Pi(\kappa(\mathfrak{n})))$ generated by the extensions of admissible unitary $\kappa(\mathfrak{n})$-Banach spaces $0 \to \Pi(\kappa(\mathfrak{n})) \to \mathrm{B} \to \Pi(\kappa(\mathfrak{n})) \to 0$ such that the induced map between the subspaces of locally algebraic vectors $\mathrm{B}^{\mathrm{alg}} \to \Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}}$ is surjective, is at most one-dimensional; see the proof of [Paškūnas 2015b, Corollary 4.21].

If τ does not extend to an irreducible representation of $W_{\mathbb{Q}_p}$ then the proof of [Paškūnas 2015b, Theorem 4.19] carries over: the key input into that proof is that the closure of $\Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}}$ in $\Pi(\kappa(\mathfrak{n}))$ is equal to the universal unitary completion of $\Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}}$ 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 $\Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}} \cong \big(\mathrm{Ind}_B^G \chi \otimes \chi|\cdot|^{-1}\big)_{\mathrm{sm}} \otimes W$, where W is an algebraic representation of G and $\chi : \mathbb{Q}_p^{\times} \to \kappa(\mathfrak{n})^{\times}$ 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 $\Pi(\kappa(\mathfrak{n}))^{\mathrm{alg}}$ 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 τ extends to an irreducible representation of $W_{\mathbb{Q}_p}$ then the assertion is proved by Dospinescu [2015]. Although[2] the main theorem of [Dospinescu 2015] is stated under the assumption $p \geq 5$, the argument only uses that assumption if we let $\Pi = \Pi(\kappa(\mathfrak{n}))$, in which case $\det \check{V}(\Pi) = \psi\varepsilon$ and $\dim_L \operatorname{Ext}^1_{G,\psi}(\Pi, \Pi) = 3$. This is given by Corollaries 2.24 and 2.26. □

Theorem 2.34. *There is a finite set $\{\mathcal{C}_\sigma\}_\sigma \subset \mathcal{Z}_1(R^\psi/\varpi)$, indexed by the irreducible smooth k-representations σ of K, such that for all p-adic Hodge types $(\boldsymbol{w}, \tau)$ we have equalities*

$$z_1(R^\psi(\boldsymbol{w}, \tau)/\varpi) = \sum_\sigma m_\sigma(\boldsymbol{w}, \tau)\mathcal{C}_\sigma,$$

$$z_1(R^{\psi,\mathrm{cr}}(\boldsymbol{w}, \tau)/\varpi) = \sum_\sigma m^{\mathrm{cr}}_\sigma(\boldsymbol{w}, \tau)\mathcal{C}_\sigma.$$

The cycle $\mathcal{C}_\sigma$ is nonzero if and only if $\operatorname{Hom}_K(\sigma, \pi) \neq 0$, in which case its Hilbert–Samuel multiplicity is equal to 1.

Proof. Let $\mathfrak{a}$ be the R^ψ-annihilator of $M(\Theta)$, where Θ is a K-invariant $\mathcal{O}$-lattice in $\sigma(\boldsymbol{w}, \tau)$. Corollary 2.31 and Proposition 2.33 imply that

$$z_1(R^\psi(\boldsymbol{w}, \tau)/\varpi) = z_1(R^\psi/(\sqrt{\mathfrak{a}}, \varpi)) = z_1(R^\psi/(\mathfrak{a}, \varpi)).$$

Corollary 2.32 and Proposition 2.28 imply that

$$z_1(R^\psi/(\mathfrak{a}, \varpi)) = \sum_\sigma m_\sigma(\boldsymbol{w}, \tau) z_1(M(\sigma)).$$

We let $\mathcal{C}_\sigma = z_1(M(\sigma))$. The proof in the potentially crystalline case is the same. □

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 $G_{\mathbb{Q}_p}$ of a global modular representation.

Let $\mathfrak{b}$ be the kernel $R^\psi/\mathfrak{a} \twoheadrightarrow R^\psi/\sqrt{\mathfrak{a}}$. Since $M(\Theta)$ is Cohen–Macaulay, $R^\psi/\mathfrak{a}$ is equidimensional. Thus if $\mathfrak{b}$ is nonzero then it is a 2-dimensional R^ψ-module, and the cycle $z_1(\mathfrak{b}/\varpi)$ is nonzero. Since

$$z_1(R^\psi/(\mathfrak{a}, \varpi)) = z_1(R^\psi/(\sqrt{\mathfrak{a}}, \varpi)) + z_1(\mathfrak{b}/\varpi),$$

if $R^\psi/\mathfrak{a}$ is not reduced then we would conclude that $e(R^\psi/(\mathfrak{a}, \varpi)) > e(R^\psi(\boldsymbol{w}, \tau)/\varpi)$. Since $e(R^\psi/(\mathfrak{a}, \varpi)) = e(M(\Theta)/\varpi) = \sum_\sigma m_\sigma(\boldsymbol{w}, \tau) e(\mathcal{C}_\sigma)$, 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 $\bar{\rho}|_{G_{\mathbb{Q}_p}}$, where $\bar{\rho}$ satisfies the assumptions made in Section 3B) then a global argument gives an inequality in the opposite direction, thus allowing

[2] I thank G. Dospinescu for pointing this out to me.

us to conclude that $R^{\psi}/\mathfrak{a}$ 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 ρ and R is the universal deformation ring of ρ then $R^{\square} \cong R[[x_1, x_2, x_3]]$. Thus we have a map of cycle groups

$$f : \mathcal{Z}_i(R) \to \mathcal{Z}_{i+3}(R^{\square}), \quad \mathfrak{p} \mapsto \mathfrak{p}[[x_1, x_2, x_3]],$$

which preserves Hilbert–Samuel multiplicities. The extra variables only keep track of a choice of basis. This implies that if $R^{\psi,\square}(\boldsymbol{w}, \tau)$ is the quotient of $R^{\square}$ parametrizing potentially semistable framed deformations of type $(\boldsymbol{w}, \tau, \psi)$ then $R^{\psi,\square}(\boldsymbol{w}, \tau) \cong R^{\psi}(\boldsymbol{w}, \tau)[[x_1, x_2, x_3]]$, so that the cycle of $R^{\psi,\square}(\boldsymbol{w}, \tau)/\varpi$ is the image of the cycle of $R^{\psi}(\boldsymbol{w}, \tau)/\varpi$ under f. Using this, one may deduce a version of Theorem 2.34 for framed deformation rings.

Let $\rho = \left(\begin{smallmatrix} \chi_1 & 0 \\ 0 & \chi_2 \end{smallmatrix}\right)$, and let $R^{\square}$ be the universal framed deformation ring of ρ. Let $R^{\psi,\square}(\boldsymbol{w}, \tau)$ (resp. $R^{\psi,\square,\mathrm{cr}}(\boldsymbol{w}, \tau)$) be the reduced, $\mathcal{O}$-torsion-free quotient of $R^{\square}$ parametrizing potentially semistable (resp. potentially crystalline) lifts of p-adic Hodge type $(\boldsymbol{w}, \tau, \psi)$.

Theorem 2.37. *There is a subset $\{\mathcal{C}_{1,\sigma}, \mathcal{C}_{2,\sigma}\}_{\sigma}$ of $\mathcal{Z}_4(R^{\psi,\square}/\varpi)$ indexed by the irreducible smooth k-representations σ of K such that for all p-adic Hodge types $(\boldsymbol{w}, \tau)$ we have equalities*

$$z_4\big(R^{\psi,\square}(\boldsymbol{w}, \tau)/\varpi\big) = \sum_{\sigma} m_{\sigma}(\boldsymbol{w}, \tau)(\mathcal{C}_{1,\sigma} + \mathcal{C}_{2,\sigma}),$$

$$z_4\big(R^{\psi,\square,\mathrm{cr}}(\boldsymbol{w}, \tau)/\varpi\big) = \sum_{\sigma} m^{\mathrm{cr}}_{\sigma}(\boldsymbol{w}, \tau)(\mathcal{C}_{1,\sigma} + \mathcal{C}_{2,\sigma}).$$

The cycle $\mathcal{C}_{1,\sigma}$ is nonzero if and only if $\mathrm{Hom}_K\big(\sigma, \big(\mathrm{Ind}_B^G \chi_1 \otimes \chi_2\omega^{-1}\big)_{\mathrm{sm}}\big) \neq 0$, and $\mathcal{C}_{2,\sigma}$ is nonzero if and only if $\mathrm{Hom}_K\big(\sigma, \big(\mathrm{Ind}_B^G \chi_2 \otimes \chi_1\omega^{-1}\big)_{\mathrm{sm}}\big) \neq 0$, 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 $p = 2$, ψ is unramified and either ρ is absolutely irreducible or $\rho^{\mathrm{ss}} = \chi_1 \oplus \chi_2$, with $\chi_1 \neq \chi_2$. If $\boldsymbol{w} = (0, 1)$ and $\tau = \mathbf{1} \oplus \mathbf{1}$ then*

$$R^{\psi,\square,\mathrm{cr}}(\boldsymbol{w}, \tau) = R^{\psi,\square}(\boldsymbol{w}, \tau).$$

In other words, every semistable lift of ρ with Hodge–Tate weights (0, 1) *is crystalline.*

Proof. It is enough to prove the statement when ρ 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 $R^{\psi}(\boldsymbol{w}, \tau) = R^{\psi,\mathrm{cr}}(\boldsymbol{w}, \tau)$. By the same argument as in Remark 2.35 we see that it is enough to show that $R^{\psi}(\boldsymbol{w}, \tau)/\varpi$ and $R^{\psi,\mathrm{cr}}(\boldsymbol{w}, \tau)/\varpi$ 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: $\mathbf{1}$ and st. The K-socle of π in all the cases is isomorphic to $\mathbf{1} \oplus \mathrm{st}$, $\sigma(\boldsymbol{w}, \tau)/\varpi \cong \mathrm{st}$ and $\sigma^{\mathrm{cr}}(\boldsymbol{w}, \tau)/\varpi \cong \mathbf{1}$. The assertion follows from Theorem 2.34. □

Remark 2.39. Assume that $p = 2$, let $\xi : G_{\mathbb{Q}_p} \to \mathcal{O}^{\times}$ be unramified and congruent to ψ modulo ϖ, and let $(\boldsymbol{w}, \tau)$ be arbitrary. It follows from Theorem 2.34, Remark 2.36, Theorem 2.37 and the proof of Corollary 2.38 that

$$z_4\big(R^{\psi,\Box}(\boldsymbol{w}, \tau)/\varpi\big) = (m_{\mathbf{1}}(\boldsymbol{w}, \tau) + m_{\mathrm{st}}(\boldsymbol{w}, \tau))z_4\big(R^{\xi,\Box}((0, 1), \mathbf{1} \oplus \mathbf{1})/\varpi\big),$$

where the cycles live in $\mathcal{Z}_4(R^{\Box})$. 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 $\mathbb{Q}_2$ with the ring integers $\mathcal{O}$ 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 Σ which does not contain any primes dividing 2. We fix a maximal order $\mathcal{O}_D$ of D, and for each finite place $v \notin \Sigma$ we have an isomorphism $(\mathcal{O}_D)_v \cong M_2(\mathcal{O}_{F_v})$. For each finite place v of F we will denote by $\boldsymbol{N}(v)$ the order of the residue field at v, and by $\varpi_v \in F_v$ 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 $\prod_v (\mathcal{O}_D)_v^{\times}$. We assume that if $v \in \Sigma$ then $U_v = (\mathcal{O}_D)_v^{\times}$ and if $v \mid 2$ then $U_v = \mathrm{GL}_2(\mathcal{O}_{F_v}) = \mathrm{GL}_2(\mathbb{Z}_2)$. Let A be a topological $\mathbb{Z}_2$-algebra. For each $v \mid 2$, we fix a continuous representation $\sigma_v : U_v \to \mathrm{Aut}(W_{\sigma_v})$ on a finite free A-module. Write $W_{\sigma} = \bigotimes_{v|2,A} W_{\sigma_v}$ and denote by $\sigma : \prod_{v|2} U_v \to \mathrm{Aut}(W_{\sigma})$ the corresponding representation. We regard σ 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 $U_v \cap \mathcal{O}_{F_v}^{\times}$ on σ is given by multiplication by ψ. We extend the action of U on W_{σ} to $U(\mathbb{A}_F^f)^{\times}$ by letting $(\mathbb{A}_F^f)^{\times}$ act via ψ.

Let $S_{\sigma,\psi}(U, A)$ denote the set of continuous functions

$$f : D^\times \backslash (D \otimes_F \mathbb{A}_F^f)^\times \to W_\sigma$$

such that for $g \in (D \otimes_F \mathbb{A}_F^f)^\times$ we have $f(gu) = \sigma(u)^{-1} f(g)$, $u \in U$, and $f(gz) = \psi^{-1}(z) f(g)$, $z \in (\mathbb{A}_F^f)^\times$. If we write $(D \otimes_F \mathbb{A}_F^f)^\times = \coprod_{i\in I} D^\times t_i U(\mathbb{A}_F^f)^\times$ for some $t_i \in (D \otimes_F \mathbb{A}_F^f)^\times$ and some finite index set I, then we have an isomorphism of A-modules

$$S_{\sigma,\psi}(U, A) \xrightarrow{\cong} \bigoplus_{i\in I} W_\sigma^{(U(\mathbb{A}_F^f)^\times \cap t_i^{-1} D^\times t_i)/F^\times}, \quad f \mapsto (f(t_i))_{i\in I}. \tag{28}$$

Lemma 3.1. *Let $U_{\max} = \prod_v \mathcal{O}_{D_v}^\times$, where the product is taken over all finite places of F. Let $t \in (D \otimes_F \mathbb{A}_F^f)^\times$. Then the group $\big(U_{\max}(\mathbb{A}_F^f)^\times \cap t D^\times t^{-1}\big)/F^\times$ 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]. □

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 $(D \otimes_F \mathbb{A}_F^f)^\times$ 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 ϖ_{v_1}. Then*

$$\big(U(\mathbb{A}_F^f)^\times \cap t D^\times t^{-1}\big)/F^\times = 1 \quad \textit{for all } t \in (D \otimes_F \mathbb{A}_F^f)^\times. \tag{29}$$

Proof. Let $u \in \big(U(\mathbb{A}_F^f)^\times \cap t D^\times t^{-1}\big)$ such that $u \notin F^\times$. Then the F-subalgebra $F[u]$ of tDt^{-1} is a quadratic field extension of F. Let u' be the conjugate of u over F. Then $u' = \mathrm{Nm}(u)/u$, where Nm is the reduced norm. Consider $w = u/u' = u^2/\mathrm{Nm}(u)$. Write $u = hg$ with $h \in U$ and $g \in (\mathbb{A}_F^f)^\times$. Then $\mathrm{Nm}(g) = g^2$ and so $w = u/u' = h^2/\mathrm{Nm}(h)$. Thus w is in U and also in $tD^\times t^{-1}$.

Since $\big(U(\mathbb{A}_F^f)^\times \cap t D^\times t^{-1}\big)/F^\times$ is a subgroup of $\big(U_{\max}(\mathbb{A}_F^f)^\times \cap t D^\times t^{-1}\big)/F^\times$, u^N is in $F^\times$ and hence $w^N = u^N/(u')^N = 1$. Let l be the prime dividing $\boldsymbol{N}(v_1)$. 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 $u \in F$. □

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 $v \nmid 2$ acts trivially and $U \cap (\mathbb{A}_F^f)^\times$ acts by ψ, to the category of finitely generated A-modules.

Let S be a finite set of places of F containing Σ, all the places above 2, all the infinite places and all the places v for which U_v is not maximal. Let $\mathbb{T}_{S,A}^{\mathrm{univ}} = A[T_v, S_v]_{v\notin S}$ 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_σ-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 $\mathbb{T}_{S,A}^{\mathrm{univ}}$-module with S_v acting via the double coset $U_v\left(\begin{smallmatrix}\varpi_v & 0\\ 0 & \varpi_v\end{smallmatrix}\right)U_v$ and T_v acting via the double coset $U_v\left(\begin{smallmatrix}\varpi_v & 0\\ 0 & 1\end{smallmatrix}\right)U_v$. We write $\mathbb{T}_{\sigma,\psi}(U, A)$ or $\mathbb{T}_{\sigma,\psi}(U)$ for the image of $\mathbb{T}_{S,A}^{\mathrm{univ}}$ 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 $\bar{F}$ of F and let $G_{F,S}$ be the Galois group of the maximal extension of F in $\bar{F}$ which is unramified outside S. We view ψ as a character of $G_{F,S}$ via global class field theory, normalized so that uniformizers are mapped to geometric Frobenii. Let $\chi_{\mathrm{cyc}} : G_{F,S} \to \mathcal{O}^\times$ be the global 2-adic cyclotomic character. We note that χ_{cyc} is trivial modulo ϖ. For each place v of F, including the infinite places, we fix an embedding $\bar{F} \hookrightarrow \bar{F}_v$. This induces a continuous homomorphism of Galois groups $G_{F_v} := \mathrm{Gal}(\bar{F}_v/F_v) \to G_{F,S}$. We fix a continuous representation

$$\bar{\rho} : G_{F,S} \to \mathrm{GL}_2(k)$$

and assume that the following conditions hold:

- The image of $\bar{\rho}$ is nonsolvable.
- $\bar{\rho}$ is unramified at all finite places $v \nmid 2$.
- If $v \in S$ is a finite place, $v \notin \Sigma$, and $v \nmid 2$, then the eigenvalues of $\bar{\rho}(\mathrm{Frob}_v)$ are distinct.
- If $v \in \Sigma$ then the eigenvalues of $\bar{\rho}(\mathrm{Frob}_v)$ are equal.
- $\det \bar{\rho} \equiv \psi\chi_{\mathrm{cyc}} \pmod{\varpi}$.
- If $v \in S$ is a finite place, $v \notin \Sigma$, and $v \nmid 2$, then

$$U_v = \left\{g \in \mathrm{GL}_2(\mathcal{O}_{F_v}) : g \equiv \left(\begin{smallmatrix}1 & *\\ 0 & 1\end{smallmatrix}\right) \pmod{\varpi_v}\right\}$$

  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 $\bar{\rho}$. For each $v \in S$ let $R_v^\square$ be the framed deformation ring of $\bar{\rho}|_{G_{F_v}}$ and let $R_v^{\psi,\square}$ be the quotient of $R_v^\square$ parametrizing lifts with determinant $\psi\chi_{\mathrm{cyc}}$. We will now introduce some quotients of $R_v^{\psi,\square}$.

