Annals of Mathematics **191** (2020), 329–354
https://doi.org/10.4007/annals.2020.191.2.1

A converse to a theorem of Gross, Zagier, and Kolyvagin

By Christopher Skinner

Abstract

Let E be a semistable elliptic curve over $\mathbb{Q}$. We prove that if E has non-split multiplicative reduction at at least one odd prime or split multiplicative reduction at at least two odd primes, then

$$\mathrm{rank}_{\mathbb{Z}} E(\mathbb{Q}) = 1 \text{ and } \#\text{Ш}(E) < \infty \implies \mathrm{ord}_{s=1} L(E,s) = 1.$$

We also prove the corresponding result for the abelian variety associated with a weight 2 newform f of trivial character. These, and other related results, are consequences of our main theorem, which establishes criteria for f and $H^1_f(\mathbb{Q}, V)$, where V is the p-adic Galois representation associated with f, that ensure that $\mathrm{ord}_{s=1} L(f,s) = 1$. The main theorem is proved using the Iwasawa theory of V over an imaginary quadratic field to show that the p-adic logarithm of a suitable Heegner point is non-zero.

1. Introduction

Let $f \in S_2(\Gamma_0(N))$ be a newform with trivial nebentypus. Associated with f is an abelian variety A_f over $\mathbb{Q}$ (really an isogeny class of abelian varieties) characterized by an equality of the Hasse–Weil L-function $L(A_f, s)$ of A_f and the product of the L-functions $L(f^\sigma, s)$ of all the Galois conjugates f^σ of f:

$$L(A_f, s) = \prod L(f^\sigma, s).$$

The endomorphism ring $\mathrm{End}^0_{\mathbb{Q}}(A_f)$ is a totally real field M_f of degree equal to the dimension of A_f; this is naturally identified with the subfield of $\mathbb{C}$ generated over $\mathbb{Q}$ by the Hecke eigenvalues of f (equivalently, the Fourier coefficients of the q-expansion of f at ∞), and so its degree $[M_f : \mathbb{Q}]$ is equal to the number of conjugate forms f^σ.

The celebrated Birch–Swinnerton-Dyer conjecture, as formulated for abelian varieties, predicts that the rank of $A_f(\mathbb{Q})$ equals the order of vanishing

Keywords: Birch–Swinnerton-Dyer Conjecture, Iwasawa Theory, Heegner points
AMS Classification: Primary: 11G40, 11R23, 11G05, 11G07.

at $s = 1$ of the L-function $L(A_f, s)$:

$$\mathrm{rank}_{\mathbb{Z}} A_f(\mathbb{Q}) \stackrel{?}{=} \mathrm{ord}_{s=1} L(A_f, s).$$

The most spectacular result to date in the direction of this conjecture follows from the combination of the work of Gross and Zagier [11] and Kolyvagin [15] [16], which together prove the following: Let $r = 0$ or 1. Then $\mathrm{ord}_{s=1} L(f, s) = r$ if and only if $\mathrm{ord}_{s=1} L(A_f, s) = [M_f : \mathbb{Q}]r$, and

$$\mathrm{ord}_{s=1} L(f, s) = r \implies \mathrm{rank}_{\mathbb{Z}} A_f(\mathbb{Q}) = [M_f : \mathbb{Q}]r \text{ and } \#Ш(A_f) < \infty,$$

where $Ш(A_f)$ is the Tate–Shafarevich group of A_f (conjecturally always finite). For $r = 0$, the converse to this implication was established in [24] (for N not squarefull) and in [25] (for all N). In the proofs of these converses it is only needed that the p-primary part $Ш(A_f)[p^\infty]$ be finite for a sufficiently large prime p such that f is ordinary with respect to some prime $\lambda \mid p$ of M_f. In particular, if A_f is a semistable elliptic curve, then it suffices to take $p \geq 11$ of good, ordinary reduction.

For $r = 1$ and A_f an elliptic curve having complex multiplication, a converse to the above implication is a consequence of the combination of results of Rubin, Bertrand, and Perrin-Riou; this is explained in Theorems 8.1 and 8.2 of [22]. In this paper we prove a converse for the $r = 1$ case for those A_f with N squarefree. If N is squarefree, then A_f does not have complex multiplication, so the cases covered by the theorems in this paper are disjoint from those covered by the results recalled in [22].

Let $\pi = \otimes \pi_v$ be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A})$ such that $L(\pi, s - 1/2) = L(f, s)$ (so N is the conductor of π).

Theorem A. *Suppose N is squarefree. If there is at least one odd prime ℓ such that π_ℓ is the twist of the special representation by the unique unramified quadratic character or at least two odd primes ℓ_1 and ℓ_2 such that π_{ℓ_1} and π_{ℓ_2} are special, then*

$$\mathrm{rank}_{\mathbb{Z}} A_f(\mathbb{Q}) = [M_f : \mathbb{Q}] \text{ and } \#Ш(A_f) < \infty \implies \mathrm{ord}_{s=1} L(f, s) = 1.$$

The hypotheses on the local representations can also be formulated in terms of the eigenvalues of the Atkin–Lehner involutions w_ℓ acting on f: either there is at least one odd prime ℓ such that the eigenvalue of w_ℓ is $+1$ or there are at least two odd primes ℓ_1 and ℓ_2 such that the eigenvalues of w_{ℓ_1} and w_{ℓ_2} are both -1.