For $v \mid 2$ let τ_v be a 2-dimensional representation of the inertia group I_v with an open kernel, and let $\boldsymbol{w}_v = (a_v, b_v)$ be a pair of integers with $b_v > a_v$. Let $\sigma(\tau_v)$ be any absolutely irreducible representation of $U_v = \mathrm{GL}_2(\mathbb{Z}_2)$ with the property that, for all irreducible infinite-dimensional smooth representations π of $\mathrm{GL}_2(\mathbb{Q}_2)$, $\mathrm{Hom}_{U_v}(\sigma(\tau_v), \pi) \neq 0$ if and only if the restriction to I_v of the Weil–Deligne representation $\mathrm{LL}(\pi)$ associated to π via the local Langlands correspondence is isomorphic to τ. The existence of such $\sigma(\tau_v)$ is shown in [Henniart 2002], where it is also shown that if $\mathrm{Hom}_{U_v}(\sigma(\tau_v), \pi) \neq 0$ then it is one-dimensional. We choose a U_v-invariant $\mathcal{O}$-lattice $\sigma(\tau_v)^0$ in $\sigma(\tau_v)$ and let

$$\sigma_v := \sigma(\tau_v)^0 \otimes_{\mathcal{O}} \mathrm{Sym}^{b_v - a_v - 1} \mathcal{O}^2 \otimes_{\mathcal{O}} \det{}^{a_v} . \tag{30}$$

We let $R_v^{\psi,\square}(\sigma_v)$ be the reduced, $\mathcal{O}$-flat quotient of $R_v^{\psi,\square}$ parametrizing potentially semistable lifts with Hodge–Tate weights $\boldsymbol{w}_v$ and inertial type τ_v. This ring is denoted by $R^{\psi,\square}(\boldsymbol{w}, \tau)$ in the local part of the paper.

We similarly define $\sigma^{\mathrm{cr}}(\tau_v)$ by additionally requiring that $\mathrm{Hom}_{U_v}(\sigma^{\mathrm{cr}}(\tau_v), \pi) \neq 0$ if and only if the monodromy operator N in $\mathrm{LL}(\pi)$ is zero and $\mathrm{LL}(\pi)|_{I_v} \cong \tau_v$. In this case we let

$$\sigma_v := \sigma^{\mathrm{cr}}(\tau_v)^0 \otimes_{\mathcal{O}} \mathrm{Sym}^{b_v - a_v - 1} \mathcal{O}^2 \otimes_{\mathcal{O}} \det{}^{a_v}. \tag{31}$$

We let $R_v^{\psi,\square}(\sigma_v)$ be the quotient of $R_v^{\psi,\square}$ parametrizing potentially crystalline lifts with Hodge–Tate weights $\boldsymbol{w}_v$ and inertial type τ_v. This ring is denoted by $R^{\psi,\square,\mathrm{cr}}(\boldsymbol{w}, \tau)$ 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 $\sigma(\tau_v)^0 \otimes_{\mathcal{O}} k$ is the same in all cases.

If v is infinite then $R_v^{\psi,\square}$ is a domain of Krull dimension 3 and $R_v^{\psi,\square}\big[\frac{1}{2}\big]$ is regular [Kisin 2009b, Proposition 2.5.6; Khare and Wintenberger 2009b, Proposition 3.1].

If v is finite, $\bar{\rho}$ is unramified at v and $\bar{\rho}(\mathrm{Frob}_v)$ has distinct Frobenius eigenvalues, then $R_v^{\psi,\square}$ has Krull dimension 4 and $R_v^{\psi,\square}\big[\frac{1}{2}\big]$ 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 $\bar{\rho}(\mathrm{Frob}_v)$ are distinct, $\bar{\rho}$ cannot have a lift of the form $\gamma \oplus \gamma\chi_{\mathrm{cyc}}$. It follows from the proof of [Kisin 2009b, Proposition 2.5.4] that different irreducible components of $R_v^{\psi,\square}\big[\frac{1}{2}\big]$ do not intersect.

If v is finite, ψ and $\bar{\rho}$ are unramified at v and $\bar{\rho}(\mathrm{Frob}_v)$ has equal eigenvalues, then for an unramified character $\gamma : G_{F_v} \to \mathcal{O}^\times$ such that $\gamma^2 = \psi|_{G_{F_v}}$ we let $R_v^{\psi,\square}(\gamma)$ be a reduced $\mathcal{O}$-torsion-free quotient of $R_v^{\psi,\square}$ with the property that if L'/L is a finite extension then a map $x : R_v^{\psi,\square} \to L'$ factors through $R_v^{\psi,\square}(\gamma)$ if and only if V_x is isomorphic to $\left(\begin{smallmatrix} \gamma\chi_{\mathrm{cyc}} & * \\ 0 & \gamma \end{smallmatrix}\right)$. 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 $R_v^{\psi,\square}(\gamma)\big[\frac{1}{2}\big]$ is regular. If L is large enough then there are precisely two such characters, which we denote by γ_1

and γ_2. We let $\bar{R}_v^{\psi,\square}$ be the image of

$$R_v^{\psi,\square} \to R_v^{\psi,\square}(\gamma_1)\bigl[\tfrac{1}{2}\bigr] \times R_v^{\psi,\square}(\gamma_2)\bigl[\tfrac{1}{2}\bigr].$$

Then $\bar{R}_v^{\psi,\square}$ 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 $\bar{R}_v^{\psi,\square}$ if and only if V_x is isomorphic to $\left(\begin{smallmatrix} \gamma\chi_{\mathrm{cyc}} & * \\ 0 & \gamma \end{smallmatrix}\right)$ for an unramified character γ. Moreover,

$$\bar{R}_v^{\psi,\square}\bigl[\tfrac{1}{2}\bigr] \cong R_v^{\psi,\square}(\gamma_1)\bigl[\tfrac{1}{2}\bigr] \times R_v^{\psi,\square}(\gamma_2)\bigl[\tfrac{1}{2}\bigr].$$

Thus $\bar{R}_v^{\psi,\square}\bigl[\tfrac{1}{2}\bigr]$ is regular and equidimensional and the Krull dimension of $\bar{R}_v^{\psi,\square}$ is 4.

We let

$$R_S^{\square} = \widehat{\bigotimes_{v\in S}} R_v^{\square}, \qquad R_S^{\psi,\square} = \widehat{\bigotimes_{v\in S}} R_v^{\psi,\square}, \qquad \sigma := \widehat{\bigotimes_{v\mid 2}} \sigma_v,$$

and

$$R_S^{\psi,\square}(\sigma) := \widehat{\bigotimes_{v\mid 2}} R_v^{\psi,\square}(\sigma_v) \widehat{\bigotimes_{v\in\Sigma}} \bar{R}_v^{\psi,\square} \widehat{\bigotimes_{\substack{v\in S\setminus\Sigma \\ v\nmid 2\infty}}} R_v^{\psi,\square} \widehat{\bigotimes_{v\mid\infty}} R_v^{\psi,\square}.$$

It follows from above that $R_S^{\psi,\square}(\sigma)$ is equidimensional of Krull dimension equal to

$$1+4\sum_{v\mid 2} 1 + 3|\Sigma| + 3\Bigl(|S|-|\Sigma|-\sum_{v\mid 2} 1 - \sum_{v\mid\infty} 1\Bigr) + 2\sum_{v\mid\infty} 1 = 1+3|S|. \tag{32}$$

3B2. *Global deformation rings.* Since $\bar{\rho}$ is assumed to have nonsolvable image, $\bar{\rho}$ is absolutely irreducible. We define $R_{F,S}^{\psi}$ to be the quotient of the universal deformation ring of $\bar{\rho}$ parametrizing deformations with determinant $\psi\chi_{\mathrm{cyc}}$. 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}^{ψ} in the same way by viewing $\bar{\rho}$ as a representation of G_{F,S_Q}.

Denote by $R_{F,S_Q}^{\psi,\square}$ 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 $\bar{\rho}$ to A with determinant $\psi\chi_{\mathrm{cyc}}$ and β_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 $R_v^{\psi,\square} \to R_{F,S_Q}^{\psi,\square}$ for every $v \in S$ and hence a homomorphism of $\mathcal{O}$-algebras $R_S^{\psi,\square} \to R_{F,S_Q}^{\psi,\square}$.

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 $Q_0 = \varnothing$, so that $S_{Q_n} = S$ for $n = 0$. Let $U_{Q_n} = \prod_v (U_{Q_n})_v$ be a compact open subgroup of $(D \otimes_F \mathbb{A}_F^f)^\times$ 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 $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}_{S,\mathcal{O}}^{\mathrm{univ}}$ such that the residue field is k, T_v is mapped to $\operatorname{tr}\bar{\rho}(\mathrm{Frob}_v)$ and S_v is mapped to the image of $\psi(\mathrm{Frob}_v)$ in k for all $v \notin S$. We define

$\mathfrak{m}_{Q_n}$ in $\mathbb{T}^{\mathrm{univ}}_{S_{Q_n},\mathcal{O}}$ in the same manner. Let $\sigma = \bigotimes_{v|2} \sigma_v$, where each σ_v is given by either (30) or (31). We assume that $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$. Then for all $n \geq 0$ there is a surjective homomorphism of $\mathcal{O}$-algebras $R^{\psi}_{F,S_{Q_n}} \to \mathbb{T}_{\sigma,\psi}(U_{Q_n})_{\mathfrak{m}_{Q_n}}$ such that for all $v \notin S_{Q_n}$ the trace of Frob_v of the tautological $R^{\psi}_{F,S_{Q_n}}$-representation of $G_{F,S_{Q_n}}$ is mapped to T_v. Set

$$M_n(\sigma) = R^{\psi,\square}_{F,S_{Q_n}} \otimes_{R^{\psi}_{F,S_{Q_n}}} S_{\sigma,\psi}(U_{Q_n}, \mathcal{O})_{\mathfrak{m}_{Q_n}},$$

with the convention that if $n = 0$ then $Q_n = \varnothing$, $S_{Q_n} = S$, $\mathfrak{m}_{Q_n} = \mathfrak{m}$, so that

$$M_0(\sigma) = R^{\psi,\square}_{F,S} \otimes_{R^{\psi}_{F,S}} S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}}.$$

It follows from the local-global compatibility of Jacquet–Langlands and Langlands correspondences that the action of $R^{\psi,\square}_{F,S_{Q_n}}$ on $M_n(\sigma)$ factors through the quotient

$$R^{\psi,\square}_{F,S_{Q_n}}(\sigma) := R^{\psi,\square}_{S}(\sigma) \otimes_{R^{\psi,\square}_{S}} R^{\psi,\square}_{F,S_{Q_n}}.$$

Let $h = \dim_k H^1(G_{F,S}, \mathrm{ad}\bar{\rho}) - 2 = |Q_n|$. Let $\mathfrak{a}_\infty$ denote the ideal of $\mathcal{O}[\![y_1, \dots, y_h]\!]$ generated by $(y_1, \dots, y_h)$. Since $R^{\psi,\square}_{F,S_{Q_n}}$ is formally smooth over $R^{\psi}_{F,S_{Q_n}}$ of relative dimension $j = 4|S| - 1$ we may choose an identification

$$R^{\psi,\square}_{F,S_{Q_n}} = R^{\psi}_{F,S_{Q_n}}[\![y_{h+1}, \dots, y_{h+j}]\!]$$

and regard $M_n(\sigma)$ as an $\mathcal{O}[\![y_1, \dots, y_{h+j}]\!]$-module. This allows us to consider $R^{\psi}_{F,S_{Q_n}}$ as an $R^{\psi,\square}_{S}$-algebra via the map $R^{\psi,\square}_{S} \to R^{\psi,\square}_{F,S_{Q_n}}/(y_{h+1}, \dots, y_{h+j}) = R^{\psi}_{F,S_{Q_n}}$. We let

$$R^{\psi}_{F,S_{Q_n}}(\sigma) := R^{\psi,\square}_{S}(\sigma) \otimes_{R^{\psi,\square}_{S}} R^{\psi}_{F,S_{Q_n}}.$$

Let $g = 2|Q_n|+1$ and $t = 2-|S|+|Q_n|$ and let $\widehat{\mathbb{G}}_m$ 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 $\mathcal{O}[\![y_1, \dots, y_{h+j}]\!]$-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

$$R^{\psi,\square}_{S}(\sigma)[\![x_1, \dots, x_g]\!] \twoheadrightarrow R'_\infty(\sigma) \twoheadrightarrow R_\infty(\sigma).$$

(P2) There is an isomorphism of $R^{\psi,\square}_{S}(\sigma)$-algebras

$$R_\infty(\sigma)/\mathfrak{a}_\infty R_\infty(\sigma) \xrightarrow{\cong} R^{\psi,\square}_{F,S}(\sigma)$$

and an isomorphism of $R^{\psi,\square}_{F,S}(\sigma)$-modules

$$M_\infty(\sigma)/\mathfrak{a}_\infty M_\infty(\sigma) \xrightarrow{\cong} M_0(\sigma).$$

(P3) $M_\infty(\sigma)$ is finite flat over $\mathcal{O}[\![y_1, \dots, y_{h+j}]\!]$.

(P4) $\operatorname{Spf} R'_\infty(\sigma)$ is equipped with a free action of $(\widehat{\mathbb{G}}_m)^t$, and a $(\widehat{\mathbb{G}}_m)^t$-equivariant morphism $\delta : \operatorname{Spf} R'_\infty(\sigma) \to (\widehat{\mathbb{G}}_m)^t$, where $(\widehat{\mathbb{G}}_m)^t$ acts on itself by the square of the identity map.

(P5) We have $\delta^{-1}(1) = \operatorname{Spf} R_\infty(\sigma) \subset \operatorname{Spf} R'_\infty(\sigma)$, and the induced action of $(\widehat{\mathbb{G}}_m[2])^t$ on $\operatorname{Spf} R_\infty(\sigma)$ 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 complete local noetherian $\mathcal{O}$-algebra $\big(R^{\mathrm{inv}}_\infty(\sigma), \mathfrak{m}^{\mathrm{inv}}_\sigma\big)$ with residue field k such that $\operatorname{Spf} R^{\mathrm{inv}}_\infty(\sigma) = \operatorname{Spf} R'_\infty(\sigma)/(\widehat{\mathbb{G}}_m)^t$. Moreover,

$$R'_\infty(\sigma) = R^{\mathrm{inv}}_\infty(\sigma)\,\widehat{\otimes}_{\mathcal{O}}\,\mathcal{O}[\![\mathbb{Z}_2^t]\!] \cong R^{\mathrm{inv}}_\infty(\sigma)[\![z_1, \dots, z_t]\!]. \tag{33}$$

This implies that

$$\dim R'_\infty(\sigma) = \dim R^{\mathrm{inv}}_\infty(\sigma) + t, \quad e(R'_\infty(\sigma)/\varpi) = e(R^{\mathrm{inv}}_\infty(\sigma)/\varpi). \tag{34}$$

Lemma 3.3. *There are $a_1, \dots, a_t \in \mathfrak{m}^{\mathrm{inv}}_\sigma$ such that*

$$R_\infty(\sigma) \cong \frac{R^{\mathrm{inv}}_\infty(\sigma)[\![z_1]\!]}{((1+z_1)^2 - (1+a_1))} \otimes_{R^{\mathrm{inv}}_\infty(\sigma)} \cdots \otimes_{R^{\mathrm{inv}}_\infty(\sigma)} \frac{R^{\mathrm{inv}}_\infty(\sigma)[\![z_t]\!]}{((1+z_t)^2 - (1+a_t))}. \tag{35}$$

In particular, $R_\infty(\sigma)$ is a free $R^{\mathrm{inv}}_\infty(\sigma)$-module of rank 2^t.

Proof. It follows from [Khare and Wintenberger 2009b, Lemma 9.4] that $\operatorname{Spf} R_\infty(\sigma)$ is a $(\widehat{\mathbb{G}}_m[2])^t$-torsor over $\operatorname{Spf} R^{\mathrm{inv}}_\infty(\sigma)$. The assertion follows from [SGA 3_{II} 1970, Exposé VIII, Proposition 4.1]. □

Lemma 3.4. *Let $\mathfrak{p} \in \operatorname{Spec} R^{\mathrm{inv}}_\infty(\sigma)$. The group $(\widehat{\mathbb{G}}_m[2])^t(\mathcal{O})$ acts transitively on the set of prime ideals of $R_\infty(\sigma)$ lying above $\mathfrak{p}$.*

Proof. Let us write X for $\operatorname{Spf} R_\infty(\sigma)$ and G for $(\widehat{\mathbb{G}}_m[2])^t$. 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 $(g, \mathrm{id}_{R_\infty(\sigma)})$. The map $g \mapsto \phi_g$ induces a homomorphism of groups $G(\mathcal{O}) \to \operatorname{Aut}(R_\infty(\sigma))$. Explicitly, if $g = (\epsilon_1, \dots, \epsilon_t)$, where ϵ_i is either 1 or -1, then ϕ_g is $R^{\mathrm{inv}}_\infty(\sigma)$-linear and maps $1 + z_i$ to $\epsilon_i(1 + z_i)$ for $1 \le i \le t$. It follows from (35) that $G(\mathcal{O})$ acts transitively on the set of maximal ideals of $\kappa(\mathfrak{p}) \otimes_{R^{\mathrm{inv}}_\infty(\sigma)} R_\infty(\sigma)$. □

Lemma 3.5. *The support of $M_\infty(\sigma)$ in* $\operatorname{Spec} R_\infty(\sigma)$ *is a union of irreducible components. The Krull dimension of* $\operatorname{Spec} R_\infty(\sigma)$ *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 $\dim R_\infty(\sigma) \le \dim R_S^{\psi,\square}(\sigma) + g - t = 3|S|+1+g-t = h+j+1$. $\square$

Lemma 3.6. $e(R'_\infty(\sigma)/\varpi) \le e\big(R_S^{\psi,\square}(\sigma)/\varpi\big)$.

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 $R_S^{\psi,\square}(\sigma)[\![x_1, \dots, x_g]\!]$ by (32). The surjection in (P1) above implies that

$$e(R'_\infty(\sigma)/\varpi) \le e\big(R_S^{\psi,\square}(\sigma)[\![x_1, \dots, x_g]\!]/\varpi\big) = e\big(R_S^{\psi,\square}(\sigma)/\varpi\big). \qquad \square$$

Lemma 3.7. *If $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}}$ is supported on a closed point $\mathfrak{n} \in \operatorname{Spec} R_S^{\psi,\square}(\sigma)\big[\frac{1}{2}\big]$ then the localization $R_S^{\psi,\square}(\sigma)_{\mathfrak{n}}$ is a regular ring.*

Proof. Since the rings $R_v^{\square}\big[\frac{1}{2}\big]$ are regular for all $v \nmid 2$ it is enough to show that $\mathfrak{n}$ defines a regular point in $\operatorname{Spec} R_v^{\psi,\square}(\sigma)$ for all $v \mid 2$. 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 $\operatorname{Spec} R_v^{\psi,\square}(\sigma)$, but this would violate the weight–monodromy conjecture for $\mathrm{WD}(\rho_{\mathfrak{n}}|_{G_{F_v}})$; see [Gee and Kisin 2014] for details. $\square$

Lemma 3.8. *If $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}}$ is supported on a closed point $\mathfrak{n} \in \operatorname{Spec} R_\infty(\sigma)\big[\frac{1}{2}\big]$ then the localization $R_\infty(\sigma)_{\mathfrak{n}}$ is a regular ring.*

Proof. Let $\mathfrak{n}_S$ be the image of $\mathfrak{n}$ in $\operatorname{Spec} R_S^{\psi,\square}[\![x_1, \dots, x_g]\!]$, let $\mathfrak{n}'$ be the image of $\mathfrak{n}$ in $\operatorname{Spec} R'_\infty(\sigma)$ via the maps in (P1), and let $\mathfrak{n}^{\mathrm{inv}}$ be the image of $\mathfrak{n}$ in $\operatorname{Spec} R_\infty^{\mathrm{inv}}(\sigma)$ via (35). It follows from Lemma 3.7 that $R_S^{\psi,\square}(\sigma)[\![x_1, \dots, x_g]\!]_{\mathfrak{n}_S}$ is a regular ring. If the map

$$R_S^{\psi,\square}(\sigma)[\![x_1, \dots, x_g]\!]_{\mathfrak{n}_S} \twoheadrightarrow R'_\infty(\sigma)_{\mathfrak{n}'} \tag{36}$$

is an isomorphism, then $R'_\infty(\sigma)_{\mathfrak{n}'}$ is a regular ring. We may assume that L is sufficiently large, so that using (33) we may write $\mathfrak{n}' = (\mathfrak{n}^{\mathrm{inv}}, z_1 - a_1, \dots, z_t - a_t)$ with $a_i \in \varpi\mathcal{O}$ for $1 \le i \le t$. The images of $z_1 - a_1, \dots, z_t - a_t$ in $\mathfrak{n}'/(\mathfrak{n}')^2$ are linearly independent. Since

$$R_\infty^{\mathrm{inv}}(\sigma)_{\mathfrak{n}^{\mathrm{inv}}} \cong R'_\infty(\sigma)_{\mathfrak{n}'}/(z_1 - a_1, \dots, z_t - a_t)R'_\infty(\sigma)_{\mathfrak{n}'},$$

we deduce that $R_\infty^{\mathrm{inv}}(\sigma)_{\mathfrak{n}^{\mathrm{inv}}}$ is regular. It follows from (35) that the map

$$R_\infty^{\mathrm{inv}}(\sigma)\big[\tfrac{1}{2}\big] \to R_\infty(\sigma)\big[\tfrac{1}{2}\big]$$

is étale. Hence $R_\infty(\sigma)_{\mathfrak{n}}$ is a regular ring.

If (36) is not an isomorphism then the dimension of the quotient must decrease. This leads to the inequality $\dim R_\infty(\sigma)_\mathfrak{n} < \dim R_\infty(\sigma) - 1$. Since $M_\infty(\sigma)$ is a Cohen–Macaulay module, as follows from (P3), its support cannot contain embedded components, hence $\dim M_\infty(\sigma)_\mathfrak{n} = \dim M_\infty(\sigma) - 1$. This leads to a contradiction, as $M_\infty(\sigma)_\mathfrak{n}$ is a finitely generated $R_\infty(\sigma)_\mathfrak{n}$-module. □

Lemma 3.9. *Let A be a local noetherian ring and let $(x_1, \dots, x_d)$ be a system of parameters of A. If A is equidimensional then every irreducible component of A contains a closed point of $(A/(x_2, \dots, x_d))[1/x_1]$.*

Proof. Let $\mathfrak{p}$ be an irreducible component of A. If $A/(\mathfrak{p}, x_2, \dots, x_d)[1/x_1]$ is zero then x_1 is nilpotent in $A/(\mathfrak{p}, x_2, \dots, x_d)$. Since $(x_1, \dots, x_d)$ is a system of parameters of A, we conclude that $A/(\mathfrak{p}, x_2, \dots, x_d)$ is zero dimensional, which implies that $\dim A/\mathfrak{p} \le d - 1$, contradicting equidimensionality of A. □

Lemma 3.10. *There is an integer r, independent of σ and the choices made in the patching process, such that for all $\mathfrak{p} \in \operatorname{Spec} R_\infty(\sigma)$ in the support of $M_\infty(\sigma)$ we have*

$$\dim_{\kappa(\mathfrak{p})} M_\infty(\sigma) \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{p}) \ge r,$$

with equality if $\mathfrak{p}$ is a minimal prime of $R_\infty(\sigma)$ in the support of $M_\infty(\sigma)$.

Proof. Let $\mathfrak{q}$ be a minimal prime of $R_\infty(\sigma)$ in the support of $M_\infty(\sigma)$. It is enough to show that $\dim_{\kappa(\mathfrak{q})} M_\infty(\sigma) \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{q})$ is independent of $\mathfrak{q}$ and σ. Since

$$M_\infty(\sigma)/(y_1, \dots, y_{h+j})M_\infty(\sigma) \cong S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m}$$

and $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m}$ is a finitely generated $\mathcal{O}$-module, $y_1, \dots, y_{h+j}, \varpi$ is a system of parameters for $R_\infty(\sigma)/\mathfrak{q}$ and it follows from Lemma 3.9 that there is a maximal ideal $\mathfrak{n}$ of $R_\infty(\sigma)\left[\frac{1}{2}\right]$, contained in $V(\mathfrak{q})$, such that $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{n} \neq 0$. It follows from (P3) that $M_\infty(\sigma)$ is a Cohen–Macaulay module. The same holds for the localization at $\mathfrak{n}$. Since $R_\infty(\sigma)_\mathfrak{n}$ is a regular ring by Lemma 3.8, a standard argument with the Auslander–Buchsbaum theorem shows that $M_\infty(\sigma)_\mathfrak{n}$ is a free $R_\infty(\sigma)_\mathfrak{n}$-module. By localizing further at $\mathfrak{q}$ we deduce that

$$\begin{aligned}\dim_{\kappa(\mathfrak{q})} M_\infty(\sigma) \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{q}) &= \dim_{\kappa(\mathfrak{n})} M_\infty(\sigma) \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{n}) \\ &= \dim_{\kappa(\mathfrak{n})} S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{n}). \qquad (37)\end{aligned}$$

So it is enough show that $\dim_{\kappa(\mathfrak{n})} S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{n})$ is independent of $\mathfrak{n}$ and σ. The action of $R_\infty(\sigma)$ on $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m}$ factors through the action of the Hecke algebra $\mathbb{T}_{\sigma,\psi}(U)$, which is reduced. Thus $\mathbb{T}_{\sigma,\psi}(U)\left[\frac{1}{2}\right]$ is a product of finite field extensions of L and we have

$$S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_{R_\infty(\sigma)} \kappa(\mathfrak{n}) = S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{n} = (S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_\mathcal{O} L)[\mathfrak{n}].$$

Let $\pi = \otimes'_v \pi_v$ be the automorphic representation of $(D \otimes_F \mathbb{A}_F^f)^\times$ corresponding to $f^D \in (S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_\mathcal{O} L)[\mathfrak{n}]$. 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

$$\dim_L (S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \otimes_\mathcal{O} L)[\mathfrak{n}] = \prod_{\substack{v \in S \\ v \nmid 2\infty}} \dim_L \pi_v^{U_v} \prod_{v \mid 2} \dim_L \operatorname{Hom}_{U_v}(\sigma(\tau_v), \pi_v).$$

We claim that the right-hand side of the above equation is equal to $2^{|S \setminus (\Sigma \cup \{v | 2\infty\})|}$. The claim will follow from the local-global compatibility of Langlands and Jacquet–Langlands correspondences. Let $\rho_\mathfrak{n}$ be the representation of $G_{F,S}$ corresponding to $\mathfrak{n}$, considered as a maximal ideal of $R^\psi_{F,S}(\sigma)\big[\tfrac{1}{2}\big]$. If $v \mid 2$ then the results of [Henniart 2002] imply that $\dim_L \operatorname{Hom}_{U_v}(\sigma(\tau_v), \pi_v) = 1$. If $v \in \Sigma$ then π_v is an unramified character of $D_v^\times$, and hence $\dim_L \pi_v^{U_v} = 1$. If $v \in S$, $v \nmid 2\infty$ and $v \notin \Sigma$ then D is split at v, $\bar\rho|_{G_{F_v}}$ is unramified and $\bar\rho(\mathrm{Frob}_v)$ has distinct eigenvalues. This implies that $\rho_\mathfrak{n}|_{G_{F_v}}$ is an extension of distinct tamely ramified characters ψ_1, ψ_2 such that $\psi_1\psi_2^{-1} \neq \chi_{\mathrm{cyc}}^{\pm 1}$. We deduce that π_v is a tamely ramified principal series. Since U_v is equal to the subgroup of unipotent upper-triangular matrices modulo ϖ_v in this case, we deduce that $\dim_L \pi_v^{U_v} = 2$. $\square$

Lemma 3.11. *There is an integer r, independent of σ and the choices made in the patching process, such that for all minimal primes $\mathfrak{p}$ of $R_\infty^{\mathrm{inv}}(\sigma)$ in the support of $M_\infty(\sigma)$ we have*

$$\dim_{\kappa(\mathfrak{p})} M_\infty(\sigma) \otimes_{R_\infty^{\mathrm{inv}}(\sigma)} \kappa(\mathfrak{p}) = 2^t r.$$

Proof. To ease the notation, let us drop σ from it in this proof. Since $\mathfrak{p}$ is minimal, it is an associated prime and so M_∞ will contain $R_\infty^{\mathrm{inv}}/\mathfrak{p}$ as a submodule. Since M_∞ is $\mathcal{O}$-torsion-free, this implies that the quotient field $\kappa(\mathfrak{p})$ has characteristic 0. It follows from (35) that $R_\infty \otimes_{R_\infty^{\mathrm{inv}}} \kappa(\mathfrak{p})$ is étale over $\kappa(\mathfrak{p})$, and so

$$R_\infty \otimes_{R_\infty^{\mathrm{inv}}} \kappa(\mathfrak{p}) \cong \prod_\mathfrak{q} \kappa(\mathfrak{q}),$$

where the product is taken over all prime ideals $\mathfrak{q}$ of R_∞ such that $\mathfrak{q} \cap R_\infty^{\mathrm{inv}} = \mathfrak{p}$. From this we get

$$\dim_{\kappa(\mathfrak{p})} M_\infty \otimes_{R_\infty^{\mathrm{inv}}} \kappa(\mathfrak{p}) = \sum_\mathfrak{q} [\kappa(\mathfrak{q}) : \kappa(\mathfrak{p})] \dim_{\kappa(\mathfrak{q})} M_\infty \otimes_{R_\infty} \kappa(\mathfrak{q}).$$

It follows from Lemma 3.4 and (P5) that all $\mathfrak{q}$ appearing in the sum lie in the support of M_∞. Lemma 3.10 implies that $\dim_{\kappa(\mathfrak{q})} M_\infty \otimes_{R_\infty} \kappa(\mathfrak{q}) = r$. Thus

$$\dim_{\kappa(\mathfrak{p})} M_\infty \otimes_{R_\infty^{\mathrm{inv}}} \kappa(\mathfrak{p}) = r \dim_{\kappa(\mathfrak{p})} R_\infty \otimes_{R_\infty^{\mathrm{inv}}} \kappa(\mathfrak{p}) = r 2^t,$$

where the last equality follows from Lemma 3.3. $\square$

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 $\ell_{A_{\mathfrak{q}}}(M_{\mathfrak{q}}) \le \ell_{A_{\mathfrak{q}}}(N_{\mathfrak{q}})$ for all $\mathfrak{q} \in \operatorname{Spec} A$ with $\dim A/\mathfrak{q} = d$ then*

$$e(M/xM, A/xA) \le e(N/xN, A/xA).$$

If $\ell_{A_{\mathfrak{q}}}(M_{\mathfrak{q}}) = \ell_{A_{\mathfrak{q}}}(N_{\mathfrak{q}})$ for all $\mathfrak{q} \in \operatorname{Spec} A$ with $\dim A/\mathfrak{q} = d$ then

$$e(M/xM, A/xA) = e(N/xN, A/xA).$$

Proof. It follows from Proposition 2.2.13 in [Emerton and Gee 2014] that

$$e(M/xM, A/xA) = \sum_{\mathfrak{q}} \ell_{A_{\mathfrak{q}}}(M_{\mathfrak{q}}) e(A/(\mathfrak{q}, x)), \tag{38}$$

where the sum is taken over all primes $\mathfrak{q}$ in the support of M such that $\dim A/\mathfrak{q} = d$. The above formula implies both assertions. □

Lemma 3.13. $e\big(M_\infty(\sigma)/\varpi, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) \le 2^t r e\big(R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big)$.

Proof. Let $\mathbb{T}^{\mathrm{inv}}_\infty(\sigma)$ be the image of $R^{\mathrm{inv}}_\infty(\sigma)$ in $\operatorname{End}_{\mathcal{O}}(M_\infty(\sigma))$. Then

$$e\big(\mathbb{T}^{\mathrm{inv}}_\infty(\sigma)/\varpi, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) \le e\big(R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big).$$

If $\mathfrak{q}$ is a minimal prime of $R^{\mathrm{inv}}_\infty(\sigma)$ in the support of $M_\infty(\sigma)$ then it follows from Lemma 3.11 that there are surjections $\mathbb{T}^{\mathrm{inv}}_\infty(\sigma)^{\oplus 2^t r}_{\mathfrak{q}} \twoheadrightarrow M_\infty(\sigma)_{\mathfrak{q}}$. Thus $\ell(M_\infty(\sigma)_{\mathfrak{q}}) \le 2^t r \ell\big(\mathbb{T}^{\mathrm{inv}}_\infty(\sigma)_{\mathfrak{q}}\big)$. The assertion follows from Lemma 3.12 applied with $x = \varpi$, $M = M_\infty(\sigma)$ and $N = \mathbb{T}^{\mathrm{inv}}_\infty(\sigma)^{\oplus 2^t r}$. □

Lemma 3.14. *If the support of $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}}$ meets every irreducible component of $R^{\psi,\square}_S(\sigma)$ then the following hold:*

(i) *$R^{\psi,\square}_S(\sigma)[[x_1, \dots, x_g]] \twoheadrightarrow R'_\infty(\sigma)$ is an isomorphism.*

(ii) *$R^{\mathrm{inv}}_\infty(\sigma)$ is reduced, equidimensional and $\mathcal{O}$-flat.*

(iii) *$R_\infty(\sigma)$ is reduced, equidimensional and $\mathcal{O}$-flat.*

(iv) *The support of $M_\infty(\sigma)$ meets every irreducible component of $R_\infty(\sigma)$.*

(v) $2^t r e\big(R^{\psi,\square}_S(\sigma)/\varpi\big) = e\big(M_\infty(\sigma)/\varpi, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big)$.

Proof. Since $R^{\psi,\square}_S(\sigma)[[x_1, \dots, x_g]]$ is reduced and equidimensional and has the same dimension as $R'_\infty(\sigma)$, to prove (i) it is enough to show that $R'_\infty(\sigma)_{\mathfrak{q}} \ne 0$ for every irreducible component $V(\mathfrak{q})$ of $\operatorname{Spec} R^{\psi,\square}_S(\sigma)[[x_1, \dots, x_g]]$. Since the diagram

$$\begin{tikzcd}
R^{\psi,\square}_S(\sigma)[[x_1, \dots, x_g]] \arrow[r] & R_\infty(\sigma) \arrow[d] \\
R^{\psi,\square}_S(\sigma) \arrow[u] \arrow[r] & R^{\psi}_{F,S}(\sigma)
\end{tikzcd}$$

commutes and the support of $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ meets every irreducible component of Spec $R_S^{\psi,\square}$, $V(\mathfrak{q})$ will contain a maximal ideal $\mathfrak{n}_S$ of $R_S^{\psi,\square}(\sigma)[\![x_1,\dots,x_g]\!][\frac{1}{2}]$, which lies in the support of $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$. It follows from the proof of Lemma 3.8 that (36) is an isomorphism in this case. Thus $R'_\infty(\sigma)_{\mathfrak{q}}\neq 0$.

From part (i) we deduce that $R'_\infty(\sigma)$ is reduced, equidimensional and $\mathcal{O}$-flat. It follows from (33) that the same holds for $R_\infty^{\mathrm{inv}}(\sigma)$. Since $R_\infty(\sigma)$ is a free $R_\infty^{\mathrm{inv}}(\sigma)$-module by Lemma 3.3, it is $\mathcal{O}$-flat. Hence, it is enough to show that $R_\infty(\sigma)[\frac{1}{2}]$ is reduced and equidimensional. It follows from Lemma 3.3 that $R_\infty(\sigma)[\frac{1}{2}]$ is étale over $R_\infty^{\mathrm{inv}}(\sigma)[\frac{1}{2}]$, 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

$$e\big(R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big)=e\big(R_S^{\psi,\square}/\varpi\big). \tag{39}$$

It follows from our assumption that the support of $M_\infty(\sigma)$ meets every irreducible component of $R_S^{\psi,\square}(\sigma)[\![x_1,\dots,x_g]\!]$. Part (i) and (33) imply that the support of $M_\infty(\sigma)$ meets every irreducible component of $R_\infty^{\mathrm{inv}}(\sigma)$. It follows from Lemma 3.4 that the group $(\widehat{\mathbb{G}}_m[2])^t(\mathcal{O})$ acts transitively on the set of irreducible components of $R_\infty(\sigma)$ lying above a given irreducible component of $R_\infty^{\mathrm{inv}}(\sigma)$. Thus for part (iii) it is enough to show that the support of $M_\infty(\sigma)$ in Spec $R_\infty(\sigma)$ is stable under the action of $(\widehat{\mathbb{G}}_m[2])^t(\mathcal{O})$. This is given by (P5) and can be proved in the same way as [Khare and Wintenberger 2009b, Lemma 9.6].

Let $V(\mathfrak{q})$ be an irreducible component of Spec $R_\infty(\sigma)$. It follows from (iii) that the localization $R_\infty(\sigma)_{\mathfrak{q}}$ is a reduced artinian ring, and hence is equal to the quotient field $\kappa(\mathfrak{q})$. Thus $M_\infty(\sigma)_{\mathfrak{q}}\cong M_\infty(\sigma)\otimes_{R_\infty(\sigma)}\kappa(\mathfrak{q})$. It follows from Lemma 3.10 that $M_\infty(\sigma)_{\mathfrak{q}}$ has length r as an $R_\infty(\sigma)_{\mathfrak{q}}$-module. By part (iv) $M_\infty(\sigma)$ is supported on every irreducible component of $R_\infty(\sigma)$, and thus the cycle of $M_\infty(\sigma)$ is equal to r times the cycle of $R_\infty(\sigma)$. Since both are $\mathcal{O}$-torsion-free, we deduce that the cycle of $M_\infty(\sigma)/\varpi$ is equal to r times the cycle of $R_\infty(\sigma)/\varpi$, which implies that

$$e\big(M_\infty(\sigma)/\varpi,R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big)=re\big(R_\infty(\sigma)/\varpi,R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big)=2^tre\big(R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big). \tag{40}$$

Part (v) follows from (39) and (40). □

Proposition 3.15. *For some $s\geq 0$ there is an isomorphism of $R_S^{\psi,\square}$-algebras*

$$R_{F,S}^{\psi,\square}\cong R_S^{\psi,\square}[\![x_1,\dots,x_{s+|S|-1}]\!]/(f_1,\dots,f_s).$$

Proof. The assertion follows from the proof of [Khare and Wintenberger 2009b, Proposition 4.5], where $s=\dim_k H^1_{\{L_v^\perp\}}(S,(\mathrm{Ad}^0)^*(1))$ 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 $R_S^{\psi,\square}(\sigma)$-algebras*

$$R_{F,S}^{\psi,\square}(\sigma)\cong R_S^{\psi,\square}(\sigma)[\![x_1,\dots,x_{s+|S|-1}]\!]/(f_1,\dots,f_s).$$

In particular, $\dim R^{\psi,\square}_{F,S}(\sigma) \geq 4|S|$ *and* $\dim R^{\psi}_{F,S}(\sigma) \geq 1$.

Proof. Since

$$R^{\psi,\square}_{F,S}(\sigma) \cong R^{\psi,\square}_{F,S} \otimes_{R^{\psi,\square}_S} R^{\psi,\square}_S(\sigma)$$

the assertion follows from Proposition 3.15. Since $\dim R^{\psi,\square}_S(\sigma) = 3|S|+1$ by (32), the isomorphism implies that

$$\dim R^{\psi,\square}_{F,S}(\sigma) \geq 3|S|+1+s+|S|-1-s = 4|S|.$$

Since $R^{\psi,\square}_{F,S}(\sigma)$ is formally smooth over $R^{\psi}_{F,S}(\sigma)$ of relative dimension $4|S|-1$, we conclude that $\dim R^{\psi}_{F,S}(\sigma) \geq 1$. $\square$

Proposition 3.17. *If* $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}} \neq 0$ *then the following are equivalent*:

(a) $2^t r e\big(R^{\psi,\square}_S(\sigma)/\varpi\big) = e\big(M_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big)$.

(b) $2^t r e\big(R^{\psi,\square}_S(\sigma)/\varpi\big) \leq e\big(M_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big)$.

(c) *the support of* $M_\infty(\sigma)$ *meets every irreducible component of* $R_\infty(\sigma)$.

(d) $R^{\psi}_{F,S}(\sigma)$ *is a finitely generated* $\mathcal{O}$*-module of rank at least* 1 *and*

$$S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{n}} \neq 0 \quad \textit{for all } \mathfrak{n} \in \text{m-Spec}\, R^{\psi}_{F,S}(\sigma)\big[\tfrac{1}{2}\big].$$

In this case any representation $\rho : G_{F,S} \to \mathrm{GL}_2(\mathcal{O})$ *corresponding to a maximal ideal of* $R^{\psi}_{F,S}(\sigma)\big[\tfrac{1}{2}\big]$ *is modular.*

Proof. Lemmas 3.6 and 3.13 and (33) imply that

$$e\big(M_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) \leq 2^t r e\big(R^{\psi,\square}_S(\sigma)/\varpi\big). \tag{41}$$

Thus (a) is equivalent to (b). Moreover, if (a) holds then the inequalities in the lemmas cited above have to be equalities. Since $R^{\psi,\square}_S(\sigma)$ is reduced and $\mathcal{O}$-torsion-free, we deduce that $R'_\infty(\sigma) \cong R^{\psi,\square}_S(\sigma)[\![x_1,\ldots,x_g]\!]$. Hence, $R'_\infty(\sigma)$ is reduced, equidimensional and $\mathcal{O}$-torsion-free. The isomorphism (33) implies that the same holds for $R^{\mathrm{inv}}_\infty(\sigma)$, which implies that $R_\infty(\sigma)$ is reduced, equidimensional, and $\mathcal{O}$-torsion-free; see the proof of Lemma 3.14. Since we have assumed (a), we have

$$2^t r e\big(R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) = e\big(M_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big). \tag{42}$$

Let $V(\mathfrak{q}_1),\ldots,V(\mathfrak{q}_m)$ be the irreducible components of the support of $M_\infty(\sigma)$ in $\operatorname{Spec} R_\infty(\sigma)$. Since $R_\infty(\sigma)$ is reduced, if $V(\mathfrak{q})$ is an irreducible component of $\operatorname{Spec} R_\infty(\sigma)$ then $\ell(R_\infty(\sigma)_{\mathfrak{q}}) = 1$. It follows from Lemma 3.10 that if $V(\mathfrak{q})$ is an irreducible component of $\operatorname{Spec} R_\infty(\sigma)$ in the support of $M_\infty(\sigma)$ then $\ell(M_\infty(\sigma)_{\mathfrak{q}}) = r$.

It follows from (38) that

$$e\big(M_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) = r\sum_{i=1}^{m} e\big(R_\infty(\sigma)/(\varpi,\mathfrak{q}_i),\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big), \tag{43}$$

$$e\big(R_\infty(\sigma)/\varpi,\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) = \sum_{\mathfrak{q}} e\big(R_\infty(\sigma)/(\varpi,\mathfrak{q}),\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big), \tag{44}$$

where the last sum is taken over all the irreducible components $V(\mathfrak{q})$. Since $e\big(R_\infty(\sigma)/(\varpi,\mathfrak{q}), R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) \neq 0$ we deduce from (42)–(44) that (b) implies (c). We have

$$R_\infty(\sigma)/(y_1,\dots,y_{h+j}) \cong R^{\psi}_{F,S}(\sigma),$$
$$M_\infty(\sigma)/(y_1,\dots,y_{h+j})M_\infty(\sigma) \cong S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}.$$

Thus, if $M_\infty(\sigma)$ is supported on the whole of Spec $R_\infty(\sigma)$ then $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ is supported on the whole of Spec $R^{\psi}_{F,S}(\sigma)$. Since $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ 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 $f_1,\dots,f_s,\varpi$ is a part of a system of parameters of $R^{\psi,\square}_S(\sigma)[[x_1,\dots,x_{s+|S|-1}]]$, and Lemma 3.9 implies that every irreducible component of that ring contains a closed point of $R^{\psi}_{F,S}(\sigma)\big[\frac{1}{2}\big]$. Since every such component is of the form $\mathfrak{q}[[x_1,\dots,x_{s+|S|-1}]]$, we deduce that every irreducible component of $R^{\psi,\square}_S(\sigma)$ contains a closed point of $R^{\psi}_{F,S}(\sigma)\big[\frac{1}{2}\big]$. It follows from the second part of (d) that the support of $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ meets every irreducible component of $R^{\psi,\square}_S(\sigma)$. It follows from Lemma 3.14 that (d) implies (a). Since $S_{\sigma,\psi}(U,\mathcal{O})\big[\frac{1}{2}\big]$ is a finite-dimensional L-vector space, the last assertion is a direct consequence of (d). $\square$

3D. *Small weights.* Let $\tilde{\mathbf{1}}$ be the trivial representation of $\mathrm{GL}_2(\mathbb{Z}_2)$ on a free $\mathcal{O}$-module of rank 1. We let $\tilde{\mathrm{st}}$ be the space of functions $f : \mathbb{P}^1(\mathbb{F}_2) \to \mathcal{O}$ such that $\sum_{x\in\mathbb{P}^1(\mathbb{F}_2)} f(x) = 0$ equipped with the natural action of $\mathrm{GL}_2(\mathbb{Z}_2)$. The reduction of $\tilde{\mathbf{1}}$ modulo ϖ is the trivial representation, the reduction of $\tilde{\mathrm{st}}$ modulo ϖ is isomorphic to k^2, which we will also denote by st. These are the only smooth irreducible k-representations of $\mathrm{GL}_2(\mathbb{Z}_2)$.

The purpose of this subsection is to verify that the equivalent conditions of Proposition 3.17 hold when, for all $v \mid 2$, σ_v is either $\tilde{\mathbf{1}}$ or $\tilde{\mathrm{st}}$, under the assumption that $\bar{\rho}|_{G_v}$ does not have scalar semisimplification at any place $v \mid 2$. If σ 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 $\bar{\rho}|_{G_{F_v}}$ with Hodge–Tate weights $(0, 1)$ is crystalline (see Corollary 2.38). This implies that $S_{\tilde{\mathbf{1}},\psi}(U,\mathcal{O})_{\mathfrak{m}}$ and $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ and $R^{\psi}_{F,S}(\tilde{\mathbf{1}})$ and $R^{\psi}_{F,S}(\sigma)$ coincide.

If $p > 2$, the results of this section are proved in [Gee 2011] by a characteristic-p argument.

Proposition 3.18. *Assume that ψ is trivial on $U \cap (\mathbb{A}_F^f)^\times$, $\sigma_v = \tilde{\mathbf{1}}$ for all $v \mid 2$ and $\bar\rho|_{G_v}$ does not have scalar semisimplification for any $v \mid 2$. Then $R^\psi_{F,S}(\sigma)$ 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 $\bar\rho|_{G_v}$ is unramified, in which case F'_w is an unramified extension of $\mathbb{Q}_2$ and $\bar\rho|_{G_{F'_w}}$ 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 $R^\psi_{F',S'}(\sigma)$ 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 $G_{F',S'}$ induces a map between the deformation functors and hence a homomorphism $R^\psi_{F',S'}(\sigma) \to R^\psi_{F,S}(\sigma)$. Let $\rho^\psi_{F,S} : G_{F,S} \to \mathrm{GL}_2\big(R^\psi_{F,S}(\sigma)\big)$ be the universal deformation. Since $R^\psi_{F',S'}(\sigma)/\varpi$ is finite, the image of $G_{F',S'}$ in $\mathrm{GL}_2\big(R^\psi_{F,S}(\sigma)/\varpi\big)$ under $\rho^\psi_{F,S}$ is a finite group. Since F'/F is finite the image of $G_{F,S}$ in $\mathrm{GL}_2\big(R^\psi_{F,S}(\sigma)/\varpi\big)$ is a finite group. Lemma 3.6 in [Khare and Wintenberger 2009a] implies that $R^\psi_{F,S}(\sigma)/\varpi$ is finite. Since $\dim R^\psi_{F,S}(\sigma) \ge 1$ by Corollary 3.16, we conclude that $\dim R^\psi_{F,S}(\sigma) = 1$ and ϖ is a system of parameters for $R^\psi_{F,S}(\sigma)$, which implies that $R^\psi_{F,S}(\sigma)$ is a finite $\mathcal{O}$-module of rank at least 1. □

Corollary 3.19. *Assume that ψ is trivial on $U \cap (\mathbb{A}_F^f)^\times$, $\sigma_v = \tilde{\mathbf{1}}$ for all $v \mid 2$ and $\bar\rho|_{G_v}$ does not have scalar semisimplification for any $v \mid 2$. If $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m} \neq 0$ then the equivalent conditions of Proposition 3.17 hold.*

Proof. Since $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m}$ is nonzero and $\mathcal{O}$-torsion-free, there is a maximal ideal $\mathfrak{n}$ of $R^\psi_{F,S}\big[\frac{1}{2}\big]$ such that $S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{n} \neq 0$. This implies that $\bar\rho$ satisfies hypotheses (α) and (β) made in Section 8.2 of [Khare and Wintenberger 2009b].

Let $\mathfrak{n}$ be any maximal ideal of $R^{\psi,\square}_{F,S}(\sigma)\big[\frac{1}{2}\big]$, and let $\rho_\mathfrak{n}$ 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 $\rho_\mathfrak{n} \cong \rho_f$. Let $\pi = \bigotimes'_v \pi_v$ be the corresponding automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F^f)$. If v is a finite place, where D ramifies, then, because of the way we have set up our deformation problem, $\rho_\mathfrak{n}|_{G_{F_v}}$ is isomorphic to $\left(\begin{smallmatrix} \gamma_v\chi_{\mathrm{cyc}} & * \\ 0 & \gamma_v \end{smallmatrix}\right)$, where γ_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 $\rho_\mathfrak{n}$ arises from a Hilbert modular form, the representation $\rho_\mathfrak{n}|_{G_{F_v}}$ cannot be split, as in this case we would obtain a contradiction to the purity of $\rho_\mathfrak{n}$; see [Blasius 2006, §2.2]. Hence, $\rho_\mathfrak{n}|_{G_{F_v}}$ is nonsplit, and this implies that π_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 $f^D \in S_{\sigma,\psi}(U, \mathcal{O})_\mathfrak{m}$

with the same Hecke eigenvalues as f. This implies that $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ is supported on $\mathfrak{n}$. Proposition 3.18 implies that part (d) of Proposition 3.17 holds. $\square$

Lemma 3.20. *Fix a place w of F above 2. Let σ and σ' be such that for all $v \mid 2$, $v \neq w$, we have $\sigma_v = \sigma'_v$, which is equal to either $\tilde{\mathbf{1}}$ or $\tilde{\mathrm{st}}$, and $\sigma_w = \tilde{\mathbf{1}}$ and $\sigma'_w = \tilde{\mathrm{st}}$. Assume that ψ is trivial on $U \cap (\mathbb{A}_F^f)^\times$, and $\bar{\rho}|_{G_{F_w}}$ does not have scalar semisimplification. Then the rings $R^{\psi}_{F,S}(\sigma)$ and $R^{\psi}_{F,S}(\sigma')$ are equal. Moreover, if $\mathfrak{n}$ is a maximal ideal of $R^{\psi}_{F,S}(\sigma)\left[\frac{1}{2}\right]$ then $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ is supported on $\mathfrak{n}$ if and only if $S_{\sigma',\psi}(U,\mathcal{O})_{\mathfrak{m}}$ is supported on $\mathfrak{n}$.*

Proof. The ring $R^{\psi,\square}_w(\tilde{\mathbf{1}})$ parametrizes crystalline lifts of $\bar{\rho}|_{G_{F_w}}$ with Hodge–Tate weights $(0, 1)$. The ring $R^{\psi,\square}_w(\tilde{\mathrm{st}})$ parametrizes semistable lifts of $\bar{\rho}|_{G_{F_w}}$ with Hodge–Tate weights $(0, 1)$. Since both rings are reduced and $\mathcal{O}$-torsion-free, we have a surjection $R^{\psi,\square}_w(\tilde{\mathrm{st}}) \twoheadrightarrow R^{\psi,\square}_w(\tilde{\mathbf{1}})$. The assumption that $\bar{\rho}|_{G_{F_w}}$ 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 τ be either σ or σ'. We fix an isomorphism $i : \overline{\mathbb{Q}}_p \cong \mathbb{C}$, let $\tau_{\mathbb{C}} = \tau \otimes_{\mathcal{O}} \mathbb{C}$ and let $\tau^*_{\mathbb{C}}$ be the $\mathbb{C}$-linear dual of τ. Since $U \cap (\mathbb{A}_F^f)^\times$ acts trivially on τ by assumption, we may consider $\tau^*_{\mathbb{C}}$ as a representation of $U(\mathbb{A}_F^f)^\times$, on which $(\mathbb{A}_F^f)^\times$ acts by ψ. Let $U' = \prod_v U'_v$ be an open subgroup of U such that $U'_v = U_v$, if $v \nmid 2$ and $U'_v = \{g \in U_v : g \equiv 1 \pmod 2\}$ for all $v \mid 2$. Then U' acts trivially on τ. Let $C^\infty(D^\times \backslash (D \otimes_F \mathbb{A}_F)^\times / U')$ be the space of smooth $\mathbb{C}$-valued functions on $D^\times \backslash (D \otimes_F \mathbb{A}_F)^\times$ 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

$$S_{\tau,\psi}(U,\mathcal{O}) \otimes_{\mathcal{O}} \mathbb{C} \cong \mathrm{Hom}_{U(\mathbb{A}_F^f)^\times}(\tau, C^\infty(D^\times \backslash (D \otimes_F \mathbb{A}_F)^\times / U' D^\times_\infty)).$$

This isomorphism is equivariant for the Hecke operators at $v \notin S$. The action of $R^{\psi,\square}_{F,S}(\tau)$ on $S_{\tau,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ factors through the action of the Hecke algebra $\mathbb{T}_{\tau,\psi}(U)$. Let $\mathfrak{n}$ be a maximal ideal of $\mathbb{T}(U)_{\tau,\psi}\left[\frac{1}{2}\right]$. The isomorphism above implies that $S_{\tau,\psi}(U,\mathcal{O})_{\mathfrak{n}}$ is nonzero if and only if there is an automorphic form

$$f^D \in C^\infty(D^\times \backslash (D \otimes_F \mathbb{A}_F)^\times / U' D^\times_\infty),$$

on which the Hecke operators for $v \notin S$ act by the eigenvalues given by the map $\mathbb{T}_{\tau,\psi}(U) \to \kappa(n) \xrightarrow{i} \mathbb{C}$. Additionally, $\mathrm{Hom}_{U(\mathbb{A}_F^f)^\times}(\tau^*_{\mathbb{C}}, \pi) \neq 0$, where $\pi = \otimes'_v \pi_v$ is the automorphic representation corresponding to f^D.

If $S_{\sigma,\psi}(U,\mathcal{O})_{\mathfrak{n}}$ is nonzero then the above implies that $\mathrm{Hom}_{U_w}(\mathbf{1}, \pi_w) \neq 0$, which implies that π_w is an unramified principal series representation, which implies that

$\mathrm{Hom}_{U_w}(\tilde{\mathrm{st}}, \pi_w) \neq 0$. Since $\sigma_v = \sigma'_v$ for all $v \neq w$, we conclude that $S_{\sigma',\psi}(U, \mathcal{O})_{\mathfrak{n}}$ is nonzero.

If $S_{\sigma',\psi}(U, \mathcal{O})_{\mathfrak{n}}$ is nonzero then the same argument shows that $\mathrm{Hom}_{U_w}(\tilde{\mathrm{st}}, \pi_w) \neq 0$, which implies that π_w is either an unramified principal series representation, in which case $\mathrm{Hom}_{U_w}(\mathbf{1}, \pi_w) \neq 0$ and thus $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{n}} \neq 0$, or π_w is a special series. We would like to rule the last case out. By Jacquet–Langlands correspondence to π we may associate an automorphic representation $\pi' = \otimes'_v \pi'_v$ of $\mathrm{GL}_2(\mathbb{A}_F)$ such that $\pi_v = \pi'_v$ for all v, where D is split. In particular, $\pi'_w = \pi_w$. Let $\rho_{\mathfrak{n}}$ be the representation of $G_{F,S}$ corresponding to the maximal ideal $\mathfrak{n}$ of $R^{\psi}_{F,S}\left[\frac{1}{2}\right]$. By the compatibility of local and global Langlands correspondence, if π'_w is special then $\rho|_{G_{F_w}}$ is semistable noncrystalline. However, this cannot happen, as explained above. $\square$

Corollary 3.21. *Assume that ψ is trivial on $U \cap (\mathbb{A}_F^f)^\times$, σ_v is either $\tilde{\mathbf{1}}$ or $\tilde{\mathrm{st}}$ for all $v \mid 2$, and $\bar{\rho}|_{G_v}$ does not have scalar semisimplification for any $v \mid 2$. If $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ then the equivalent conditions of Proposition 3.17 hold.*

Proof. If $\sigma_v = \tilde{\mathbf{1}}$ for all $v \mid 2$ 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 σ as above. $\square$

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_σ, where the σ_v are of the form (30) or (31). Let $\psi : (\mathbb{A}_F^f)^\times / F^\times \to \mathcal{O}^\times$ be a continuous character such that $U \cap (\mathbb{A}_F^f)^\times$ acts on W_σ by the character ψ. Let $\bar{\sigma}$ and $\bar{\psi}$ be representations obtained by reducing σ and ψ modulo ϖ. We 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 $R^{\psi,\square}_{F,S}$ is formally smooth over $R^{\psi}_{F,S}$, it is a flat $R^{\psi}_{F,S}$-module; therefore, the functor $\otimes_{R^{\psi}_{F,S}} R^{\psi,\square}_{F,S}$ is exact, and so is the localization at $\mathfrak{m}_{Q_n}$. Hence the functor

$$\sigma \mapsto M_n(\sigma) = R^{\psi,\square}_{F,S_{Q_n}} \otimes_{R^{\psi}_{F,S_{Q_n}}} S_{\sigma,\psi}(U_{Q_n}, \mathcal{O})_{\mathfrak{m}_{Q_n}} \tag{45}$$

is exact. Following [Kisin 2009a, §2.2.5] we fix a U-invariant filtration on $\bar{\sigma}$ 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

$$0 = M_n^0(\sigma) \subset M_n^1(\sigma) \subset \cdots \subset M_n^s(\sigma) = M_n(\sigma) \otimes_{\mathcal{O}} k, \tag{46}$$

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 σ_i is of the form $\bigotimes_{v|2} \sigma_{i,v}$, where $\sigma_{i,v}$ is either the trivial representation, in which case we let $\tilde{\sigma}_{i,v} = \tilde{\mathbf{1}}$, or st, in which case we let $\tilde{\sigma}_{i,v} := \tilde{\mathrm{st}}$. We let $\tilde{\sigma}_i := \bigotimes_{v|2} \tilde{\sigma}_{i,v}$ and consider it as a representation of U by letting U_v for v not above 2 act trivially. We note that, since both $\tilde{\mathbf{1}}$ and $\tilde{\mathrm{st}}$ 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 $\psi \equiv \xi \pmod{\varpi}$ and the restriction of ξ to $U \cap (\mathbb{A}_F^f)^\times$ is trivial. For example, we could choose ξ to be a Teichmüller lift of $\bar{\psi}$. Let

$$M_n(\tilde{\sigma}_i) = R^{\xi,\square}_{F,S_{Q_n}} \otimes_{R^{\xi}_{F,S_{Q_n}}} S_{\tilde{\sigma}_i,\xi}(U_{Q_n}, \mathcal{O})_{\mathfrak{m}_{Q_n}}.$$

The exactness of the functor in (45), used with $\tilde{\sigma}_i$ and ξ instead of σ and ψ, 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 $\mathfrak{a}_{i,n}$ be the $R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)$-annihilator of $M_n(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k$. Since the action of $R_S^\square[[x_1, \ldots, x_g]]$ on $M_n(\sigma)$ and $M_n(\tilde{\sigma}_i)$ factors through $R^{\psi,\square}_{F,S_{Q_n}}(\sigma)$ and $R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)$, respectively, we obtain a surjection

$$\varphi_{i,n} : R^{\psi,\square}_{F,S_{Q_n}}(\sigma) \twoheadrightarrow R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)/\mathfrak{a}_{i,n}. \tag{49}$$

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 $1 \le i \le s$ there is an $R_\infty(\tilde{\sigma}_i)$-module $M_\infty(\tilde{\sigma}_i)$ as in Section 3C and a surjection $\varphi_i : R_\infty(\sigma) \twoheadrightarrow R_\infty(\tilde{\sigma}_i)/\mathfrak{a}_i$, where $\mathfrak{a}_i$ 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 $1 \le i \le s$ 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 σ indicates the fixed weight and inertial type we are working with. In particular, $D(\sigma)_m$ and $D'(\sigma)_m$ are finite $R_S^{\psi,\square}(\sigma)[[x_1,\dots,x_g]]$-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

$$\Delta_m := \bigl(\Delta(\sigma)_m,\ \{L(\sigma)_m^i\}_{i=0}^s,\ \{\Delta(\tilde{\sigma}_i)_m\}_{i=1}^s,\ \{\varphi_{i,m}\}_{i=1}^s,\ \{\alpha_{i,m}\}_{i=1}^s\bigr),$$

where $\{L(\sigma)_m^i\}_{i=0}^s$ is a filtration of $L(\sigma)_m\otimes_{\mathcal{O}} k$ by $D(\sigma)_m$-submodules, $\varphi_{i,m}: D(\sigma)_m \twoheadrightarrow D(\tilde{\sigma}_i)_m/\mathfrak{a}_{i,m}$ is a surjection of $R_S^{\square}[[x_1,\dots,x_g]]$-algebras, where $\mathfrak{a}_{i,m}$ is the $D(\tilde{\sigma}_i)_m$-annihilator of $L(\tilde{\sigma}_i)\otimes_{\mathcal{O}} k$, and $\alpha_{i,m}$ is an isomorphism of $D(\sigma)_m$-modules between $L(\sigma)_m^i/L(\sigma)_m^{i-1}$ and $L(\tilde{\sigma}_i)\otimes_{\mathcal{O}} k$, where the action of $D(\sigma)_m$ on this last module is given by $\varphi_{i,m}$.

An isomorphism of patching data between Δ_m and Δ'_m is a tuple $\bigl(\beta,\{\beta_i\}_{i=1}^s\bigr)$, where $\beta:\Delta_m(\sigma)\cong\Delta'_m(\sigma)$ and $\beta_i:\Delta_m(\tilde{\sigma}_i)\cong\Delta_m(\tilde{\sigma}_i)$ are isomorphisms of patching data, in the sense of [Khare and Wintenberger 2009b, Proposition 9.3], which respect the filtration and the maps $\{\varphi_{i,m}\}_{i=1}^s$, $\{\alpha_{i,m}\}_{i=1}^s$. 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 $\mathfrak{c}_m$ and $\mathfrak{b}_n$ are those defined in [loc. cit.]. For an integer $n\ge n_0+1$ and for m with $n\ge m\ge 3$, 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

$$D(\sigma)_{n,m} = R_{n+a}(\sigma)/\bigl(\mathfrak{c}_m R_{n+a}(\sigma)+\mathfrak{m}_{R_{n+a}(\sigma)}^{(r_m)}\bigr),$$
$$L(\sigma)_{n,m} = M_{n+a}(\sigma)/\mathfrak{c}_m M_{n+a}(\sigma),$$

where $R_n(\sigma):=R_{F,S_{Q_n}}^{\psi,\square}(\sigma)$. We define $\Delta_{n,m}(\tilde{\sigma}_i)$ analogously with $\tilde{\sigma}_i$ instead of σ and with ξ instead of ψ. We let $(L(\sigma)_{n,m}^i)_{i=1}^s$ be the filtration obtained by reducing (46) modulo $\mathfrak{c}_m$. Similarly, we let $\{\varphi_{i,n,m}\}_{i=1}^s$, $\{\alpha_{i,n,m}\}_{i=1}^s$ be the maps obtained by reducing (48) and (49) modulo $\mathfrak{c}_m$. Then

$$\Delta_{n,m} := \bigl(\Delta(\sigma)_{n,m},\ \{L(\sigma)_{n,m}^i\}_{i=0}^s,\ \{\Delta(\tilde{\sigma}_i)_{n,m}\}_{i=1}^s,\ \{\varphi_{i,n,m}\}_{i=1}^s,\ \{\alpha_{i,n,m}\}_{i=1}^s\}\bigr)$$

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

$$\bigl((R_{n+a}(\sigma), M_{n+a}(\sigma)),\ \{(R_{n+a}(\tilde{\sigma}_i), M_{n+a}(\tilde{\sigma}_i))\}_{i=1}^s\bigr)_{n\ge n_0+1}$$

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 $R_\infty^{\mathrm{inv}}(\sigma)$ under φ_i. Following [Khare and Wintenberger 2009b] we let $\mathrm{CNL}_{\mathcal{O}}$ 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 $A \in \mathrm{CNL}_{\mathcal{O}}$ then we let $\mathrm{Sp}_A : \mathrm{CNL}_{\mathcal{O}} \to \mathrm{Sets}$ be the functor $\mathrm{Sp}_A(B) = \mathrm{Hom}_{\mathrm{CNL}_{\mathcal{O}}}(A, B)$. 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, $G^*(A) = \mathrm{Hom}_{\mathrm{Groups}}(G, A^\times)$. Assume that we are given a free G^* action on Sp_A. This means that, for all $B \in \mathrm{CNL}_{\mathcal{O}}$, $G^*(B)$ acts on $\mathrm{Sp}_A(B)$ without fixed points. By Proposition 2.6(1) in [Khare and Wintenberger 2009b] the quotient $G^* \backslash \mathrm{Sp}_A$ exists in $\mathrm{CNL}_{\mathcal{O}}$ and is represented by $(A^{\mathrm{inv}}, \mathfrak{m}_A^{\mathrm{inv}}) \in \mathrm{CNL}_{\mathcal{O}}$. Moreover, Sp_A is a G^*-torsor over $\mathrm{Sp}_{A^{\mathrm{inv}}}$.

Lemma 3.23. *Let $(A, \mathfrak{m}_A)$ and $(B, \mathfrak{m}_B)$ be in $\mathrm{CNL}_{\mathcal{O}}$. Assume that G^* acts freely on Sp_A and Sp_B and we are given a G^*-equivariant closed immersion $\mathrm{Sp}_B \hookrightarrow \mathrm{Sp}_A$. Then the map induces a closed immersion $\mathrm{Sp}_{B^{\mathrm{inv}}} \hookrightarrow \mathrm{Sp}_{A^{\mathrm{inv}}}$.*

Proof. Since G^* acts trivially on $\mathrm{Sp}_{A^{\mathrm{inv}}}$, by the universal property of the quotient, the map $\mathrm{Sp}_B \to \mathrm{Sp}_A \to \mathrm{Sp}_{A^{\mathrm{inv}}}$ factors through $\mathrm{Sp}_{B^{\mathrm{inv}}} \to \mathrm{Sp}_{A^{\mathrm{inv}}}$. Hence, we obtain the following commutative diagram in $\mathrm{CNL}_{\mathcal{O}}$:

$$\begin{tikzcd} A^{\mathrm{inv}} \arrow[r] \arrow[d] & B^{\mathrm{inv}} \arrow[d] \\ A \arrow[r, two heads] & B \end{tikzcd}$$

Since Sp_A is a G^*-torsor over $\mathrm{Sp}_{A^{\mathrm{inv}}}$, 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 $A \twoheadrightarrow B$ induces a surjection $A/\mathfrak{m}_A^{\mathrm{inv}} A \twoheadrightarrow B/\mathfrak{m}_B^{\mathrm{inv}} B$. Since both k-vector spaces have dimension $|G|$, the map is an isomorphism and this implies that the image of $\mathfrak{m}_A^{\mathrm{inv}}$ is equal to $\mathfrak{m}_B^{\mathrm{inv}}$. Hence, the top horizontal arrow in the diagram is surjective. $\square$

Let $\mathrm{CNL}_{\mathcal{O}}^{[m]}$ be the full subcategory of $\mathrm{CNL}_{\mathcal{O}}$ consisting of objects $(A, \mathfrak{m}_A)$ such that $\mathfrak{m}_A^m = 0$. We have a truncation functor $\mathrm{CNL}_{\mathcal{O}} \to \mathrm{CNL}_{\mathcal{O}}^{[m]}$, $A \mapsto A^{[m]} := A/\mathfrak{m}_A^m$. 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 $(A, \mathfrak{m}_A)$ and $(B, \mathfrak{m}_B)$ be in $\mathrm{CNL}_{\mathcal{O}}$. Assume that G^* acts freely on $X := \mathrm{Sp}_A$ and $Y := \mathrm{Sp}_B$ 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 $\mathfrak{m}_A^{\mathrm{inv}} A$ in $A/\mathfrak{m}_A^m = B/\mathfrak{m}_B^m$ is equal to the image of $\mathfrak{m}_B^{\mathrm{inv}} B$.*

Proof. Let X^{inv} and Y^{inv} denote the quotients of X and Y by G^*. Then we have isomorphisms

$$G^* \times X \cong X \times_{X^{\mathrm{inv}}} X, \quad G^* \times Y \cong Y \times_{Y^{\mathrm{inv}}} Y,$$

where the map is given by $(g, x) \mapsto (x, gx)$. We define $Z := X^{[m]} = Y^{[m]}$ and $C := A/\mathfrak{m}_A^m = B/\mathfrak{m}_B^m$. The restriction of the above isomorphism to $\mathrm{CNL}_{\mathcal{O}}^{[m]}$ gives us isomorphisms

$$(G^* \times Z)^{[m]} \cong (Z \times_{X^{\mathrm{inv}}} Z)^{[m]}, \quad (G^* \times Z)^{[m]} \cong (Z \times_{Y^{\mathrm{inv}}} Z)^{[m]}.$$

Thus we have an isomorphism

$$(Z \times_{X^{\mathrm{inv}}} Z)^{[m]} \cong (Z \times_{Y^{\mathrm{inv}}} Z)^{[m]},$$

where the map is given by $(z_1, z_2) \mapsto (z_1, z_2)$. On rings this isomorphism reads $(C \otimes_{A^{\mathrm{inv}}} C)^{[m]} \cong (C \otimes_{B^{\mathrm{inv}}} C)^{[m]}$, $c_1 \otimes c_2 \mapsto c_1 \otimes c_2$.

Both $A/\mathfrak{m}_A^{\mathrm{inv}} A$ and $B/\mathfrak{m}_B^{\mathrm{inv}} B$ are k-vector spaces of dimension $|G|$. In particular, if $m > |G|$ then $\mathfrak{m}_A^m \subset \mathfrak{m}_A^{\mathrm{inv}} A$ and $\mathfrak{m}_B^m \subset \mathfrak{m}_B^{\mathrm{inv}} B$. So we obtain a map $C \twoheadrightarrow A/\mathfrak{m}_A^{\mathrm{inv}} A$. If $m > 2|G|$ then by base changing along this map, we obtain an isomorphism

$$A/\mathfrak{m}_A^{\mathrm{inv}} A \otimes_k A/\mathfrak{m}_A^{\mathrm{inv}} A \cong A/\mathfrak{m}_A^{\mathrm{inv}} A \otimes_{B^{\mathrm{inv}}} A/\mathfrak{m}_A^{\mathrm{inv}} A.$$

If the image of B^{inv} in $A/\mathfrak{m}_A^{\mathrm{inv}} A$ is not equal to k then, for some $b \in B^{\mathrm{inv}}$, $1 \otimes b$ 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^{inv} in $A/\mathfrak{m}_A^{\mathrm{inv}} A$ is equal to k. Thus $\mathfrak{m}_B^{\mathrm{inv}} C \subset \mathfrak{m}_A^{\mathrm{inv}} C$ and by symmetry we obtain the other inclusion. $\square$

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 $G_{n,2} \cong (\mathbb{Z}/2\mathbb{Z})^t$. Let $G_{n,2}^*$ be the group scheme defined over $\mathcal{O}$ such that, for every $\mathcal{O}$-algebra A, $G_{n,2}^*(A) = \mathrm{Hom}_{\mathrm{Groups}}(G_{n,2}, A^\times)$. For a local artinian augmented $\mathcal{O}$-algebra A and $\chi \in G_{n,2}^*(A)$, if ρ_A is a $G_{F,S_{Q_n}}$-representation lifting $\bar\rho$ to A then so is $\rho_A \otimes \chi$. Moreover, since χ^2 is trivial, ρ_A and $\rho_A \otimes \chi$ have the same determinant. This induces an action of $G_{n,2}^*$ on

$$\mathrm{Spf}\, R_{F,S_{Q_n}}^{\square}, \quad \mathrm{Spf}\, R_{F,S_{Q_n}}^{\psi,\square}(\sigma), \quad \text{and} \quad \mathrm{Spf}\, R_{F,S_{Q_n}}^{\xi,\square}(\tilde\sigma_i).$$

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 $\big(R_{F,S_{Q_n}}^{\square,\mathrm{inv}}, \mathfrak{m}_n^{\mathrm{inv}}\big)$, $\big(R_{F,S_{Q_n}}^{\psi,\square,\mathrm{inv}}(\sigma), \mathfrak{m}_{n,\sigma}^{\mathrm{inv}}\big)$ and $\big(R_{F,S_{Q_n}}^{\xi,\square,\mathrm{inv}}(\tilde\sigma_i), \mathfrak{m}_{n,\tilde\sigma_i}^{\mathrm{inv}}\big)$, respectively.

Lemma 3.25. *The map*

$$\operatorname{Spf} R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)/\mathfrak{a}_{i,n} \to \operatorname{Spf} R^{\psi,\square}_{F,S_{Q_n}}(\sigma)$$

induced by (49) is $G^*_{n,2}$-equivariant. Moreover,

$$\varphi_{i,n}\big(\mathfrak{m}^{\mathrm{inv}}_{n,\sigma} R^{\psi,\square}_{F,S_{Q_n}}(\sigma)\big) = \mathfrak{m}^{\mathrm{inv}}_{n,\tilde{\sigma}_i} R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)/\mathfrak{a}_{i,n}.$$

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

$$q_\sigma : R^{\square}_{F,S_{Q_n}} \twoheadrightarrow R^{\psi,\square}_{F,S_{Q_n}}(\sigma) \quad \text{and} \quad q_{\tilde{\sigma}_i} : R^{\square}_{F,S_{Q_n}} \twoheadrightarrow R^{\xi,\square}_{F,S_{Q_n}}(\tilde{\sigma}_i)$$

denote the natural surjections. Since $\varphi_{i,n} \circ q_\sigma = q_{\tilde{\sigma}_i} \pmod{\mathfrak{a}_{n,i}}$, it is enough to show that $q_\sigma\big(\mathfrak{m}^{\mathrm{inv}}_n R^{\square}_{F,S_{Q_n}}\big) = \mathfrak{m}^{\mathrm{inv}}_{n,\sigma} R^{\psi,\square}_{F,S_{Q_n}}(\sigma)$ for all σ and ψ as above. This follows from Lemma 3.23. ☐

Let $\mathfrak{m}^{\mathrm{inv}}_\sigma$ and $\mathfrak{m}^{\mathrm{inv}}_{\tilde{\sigma}_i}$ be the maximal ideals of $R^{\mathrm{inv}}_\infty(\sigma)$ and $R^{\mathrm{inv}}_\infty(\tilde{\sigma}_i)$, respectively.

Proposition 3.26. *The surjection $\varphi_i : R_\infty(\sigma) \twoheadrightarrow R_\infty(\tilde{\sigma}_i)/\mathfrak{a}_i$ maps $\mathfrak{m}^{\mathrm{inv}}_\sigma R_\infty(\sigma)$ onto the image of $\mathfrak{m}^{\mathrm{inv}}_{\tilde{\sigma}_i} R_\infty(\tilde{\sigma}_i)$. In particular,*

$$e\big(M^i_\infty(\sigma)/M^{i-1}_\infty(\sigma),\, R^{\mathrm{inv}}_\infty(\sigma)/\varpi\big) = e\big(M_\infty(\tilde{\sigma}_i) \otimes_{\mathcal{O}} k,\, R^{\mathrm{inv}}_\infty(\tilde{\sigma}_i)/\varpi\big). \tag{50}$$

Proof. If $(A, \mathfrak{m})$ is a complete local noetherian algebra then by $A^{[r]}$ we denote the ring $A/\mathfrak{m}^r$. 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

$$R_\infty(\sigma) \cong \varprojlim_m D''_{m,m}(\sigma),$$

where $D''_{m,n}(\sigma) = R_{n+a}(\sigma)^{[r'_m]}$. Moreover, it is shown that the map is $(\widehat{\mathbb{G}}_m[2])^t$-equivariant by fixing an identification of $G_{n+a,2}$ with $(\mathbb{Z}/2\mathbb{Z})^t$.

For each fixed $r \geq 0$ we have

$$R_\infty(\sigma)^{[r]} \cong \varprojlim_m D''_{m,m}(\sigma)^{[r]}.$$

Hence, by choosing m large enough we may assume that $R_\infty(\sigma)^{[r]} = D''_{m,m}(\sigma)^{[r]}$ with $r \leq r'_m$. Since $(\widehat{\mathbb{G}}_m[2])^t$-action on $\mathrm{Sp}_{R_\infty(\sigma)}$ and on $\mathrm{Sp}_{R_{n+a}(\sigma)}$ 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 $\mathfrak{m}^{\mathrm{inv}}_\sigma R_\infty(\sigma)$ in $D''_{m,m}(\sigma)^{[r]}$ is equal to the image of $\mathfrak{m}^{\mathrm{inv}}_{m+a,\sigma} R_{m+a}(\sigma)$. It follows from Lemma 3.25 that the composition

$$R_\infty(\sigma) \longrightarrow R_{m+a}(\sigma)^{[r]} \xrightarrow{\varphi_{i,m}} (R_{m+a}(\tilde{\sigma}_i)/\mathfrak{a}_{i,m})^{[r]}$$

maps $\mathfrak{m}_\sigma^{\mathrm{inv}} R_\infty(\sigma)$ onto the image of $\mathfrak{m}_{\tilde{\sigma}_i}^{\mathrm{inv}} R_\infty(\tilde{\sigma}_i)$. The action of $R_{m+a}(\tilde{\sigma}_i)$ on $L_{m,m}(\tilde{\sigma}_i)$ factors through $R_{m+a}(\tilde{\sigma}_i)^{[r'_m]}$. Since by construction

$$\varphi_i = \varprojlim_m \varphi_{i,m}, \quad R_\infty(\tilde{\sigma}_i) = \varprojlim_m R_{m+a}(\tilde{\sigma}_i)^{[r'_m]}, \quad M_\infty(\tilde{\sigma}_i) = \varprojlim_m L_{m,m}(\tilde{\sigma}_i),$$

we deduce that φ_i maps $\mathfrak{m}_\sigma^{\mathrm{inv}} R_\infty(\sigma)$ onto the image of $\mathfrak{m}_{\tilde{\sigma}_i}^{\mathrm{inv}} R_\infty(\tilde{\sigma}_i)$. $\square$

Corollary 3.27. *Assume that $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ and that $\bar{\rho}|_{G_{F_v}} \not\cong \left(\begin{smallmatrix} \chi & * \\ 0 & \chi \end{smallmatrix}\right)$ 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 $\rho : G_{F,S} \to \mathrm{GL}_2(\mathcal{O})$ corresponding to a maximal ideal of $R^{\psi}_{F,S}(\sigma)\left[\frac{1}{2}\right]$ is modular.*

Proof. We will verify that part (b) of Proposition 3.17 holds. We first note that, since $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ and U satisfies (29), there is an i such that $S_{\tilde{\sigma}_i,\xi}(U, k)_{\mathfrak{m}} \neq 0$. This implies that $S_{\tilde{\sigma}_i,\xi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$, and it follows from Lemma 3.20 that $S_{\tilde{\sigma}_i,\xi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ for all $1 \le i \le s$ and $S_{\tilde{\mathbf{1}},\xi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$. In particular, the rings $R_S^{\xi,\square}(\tilde{\sigma}_i)$ are nonzero and equal to $R_S^{\xi,\square}(\tilde{\mathbf{1}})$. Corollary 3.21 implies that for all $1 \le i \le s$ the equality

$$2^t r e\big(R_S^{\xi,\square}(\tilde{\sigma}_i)/\varpi\big) = e\big(M_\infty(\tilde{\sigma}_i)/\varpi,\, R_\infty^{\mathrm{inv}}(\tilde{\sigma}_i)/\varpi\big) \tag{51}$$

holds. Since the Hilbert–Samuel multiplicity is additive in short exact sequences, we have

$$e\big(M_\infty(\sigma)/\varpi,\, R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big) = \sum_{i=1}^{s} e\big(M_\infty^i(\sigma)/M_\infty^{i-1}(\sigma),\, R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big). \tag{52}$$

Proposition 3.26 implies that for all $1 \le i \le s$ we have

$$e\big(M_\infty^i(\sigma)/M_\infty^{i-1}(\sigma),\, R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big) = e\big(M_\infty(\tilde{\sigma}_i)/\varpi,\, R_\infty^{\mathrm{inv}}(\tilde{\sigma}_i)/\varpi\big). \tag{53}$$

Thus

$$e\big(M_\infty(\sigma)/\varpi,\, R_\infty^{\mathrm{inv}}(\sigma)/\varpi\big) = 2^t r \sum_{i=1}^{s} e\big(R_S^{\xi,\square}(\tilde{\sigma}_i)/\varpi\big). \tag{54}$$

Thus to verify part (b) of Proposition 3.17 it is enough to show that

$$e\big(R_S^{\psi,\square}(\sigma)/\varpi\big) \le \sum_{i=1}^{s} e\big(R_S^{\xi,\square}(\tilde{\sigma}_i)/\varpi\big). \tag{55}$$

If A and B are complete local κ-algebras with residue field κ then it is shown in [Kisin 2009a, Proposition 1.3.8] that $e(A \widehat{\otimes}_\kappa B) = e(A)e(B)$. Since ψ is congruent to ξ modulo ϖ, inequality (55) reduces to the following inequality on Hilbert–Samuel multiplicities of potentially semistable rings at all $v \mid 2$:

$$e\big(R_v^{\psi,\square}(\sigma_v)/\varpi\big) \le \sum_{i=1}^{s_v} e\big(R_v^{\xi,\square}(\tilde{\sigma}_{v,i})/\varpi\big). \tag{56}$$

Here the $\sigma_{v,i}$ are irreducible k-representation of $\mathrm{GL}_2(\mathbb{F}_2)$ which appear as graded pieces of a $\mathrm{GL}_2(\mathbb{Z}_2)$-invariant filtration on $\sigma_v \otimes_{\mathcal{O}} k$. Inequality (56) is proved in the local part of the paper; see Remark 2.39. □

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 $\bar{\rho}: G_F \to \mathrm{GL}_2(k)$ be a continuous absolutely irreducible representation. If there is a Hilbert eigenform f such that $\bar{\rho} \cong \bar{\rho}_f$ then there is a Hilbert eigenform g of parallel weight 2 such that $\bar{\rho} \cong \bar{\rho}_g$ and at $v \mid 2$ the corresponding representation π_v of $\mathrm{GL}_2(F_v)$ is either an unramified principal series or a twist of Steinberg representation by an unramified character. Moreover, if $\bar{\rho}|_{G_{F_v}} \not\cong \left(\begin{smallmatrix} \chi & * \\ 0 & \chi \end{smallmatrix}\right)$ for all $v \mid 2$ and any character $\chi: G_{F_v} \to k^\times$ then we may assume that π_v is an unramified principal series representation for all $v \mid 2$.*

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 $(D \otimes_F \mathbb{A}_F^f)^\times$ such that $U_v = \mathrm{GL}_2(\mathcal{O}_{F_v})$ for all $v \mid 2$, and U is sufficiently small, so that (29) holds, and $\tau = \bigotimes_{v\mid 2} \tau_v$ is a locally algebraic representation of U. Let $\mathfrak{m}$ be the maximal ideal of the Hecke algebra $\mathbb{T}_{S,\mathcal{O}}^{\mathrm{univ}}$ corresponding to $\bar{\rho}$. Then $f^D \in S_{\tau,\psi}(U,\mathcal{O})_{\mathfrak{m}}$, and hence $S_{\tau,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ is nonzero.

Let $\bar{\tau}$ denote the reduction of a U-invariant lattice in τ, and let $\bar{\psi}$ denote ψ modulo ϖ. 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 σ of $\bar{\tau}$ such that $S_{\sigma,\bar{\psi}}(U,k)_{\mathfrak{m}}$ is nonzero. Such a σ is of the form $\bigotimes_{v\mid 2} \sigma_v$, where σ_v is a representation of $\mathrm{GL}_2(\mathbb{F}_2)$. Thus σ_v is either trivial, in which case we let $\tilde{\sigma}_v = \tilde{\mathbf{1}}$, or k^2, in which case we let $\tilde{\sigma}_v = \tilde{\mathrm{st}}$. Then the reduction of $\tilde{\sigma}_v$ modulo ϖ_v is isomorphic to σ_v and $F_v^\times \cap U_v$ acts trivially on $\tilde{\sigma}_v$. Let $\tilde{\sigma} := \bigotimes_{v\mid 2} \tilde{\sigma}_v$. 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 $S_{\tilde{\sigma},\xi}(U,\mathcal{O})_{\mathfrak{m}}$ is nonzero, since its reduction modulo ϖ is equal to $S_{\sigma,\xi}(U,k)_{\mathfrak{m}}$. We may take any eigenform $g^D \in S_{\tilde{\sigma},\tilde{\psi}}(U,\mathcal{O})_{\mathfrak{m}}$ 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. □

Theorem 3.30. *Let F be a totally real field where 2 is totally split, and let*

$$\rho : G_{F,S} \to \mathrm{GL}_2(\mathcal{O})$$

be a continuous representation. Suppose:

(i) $\bar{\rho} : G_{F,S} \overset{\rho}{\to} \mathrm{GL}_2(\mathcal{O}) \to \mathrm{GL}_2(k)$ *is modular with nonsolvable image.*

(ii) *If* $v \mid 2$ *then* $\rho|_{G_{F_v}}$ *is potentially semistable with distinct Hodge–Tate weights.*

(iii) $\det \rho$ *is totally odd.*

(iv) *If* $v \mid 2$ *then* $\bar{\rho}|_{G_{F_v}} \ncong \left(\begin{smallmatrix} \chi & * \\ 0 & \chi \end{smallmatrix}\right)$, *for any character* $\chi : G_{F_v} \to k^\times$.

Then ρ *is modular.*

Proof. Let $\psi = \chi_{\mathrm{cyc}}^{-1} \det \rho$, where χ_{cyc} is the 2-adic cyclotomic character. By solvable base change it is enough to prove the assertion for the restriction of ρ 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 ρ is ramified at $v \nmid 2$ then the image of inertia is unipotent. Let Σ be the set of places outside 2 where ρ is ramified. If $v \in \Sigma$ then

$$\rho|_{G_{F'_v}} \cong \begin{pmatrix} \gamma_v \chi_{\mathrm{cyc}} & * \\ 0 & \gamma_v \end{pmatrix},$$

where γ_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 Σ, 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 $\bar{\rho}(\mathrm{Frob}_{v_1})$ has distinct eigenvalues. Let S be the union of infinite places, Σ, places above 2 and v_1. Let $U = \prod_v U_v$ be an open subgroup of $(D \otimes_{F''} \mathbb{A}^f_{F''})^\times$ such that $U_v = \mathcal{O}^\times_{D_v}$ if $v \neq v_1$ and U_{v_1} is unipotent upper triangular modulo ϖ_{v_1}. We note that Lemma 3.2 implies that U satisfies (29). Let $\mathfrak{m}$ be the maximal ideal in the Hecke algebra $\mathbb{T}^{\mathrm{univ}}_{S,\mathcal{O}}$ corresponding to $\bar{\rho}$.

Let g^D be the eigenform on D corresponding to g via the Jacquet–Langlands correspondence. Then $g^D \in S_{\sigma,\psi'}(U, \mathcal{O})_{\mathfrak{m}}$, where σ is the trivial representation of U and $\psi' : (\mathbb{A}^f_{F''})^\times \to \mathcal{O}^\times$ is a suitable character congruent to ψ modulo ϖ. In particular, $S_{\sigma,\psi'}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$. It follows from Lemma 3.20 that $S_{\sigma,\psi'}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ for all $\sigma = \bigotimes_{v|2} \sigma_v$, where σ_v is either $\tilde{\mathbf{1}}$ or $\tilde{\mathrm{st}}$. Since U satisfies (29), we deduce that $S_{\sigma,\psi}(U, k)_{\mathfrak{m}} \neq 0$ for any irreducible smooth k-representation σ of $\prod_{v|2} \mathrm{GL}_2(\mathbb{Z}_2)$. Since U satisfies (29), we deduce via Lemma 3.1.4 of [Kisin 2009c] that $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ for any continuous finite-dimensional representation σ of $\prod_{v|2} \mathrm{GL}_2(\mathbb{Z}_2)$ on which $U \cap (\mathbb{A}^f_{F''})^\times$ acts by ψ.

For $v \mid 2$ suppose that $\rho|_{G_{F''_v}}$ has Hodge–Tate weights $\boldsymbol{w}_v = (a_v, b_v)$ with $b_v > a_v$ and inertial type τ_v. Let σ_v be defined by (30) and let $\sigma = \bigotimes_{v|2} \sigma_v$. The above implies that $S_{\sigma,\psi}(U, \mathcal{O})_{\mathfrak{m}} \neq 0$ and, since $\rho|_{G_{F''}}$ defines a maximal ideal of $R^{\psi}_{F'',S}\bigl[\frac{1}{2}\bigr]$, the assertion follows from Corollary 3.27. $\square$

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

[Barthel and Livné 1994] L. Barthel and R. Livné, "Irreducible modular representations of GL_2 of a local field", *Duke Math. J.* **75**:2 (1994), 261–292. MR 1290194 Zbl 0826.22019

[Berger and Breuil 2010] L. Berger and C. Breuil, "Sur quelques représentations potentiellement cristallines de $\mathrm{GL}_2(\mathbb{Q}_p)$", pp. 155–211 in *p-adic representations of p-adic groups, II: Representations of* $\mathrm{GL}_2(\mathbb{Q}_p)$ *and* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2642406 Zbl 1243.11063

[Blasius 2006] D. Blasius, "Hilbert modular forms and the Ramanujan conjecture", pp. 35–56 in *Noncommutative geometry and number theory*, edited by C. Consani and M. Marcolli, Aspects Math. **E37**, Vieweg, Wiesbaden, 2006. MR 2327298 Zbl 1183.11023

[Böckle 2010] G. Böckle, "Deformation rings for some mod 3 Galois representations of the absolute Galois group of $\mathbf{Q}_3$", pp. 529–542 in *p-adic representations of p-adic groups, II: Representations of* $\mathrm{GL}_2(\mathbb{Q}_p)$ *and* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2642411 Zbl 1223.11141

[Breuil 2003a] C. Breuil, "Sur quelques représentations modulaires et p-adiques de $\mathrm{GL}_2(\mathbb{Q}_p)$, I", *Compositio Math.* **138**:2 (2003), 165–188. MR 2018825 Zbl 1044.11041

[Breuil 2003b] C. Breuil, "Sur quelques représentations modulaires et p-adiques de $\mathrm{GL}_2(\mathbb{Q}_p)$, II", *J. Inst. Math. Jussieu* **2**:1 (2003), 23–58. MR 1955206 Zbl 1165.11319

[Breuil and Emerton 2010] C. Breuil and M. Emerton, "Représentations p-adiques ordinaires de $\mathrm{GL}_2(\mathbb{Q}_p)$ et compatibilité local-global", pp. 255–315 in *p-adic representations of p-adic groups, III:*

Global and geometrical methods, edited by L. Berger et al., Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2667890 Zbl 1251.11043

[Chenevier 2009] G. Chenevier, "Sur la variété des caractères p-adiques du groupe de Galois absolu de $\mathbb{Q}_p$", preprint, 2009, available at http://gaetan.chenevier.perso.math.cnrs.fr/articles/lieugalois.pdf.

[Chenevier 2014] G. Chenevier, "The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings", pp. 221–285 in *Automorphic forms and Galois representations* (Durham, 2011), vol. 1, edited by P. K. F. Diamond and M. Kim, LMS Lecture Note Series **414**, Cambridge University Press, 2014. Zbl 1310.11002

[Colmez 2010] P. Colmez, "Représentations de $\mathrm{GL}_2(\mathbb{Q}_p)$ et (ϕ, Γ)-modules", pp. 281–509 in *p-adic representations of p-adic groups, II: Representations of* $\mathrm{GL}_2(\mathbb{Q}_p)$ *and* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2642409 Zbl 1218.11107

[Colmez et al. 2014] P. Colmez, G. Dospinescu, and V. Paškūnas, "The p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$", *Camb. J. Math.* **2**:1 (2014), 1–47. MR 3272011 Zbl 1312.11090

[Colmez et al. 2015] P. Colmez, G. Dospinescu, and V. Paškūnas, "Irreducible components of deformation spaces: wild 2-adic exercises", *Int. Math. Res. Not.* **2015**:14 (2015), 5333–5356. MR 3384443 Zbl 06513140

[Darmon et al. 1997] H. Darmon, F. Diamond, and R. Taylor, "Fermat's last theorem", pp. 2–140 in *Elliptic curves, modular forms & Fermat's last theorem* (Hong Kong, 1993), edited by J. Coates and S. T. Yau, Int. Press, Cambridge, MA, 1997. MR 1605752

[Dospinescu 2015] G. Dospinescu, "Extensions de représentations de de Rham et vecteurs localement algébriques", *Compos. Math.* **151**:8 (2015), 1462–1498. MR 3383164 Zbl 06484396

[Emerton 2010a] M. Emerton, "Ordinary parts of admissible representations of p-adic reductive groups, I: Definition and first properties", pp. 355–402 in *p-adic representations of p-adic groups, III: Global and geometrical methods*, edited by L. Berger et al., Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2667882 Zbl 1205.22013

[Emerton 2010b] M. Emerton, "Ordinary parts of admissible representations of p-adic reductive groups, II: Derived functors", pp. 403–459 in *p-adic representations of p-adic groups, III: Global and geometrical methods*, edited by L. Berger et al., Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2667883 Zbl 1205.22014

[Emerton 2011] M. Emerton, "Local-global compatibility in the p-adic Langlands programme for $\mathrm{GL}_2/\mathbb{Q}$", preprint, 2011, available at http://www.math.uchicago.edu/~emerton/pdffiles/lg.pdf.

[Emerton and Gee 2014] M. Emerton and T. Gee, "A geometric perspective on the Breuil–Mézard conjecture", *J. Inst. Math. Jussieu* **13**:1 (2014), 183–223. MR 3134019 Zbl 1318.11061

[Emerton and Paškūnas 2010] M. Emerton and V. Paškūnas, "On the effaceability of certain δ-functors", pp. 461–469 in *p-adic representations of p-adic groups, III: Global and geometrical methods*, edited by L. Berger et al., Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2667892 Zbl 1198.22010

[Fontaine 1994] J.-M. Fontaine, "Représentations p-adiques semi-stables", pp. 113–184 in *Périodes p-adiques* (Bures-sur-Yvette, 1988), Astérisque **223**, Société Mathématique de France, Paris, 1994. MR 1293972 Zbl 0865.14009

[Gabriel 1962] P. Gabriel, "Des catégories abéliennes", *Bull. Soc. Math. France* **90** (1962), 323–448. MR 0232821 Zbl 0201.35602

[Gee 2011] T. Gee, "Automorphic lifts of prescribed types", *Math. Ann.* **350**:1 (2011), 107–144. MR 2785764 Zbl 1276.11085

[Gee and Kisin 2014] T. Gee and M. Kisin, "The Breuil–Mézard conjecture for potentially Barsotti–Tate representations", *Forum Math. Pi* **2** (2014), art. ID e1. MR 3292675 Zbl 06391704

[Henniart 2002] G. Henniart, "Sur l'unicité des types pour GL_2", (2002). Appendix to C. Breuil and A. Mézard, "Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbb{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ en $l = p$", *Duke Math. J.* **115**:2 (2002), 205–310. Zbl 1042.11030

[Hu and Tan 2015] Y. Hu and F. Tan, "The Breuil–Mezard conjecture for non-scalar split residual representations", *Ann. Sci. Éc. Norm. Supér.* (4) **48**:6 (2015), 1383–1421. MR 3429471 Zbl 1334.11041

[Katz and Messing 1974] N. M. Katz and W. Messing, "Some consequences of the Riemann hypothesis for varieties over finite fields", *Invent. Math.* **23** (1974), 73–77. MR 0332791 Zbl 0275.14011

[Khare and Wintenberger 2009a] C. Khare and J.-P. Wintenberger, "On Serre's conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$", *Ann. of Math.* (2) **169**:1 (2009), 229–253. MR 2480604 Zbl 1196.11076

[Khare and Wintenberger 2009b] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture, II", *Invent. Math.* **178**:3 (2009), 505–586. MR 2551764 Zbl 1304.11042

[Kisin 2008] M. Kisin, "Potentially semi-stable deformation rings", *J. Amer. Math. Soc.* **21**:2 (2008), 513–546. MR 2373358 Zbl 1205.11060

[Kisin 2009a] M. Kisin, "The Fontaine–Mazur conjecture for GL_2", *J. Amer. Math. Soc.* **22**:3 (2009), 641–690. MR 2505297 Zbl 1251.11045

[Kisin 2009b] M. Kisin, "Modularity of 2-adic Barsotti–Tate representations", *Invent. Math.* **178**:3 (2009), 587–634. MR 2551765 Zbl 1304.11043

[Kisin 2009c] M. Kisin, "Moduli of finite flat group schemes, and modularity", *Ann. of Math.* (2) **170**:3 (2009), 1085–1180. MR 2600871 Zbl 1201.14034

[Kisin 2010] M. Kisin, "Deformations of $G_{\mathbb{Q}_p}$ and $\mathrm{GL}_2(\mathbb{Q}_p)$ representations", pp. 511–528 in *p-adic representations of p-adic groups, II: Representations of* $\mathrm{GL}_2(\mathbb{Q}_p)$ *and* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2642410 Zbl 1233.11126

[Paškūnas 2009] V. Paškūnas, "On some crystalline representations of $\mathrm{GL}_2(\mathbb{Q}_p)$", *Algebra Number Theory* **3**:4 (2009), 411–421. MR 2525557 Zbl 1173.22015

[Paškūnas 2010a] V. Paškūnas, "Admissible unitary completions of locally $\mathbb{Q}_p$-rational representations of $\mathrm{GL}_2(F)$", *Represent. Theory* **14** (2010), 324–354. MR 2608966 Zbl 1192.22009

[Paškūnas 2010b] V. Paškūnas, "Extensions for supersingular representations of $\mathrm{GL}_2(\mathbb{Q}_p)$", pp. 317–353 in *p-adic representations of p-adic groups, III: Global and geometrical methods*, edited by L. Berger et al., Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2667891 Zbl 1204.22013

[Paškūnas 2013] V. Paškūnas, "The image of Colmez's Montreal functor", *Publ. Math. Inst. Hautes Études Sci.* **118** (2013), 1–191. MR 3150248 Zbl 1297.22021

[Paškūnas 2014] V. Paškūnas, "Blocks for mod p representations of $\mathrm{GL}_2(\mathbb{Q}_p)$", pp. 231–247 in *Automorphic forms and Galois representations* (94th London Mathematical Society (LMS) - EPSRC Durham symposium, UK, July 18–28, 2011), vol. 2, edited by P. K. F. Diamond and M. Kim, LMS Lecture Note Series **415**, Cambridge University Press, 2014. Zbl 1310.11003

[Paškūnas 2015a] V. Paškūnas, "On 2-adic deformations", preprint, 2015. arXiv 1509.00320

[Paškūnas 2015b] V. Paškūnas, "On the Breuil–Mézard conjecture", *Duke Math. J.* **164**:2 (2015), 297–359. MR 3306557 Zbl 06416950

[Saito 2009] T. Saito, "Hilbert modular forms and p-adic Hodge theory", *Compos. Math.* **145**:5 (2009), 1081–1113. MR 2551990 Zbl 1259.11060

[Schneider and Teitelbaum 2002] P. Schneider and J. Teitelbaum, "Banach space representations and Iwasawa theory", *Israel J. Math.* **127** (2002), 359–380. MR 1900706 Zbl 1006.46053

[Serre 2000] J.-P. Serre, *Local algebra*, Springer, Berlin, 2000. MR 1771925 Zbl 0959.13010

[SGA 3_{II} 1970] M. Demazure and A. Grothendieck, *Schémas en groupes, Tome II: Groupes de type multiplicatif, et structure des schémas en groupes généraux, Exposés VIII–XVIII* (Séminaire de Géométrie Algébrique du Bois Marie 1962–1964), Lecture Notes in Math. **152**, Springer, Berlin, 1970. MR 43 #223b Zbl 0209.24201

[Taylor 2003] R. Taylor, "On icosahedral Artin representations, II", *Amer. J. Math.* **125**:3 (2003), 549–566. MR 1981033 Zbl 1031.11031

[Taylor 2006] R. Taylor, "On the meromorphic continuation of degree two L-functions", *Doc. Math.* **Extra Vol.** (2006), 729–779. MR 2290604 Zbl 1138.11051

[Thorne 2016] J. A. Thorne, "A 2-adic automorphy lifting theorem for unitary groups over CM fields", preprint, 2016, available at http://www.math.harvard.edu/~thorne/p_equals_2.pdf.

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*

ALGEBRA AND NUMBER THEORY 10:6 (2016)
dx.doi.org/10.2140/ant.2016.10.1359

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 Γ. By this we mean a nontrivial word w such that the probabilities that $w = 1$ in the finite quotients of Γ 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 Γ be a finitely generated linear group over any field and let G be its profinite completion. Then either Γ 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 Γ is a finitely generated linear group which is not virtually solvable then its profinite completion $\widehat{\Gamma}$ 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 Γ 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 Γ be a finitely generated linear group over any field. Then either Γ is virtually solvable or Γ is randomly free.*

The proof of this result relies on the notion and properties of probabilistic identities which we introduce below.

Let $w = w(x_1, \dots, x_n)$ be a nontrivial element of the free group F_n, and let Γ 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 Γ. We say that w is a *probabilistic identity* of Γ if there exists $\epsilon > 0$ such that, for each finite quotient $H = \Gamma/\Delta$ of Γ, the probability $P_H(w)$ that $w(h_1, \dots, h_n) = 1$ (where the $h_i \in H$ are chosen independently with respect to the uniform distribution on H) is at least ϵ. This amounts to saying that, in the profinite completion $G = \widehat{\Gamma}$ of Γ, the probability (with respect to the Haar measure) that $w(g_1, \dots, g_n) = 1$ 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 Γ 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 Γ, 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 Γ if there exists a finite index subgroup $\Delta \leq \Gamma$ and cosets $\gamma_1\Delta, \dots, \gamma_n\Delta$ (where $\gamma_i \in \Gamma$) such that $w(\gamma_1\Delta, \dots, \gamma_n\Delta) = \{1\}$.

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 Γ is not virtually solvable, and let G be the profinite completion of Γ. Note that $g_1, \dots, g_n \in G$ freely generate a free subgroup of G if and only if $w(g_1, \dots, g_n) \neq 1$ 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 σ-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 Γ. 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 Γ, though Γ need not be virtually solvable. However, we can show the following.

Theorem 1.4. *Let Γ be a finitely generated linear group over any field. Then every probabilistic identity* (*possibly with parameters*) *on Γ is a coset identity.*

It easily follows that, if w is a word in n variables (possibly with parameters from Γ), and $\gamma \in \Gamma$ is such that in all finite quotients $H = \Gamma/\Delta$ of Γ 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 Γ be a linear group and let n be a positive integer. Let us say that a subset Ξ of Γ^n is *Zariski-closed* if there is an embedding of Γ in $\mathrm{GL}_r(F)$ (for some field F and a positive integer r) and a Zariski-closed subset X of GL_r^n such that $\Xi = X(F) \cap \Gamma^n$.

Then we have the following.

Theorem 1.5. *Let Γ 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 Δ of Γ. 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* Ξ *in the profinite group* $(\widehat{\Gamma})^n$ *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 $T^k \leq G \leq \mathrm{Aut}(T^k)$ 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 Γ into a "linear algebraic group" G over an infinite product $\prod_{\mathfrak{m}} A/\mathfrak{m}$ 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 $X\left(\prod_{\mathfrak{m}} A/\mathfrak{m}\right) \cap \Gamma^n$ is positive. Every translate of X by an element of Γ^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 Γ^n with positive measure; intersecting X with a suitable translate by an element of Γ^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* Γ *is a finitely generated linear group which is not virtually solvable. Then all fibers in* $(\widehat{\Gamma})^n$ *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 $w(h_1, \dots, h_n) = h$). Its ℓ_∞-norm is defined by $\| P_{H,w} \|_\infty = \max_{h \in H} P_{H,w}(h)$. Then we have:

Theorem 1.6. *Let* Γ *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* Γ *we have* $\| P_{H,w} \|_\infty \geq \epsilon$. *Then* Γ *is virtually solvable.*

See also [Aoun 2011] for a different probabilistic Tits alternative, related to certain random walks on the discrete linear group Γ.

2. Proof of Theorems 1.4 and 1.2

If a group Γ acts on a topological space X and $Y \subseteq X$, we say Y is Γ-*finite* if its orbit under Γ is finite. We say a closed subset $Z \subseteq X$ is Γ-*covered* by Y if Z is a closed subset of some finite union of Γ-translates of Y.

Lemma 2.1. *Let Γ be a group acting on a set X. If $Y_1, \dots, Y_n$ are subsets of X which are not Γ-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 Γ. 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. □

Proposition 2.2. *Let X be a Noetherian topological space and Γ 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 Γ-covers some closed subset $W \subseteq X$ with $f(W) > 0$, then Y Γ-covers some closed Γ-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 Γ-covering a set of positive f-value. If two distinct irreducible components Y_i and Y_j of Y were Γ-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 Γ-cover a set of positive f-value. This is impossible by the minimality of Y.

If Y is Γ-finite, then
$$Z := \bigcup_{g\in\Gamma} gY$$
is a Γ-stable finite union of Γ-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 Γ-finite. We write $Y = Y_0 \cup Y'$, where no Γ-translate of Y' contains Y_0.

By condition (I), there exists a finite sequence $g_1, \dots, g_r \in \Gamma$ such that $f(Z) > 0$ for
$$Z := g_1 Y \cup \cdots \cup g_r Y.$$

We choose the g_i so that

$$f(Z) > \frac{\sup_{\Delta\subsetneq\Gamma \text{ finite}} f\big(\bigcup_{g\in\Delta} gY\big)}{2}. \tag{2-1}$$

As no Γ-translate of Y_0 is Γ-finite, Lemma 2.1 implies that there exists g such that $g_iY_0 \neq gg_jY_0$ for all i, j. Thus,

$$Y' \cup \bigcup_{i,j}(Y_0 \cap g_i^{-1}gg_jY_0) \subsetneq Y$$

Γ-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 Γ-finite. □

Now, let A be an integral domain finitely generated over $\mathbb{Z}$ with fraction field K. Let $\mathcal{G} = \operatorname{Spec} B$ 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 $\operatorname{Spec} T$-points of $\mathcal{G}\to\operatorname{Spec} A$, 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 $\operatorname{Spec} T\to\mathcal{G}$ sends $\operatorname{Spec} T$ 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

$$P(\mathcal{G},A) := \prod_{\mathfrak{m}\in\operatorname{Maxspec}(A)} \mathcal{G}(A/\mathfrak{m}),$$

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 $A/\mathfrak{m}$ is a field finitely generated over $\mathbb{Z}$ (and hence finite), it follows that each $\mathcal{G}(A/\mathfrak{m})$ is finite and $P(\mathcal{G},A)$ is a profinite group. For any closed subset $X\subseteq\mathcal{G}$, we define the closed subset

$$P(X,A) := \prod_{\mathfrak{m}\in\operatorname{Maxspec}(A)} X(A/\mathfrak{m}) \subseteq P(\mathcal{G},A).$$

Lemma 2.3. *If $X\subseteq\mathcal{G}$ does not meet the generic fiber* $\operatorname{Spec} B\otimes_A K\subset\mathcal{G}$, *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

$$\sum_i b_i \otimes \frac{a_i}{a_i'} = 1,$$

and clearing denominators we see that some nonzero element $a' := \prod_i a_i' \in A$ belongs to I. If $\mathfrak{m}$ is a maximal ideal of $A[1/a']$, then $A[1/a']/\mathfrak{m}$ is a field finitely generated over $\mathbb{Z}$, hence a finite field, and therefore $\mathfrak{m} \cap A$ is a maximal ideal of A. Thus, the image of a' in $A/(\mathfrak{m} \cap A)$ is nonzero, from which it follows that there are no A-homomorphisms $B/I \to A/(\mathfrak{m} \cap A)$, i.e., $X(A/(\mathfrak{m} \cap A)) = \varnothing$. □

For any subgroup $\Gamma \subseteq \mathcal{G}(A) \subseteq P(\mathcal{G}, A)$, we define $\bar{\Gamma}$ to be the closure of Γ in $P(\mathcal{G}, A)$. This is a closed subgroup of a profinite group and therefore a profinite group itself. We endow it with Haar measure $\mu_{\bar{\Gamma}}$, normalized so that $(\bar{\Gamma}, \mu_{\bar{\Gamma}})$ is a probability space. In particular, left translation by Γ is a continuous measure-preserving action on $(\bar{\Gamma}, \mu_{\bar{\Gamma}})$. As Haar measure is outer regular, for every Borel set B,

$$\mu_{\bar{\Gamma}}(B) = \inf_{S \subseteq \mathrm{Maxspec}(A)} \frac{|\mathrm{pr}_S B|}{|\mathrm{pr}_S \Gamma|},$$

where S ranges over all finite sets of maximal ideals of A and pr_S denotes projection onto $\prod_{\mathfrak{m} \in S} \mathcal{G}(A/\mathfrak{m})$.

For any positive integer n, we let $\mathcal{G}^n$ denote the n-th fiber power of $\mathcal{G}$ relative to A, i.e., defining

$$B_n := \underbrace{B \otimes_A B \otimes_A \cdots \otimes_A B}_{n},$$

we define $\mathcal{G}^n := \mathrm{Spec}\, B_n$, regarded as a topological space with respect to the Zariski topology. Note that in general the Zariski topology on $\mathcal{G}^n$ is *not* the product topology. However, by the universal property of tensor products, $\mathcal{G}^n(T)$ is canonically isomorphic to $\mathcal{G}(T)^n$ 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 $\mathcal{G}^n$ is a Noetherian topological space.

We consider the closure $\bar{\Gamma}^n$ of Γ^n in $P(\mathcal{G}^n, A)$. For any closed subset $Y \subseteq \mathcal{G}^n$, we define

$$P_\Gamma(Y) := \bar{\Gamma}^n \cap P(Y, A).$$

Thus, if Y and Z are closed subsets of $\mathcal{G}^n$,

$$P_\Gamma(Y \cap Z) = \bar{\Gamma}^n \cap P(Y \cap Z, A) = \bar{\Gamma}^n \cap (P(Y, A) \cap P(Z, A)) = P_\Gamma(Y) \cap P_\Gamma(Z).$$

As

$$P(Y \cup Z, A) = \prod_{\mathfrak{m} \in \mathrm{Maxspec}(A)} (Y(A/\mathfrak{m}) \cup Z(A/\mathfrak{m})) \supseteq P(Y, A) \cup P(Z, A),$$

we have

$$P_\Gamma(Y \cup Z) \supseteq P_\Gamma(Y) \cup P_\Gamma(Z).$$

Defining

$$f(Y) := \mu_{\bar{\Gamma}^n}(P_\Gamma(Y)),$$

condition (I) of Proposition 2.2 is obvious. As $\mu_{\bar{\Gamma}^n}$ is a measure, if $f(Y \cap Z) = 0$, then

$$\begin{aligned} f(Y \cup Z) &= \mu_{\bar{\Gamma}^n}(P_\Gamma(Y \cup Z)) \\ &\geq \mu_{\bar{\Gamma}^n}(P_\Gamma(Y) \cup P_\Gamma(Z)) \\ &= \mu_{\bar{\Gamma}^n}(P_\Gamma(Y)) + \mu_{\bar{\Gamma}^n}(P_\Gamma(Z)) - \mu_{\bar{\Gamma}^n}(P_\Gamma(Y) \cap P_\Gamma(Z)) \\ &= f(Y) + f(Z) - f(Y \cap Z) = f(Y) + f(Z). \end{aligned}$$

As $\mu_{\bar{\Gamma}^n}$ is Γ^n-invariant, this implies condition (II).

Proposition 2.4. *Let G denote a linear algebraic group over a field K. If Γ is Zariski-dense in $G(K)$, then a nonempty closed subset Y of G^n is Γ^n-finite if and only if it is a union of connected components of G^n.*

Proof. If Y is Γ^n-finite, its stabilizer Δ is of finite index in Γ^n, which implies that the Zariski closure D of Δ 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 Γ^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. □

We can now prove Theorem 1.4.

Proof. We fix a faithful representation $\rho : \Gamma \to \mathrm{GL}_r(F)$, where F is an algebraically closed field. Let $G \subset \mathrm{GL}_r$ denote the Zariski closure of Γ in GL_r.

We recall how to extend G to a subgroup scheme of GL_r defined over a finitely generated $\mathbb{Z}$-algebra. Let

$$R_{\mathbb{Z},r} := \mathbb{Z}[x_{ij}, y]_{i,j=1,\dots,r}/(y \det(x_{ij}) - 1)$$

denote the coordinate ring of 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 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:

$$\begin{aligned}\Delta_{A,r}(I) &\subseteq I \otimes_A R_{A,r} + R_{A,r} \otimes_A I,\\ S_{A,r}(I) &\subseteq I,\\ \epsilon_{A,r}(I) &= \{0\}.\end{aligned}$$

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 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 $\mathrm{GL}_r(F)$ of the $\rho(g_j)$, as g_j runs over some finite generating set of Γ, 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 $\mathcal{G} := \operatorname{Spec} R_{A,r}/I$, the closed group subscheme of 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 Γ (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 Γ-covers a set of positive f-value, so by Proposition 2.2 Y Γ-covers a closed Γ-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

$$\frac{|\mathrm{pr}_S\, w^{-1}(1)|}{|\mathrm{pr}_S\, \Gamma^n|} < \epsilon.$$

Defining Δ to be the kernel of pr_S, we see that, in the finite quotient Γ/Δ, the probability that the word map w attains the value $1+\Delta$ is less than ϵ. It follows that w is not a probabilistic identity on Γ. This contradiction completes the proof of Theorem 1.4. $\square$

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° be the identity component, T a maximal torus of G° and B a Borel subgroup of G° containing T. Let Φ be the root system of G with respect to T, and let Φ^+ denote the set of roots of B with respect to T. Every maximal torus of G° 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 α of Φ^+ is determined by B, so h likewise preserves α. It therefore normalizes $\ker\alpha^\circ$, and therefore the derived group G_α 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_α in G meets every connected component of G.

Suppose now that w is constant on $g_1G^\circ\times\cdots\times g_nG^\circ$ for some $g_1, \dots, g_n \in G(K)$. Without loss of generality we may assume that all g_i centralize G_α. As w is constant on $g_1G_\alpha \times \cdots \times g_nG_\alpha$, and as

$$w(g_1h_1, \dots, g_nh_n) = w(g_1, \dots, g_n)w(h_1, \dots, h_n)$$

for all $h_1, \dots, h_n \in G_\alpha(K)$, it follows that w is constant on G_α^n. This is impossible because nontrivial words give nontrivial word maps on all semisimple algebraic groups [Borel 1983]. $\square$

Proof of Theorem 1.2. Every virtually solvable linear group satisfies a nontrivial identity. In the other direction, if $\Gamma \subset \mathrm{GL}_r(K)$ 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°, 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 Γ. $\square$

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 $1 \neq w \in F_n$, any positive integer d and any real number $\epsilon > 0$, there exists a word $1 \neq v \in F_m$ (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 ϵ). 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 ϵ 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 Γ be a finitely generated residually finite group. Then either Γ satisfies an identity or Γ is randomly free.*

In particular:

(i) *If Γ does not satisfy an identity then $\widehat{\Gamma}$ has a nonabelian free subgroup.*

(ii) *If $\widehat{\Gamma}$ has a nonabelian free subgroup then almost all n-tuples in $\widehat{\Gamma}$ 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 $1 \neq w = w(x_1, \dots, x_n) \in F_n$, 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 $g_1, \dots, g_n \in G$ generating G such that $w(g_1, \dots, g_n) \neq 1$ (and then ϕ 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 $(g_1, \dots, g_n) \in G^n$, we have $w(g_1, \dots, g_n) = 1$. 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

[Abért 2005] M. Abért, "Group laws and free subgroups in topological groups", *Bull. London Math. Soc.* **37**:4 (2005), 525–534. MR 2143732 Zbl 1095.20001

[Aoun 2011] R. Aoun, "Random subgroups of linear groups are free", *Duke Math. J.* **160**:1 (2011), 117–173. MR 2838353 Zbl 1239.20051

[Borel 1983] A. Borel, "On free subgroups of semisimple groups", *Enseign. Math.* (2) **29**:1-2 (1983), 151–164. MR 702738 Zbl 0533.22009

[Breuillard and Gelander 2007] E. Breuillard and T. Gelander, "A topological Tits alternative", *Ann. of Math.* (2) **166**:2 (2007), 427–474. MR 2373146 Zbl 1149.20039

[Dixon et al. 2003] J. D. Dixon, L. Pyber, Á. Seress, and A. Shalev, "Residual properties of free groups and probabilistic methods", *J. Reine Angew. Math.* **556** (2003), 159–172. MR 1971144 Zbl 1027.20013

[Epstein 1971] D. B. A. Epstein, "Almost all subgroups of a Lie group are free", *J. Algebra* **19** (1971), 261–262. MR 0281776 Zbl 0222.22012

[Jaikin-Zapirain and Pyber 2011] A. Jaikin-Zapirain and L. Pyber, "Random generation of finite and profinite groups and group enumeration", *Ann. of Math.* (2) **173**:2 (2011), 769–814. MR 2776362 Zbl 1234.20042

[Khukhro 1986] E. I. Khukhro, "Locally nilpotent groups that admit a splitting automorphism of prime order", *Mat. Sb.* (*N.S.*) **130(172)**:1 (1986), 120–127, 128. MR 847346 Zbl 0608.20025

[Larsen and Shalev 2012] M. Larsen and A. Shalev, "Fibers of word maps and some applications", *J. Algebra* **354** (2012), 36–48. MR 2879221 Zbl 1258.20011

[Mann 1994] A. Mann, "Finite groups containing many involutions", *Proc. Amer. Math. Soc.* **122**:2 (1994), 383–385. MR 1242094 Zbl 0811.20024

[Neumann 1954] B. H. Neumann, "Groups covered by finitely many cosets", *Publ. Math. Debrecen* **3** (1954), 227–242. MR 0072138 Zbl 0057.25603

[Neumann 1989] P. M. Neumann, "Two combinatorial problems in group theory", *Bull. London Math. Soc.* **21**:5 (1989), 456–458. MR 1005821 Zbl 0695.20018

[Szegedy 2005] B. Szegedy, "Almost all finitely generated subgroups of the Nottingham group are free", *Bull. London Math. Soc.* **37**:1 (2005), 75–79. MR 2105821 Zbl 1073.20022

[Tits 1972] J. Tits, "Free subgroups in linear groups", *J. Algebra* **20** (1972), 250–270. MR 0286898 Zbl 0236.20032

[Waterhouse 1979] W. C. Waterhouse, *Introduction to affine group schemes*, Graduate Texts in Mathematics **66**, Springer, New York-Berlin, 1979. MR 547117 Zbl 0442.14017

Communicated by Efim Zelmanov
Received 2015-10-29 Revised 2016-05-01 Accepted 2016-05-31

mjlarsen@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*

Guidelines for Authors

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 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.

Algebra & Number Theory

Volume 10 No. 6 2016