Since every elliptic curve over $\mathbb{Q}$ is modular, Theorem A can be restated in this case to read

Theorem A′. *Suppose E is a semistable elliptic curve over $\mathbb{Q}$. If there is at least one odd prime at which E has nonsplit multiplicative reduction or* *at least two odd primes at which E has split multiplicative reduction, then*

$$\mathrm{rank}_{\mathbb{Z}} E(\mathbb{Q}) = 1 \text{ and } \#Ш(E) < \infty \implies \mathrm{ord}_{s=1} L(E,s) = 1.$$

As the hypotheses on $E(\mathbb{Q})$ and $Ш(E)$ ensure that the root number of E is -1 (by, for instance, Nekovář's results toward the parity conjecture), the hypotheses on the reduction types in Theorem A′ only exclude those semistable curves that have conductor equal to 2ℓ with ℓ an odd prime and that have split reduction at both 2 and ℓ. However, showing that a positive proportion of semistable elliptic curves E, when ordered by height for example, satisfy the hypotheses of Theorem A′ with $\mathrm{rank}_{\mathbb{Z}} E(\mathbb{Q}) = 1$ and $\#Ш(E) < \infty$ remains an open and interesting problem.

Our proof of Theorem A is similar in spirit to those of Theorems 8.1 and 8.2 of [22] insofar as it is essentially p-adic. We deduce Theorem A from Theorem B below, which gives a p-adic criterion for A_f to have both algebraic and analytic rank $[M_f : \mathbb{Q}]$ over an imaginary quadratic field. However, unlike the proofs in [22], our proof of this criterion does not make use of a p-adic Gross–Zagier formula for the derivative of a p-adic L-function or require non-degeneracy of p-adic heights. Instead it uses a formula expressing the value of a p-adic L-function in terms of the formal log of a rational point on A_f.

Let p be a prime and $\lambda \mid p$ a prime of M_f. Let $L = M_{f,\lambda}$. Let $\rho_{f,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_L(V)$ be the usual two-dimensional Galois representation associated with f and let $\bar{\rho}_{f,\lambda}$ be its residual representation (the semisimplification of the reduction of a Galois stable lattice in V).

THEOREM B. *Suppose N is squarefree and $p \geq 5$. Let K be an imaginary quadratic field with odd discriminant D. Suppose*

(a) *$p \nmid N$ and f is ordinary with respect to λ;*
(b) *$\bar{\rho}_{f,\lambda}$ is irreducible and ramified at an odd prime that is either inert or ramified in K;*
(c) *both 2 and p split in K;*
(d) *if $(D,N) \neq 1$, then for each $\ell \mid (D,N)$, π_ℓ is the twist of the special representation by the unique unramified quadratic character, and each prime divisor of $N/(D,N)$ splits in K;*
(e) *$\dim_L H^1_f(K,V) = 1$ and the restriction $H^1_f(K,V) \xrightarrow{\mathrm{res}} \prod_{v\mid p} H^1_f(K_v,V)$ is an injection.*

Then $\mathrm{ord}_{s=1} L(f,K,s) = 1$, $\mathrm{rank}_{\mathbb{Z}} A_f(K) = [M_f : \mathbb{Q}] = \mathrm{ord}_{s=1} L(A_f/K,s)$, *and* $Ш(A_f/K)$ *is finite.*

Here $H^1_f(K,V)$ and $H^1_f(K_v,V)$ are, respectively, the global and local Bloch–Kato Selmer groups. Also, $L(A_f/K,s)$ and $Ш(A_f/K)$ are the L-function and the Tate–Shafarevich group of A_f/K, and $L(f,K,s) = L(f,s)L(f \otimes \chi_D, s)$ with χ_D the quadratic character associated with K.

The deduction of Theorem A from Theorem B, which is explained in more detail in Section 3, goes as follows: As f is not a CM form (since N is squarefree), condition (a) holds for some $\lambda \mid p$ for a set of primes p of density one, while $\bar{\rho}_{f,\lambda}$ is irreducible if p is sufficiently large, in which case it follows from Ribet's work on level-lowering that $\bar{\rho}_{f,\lambda}$ is ramified at some odd prime $q \neq p$ and even, for p sufficiently large, ramified at all primes $q \mid\mid N$. Fix such a p and λ. From the hypotheses on π_ℓ in Theorem A it then follows that an imaginary quadratic field K can be chosen so that its discriminant D is odd; (b), (c), and (d) hold; and $L(A_f^D, 1) \neq 0$, where A_f^D is the K-twist of A_f. As $L(A_f^D, s) = \prod L(f^\sigma \otimes \chi_D, s)$, the existence of a K with the desired properties follows from [9]. From the non-vanishing of $L(A_f^D, 1)$ it follows that $A_f^D(\mathbb{Q})$ and $Ш(A_f^D)$ are finite. Together with $A_f(\mathbb{Q})$ having rank $[M_f : \mathbb{Q}]$ and $Ш(A_f)$ being finite, this implies that $A_f(K)$ has rank $[M_f : \mathbb{Q}]$ and $Ш(A_f/K)$ is finite, which in turn imply (e). It then follows from Theorem B that $\mathrm{ord}_{s=1} L(A_f/K, s) = [M_f : \mathbb{Q}]$ and $\mathrm{ord}_{s=1} L(f, K, s) = 1$. As $L(A_f/K, s) = L(A_f, s)L(A_f^D, s)$ and $L(f, K, s) = L(f, s)L(f \otimes \chi_D, s)$, it follows that $\mathrm{ord}_{s=1} L(A_f, s) = [M_f : \mathbb{Q}]$ and $\mathrm{ord}_{s=1} L(f, s) = 1$.

As the deduction of Theorem A from Theorem B shows, the hypothesis that $Ш(A_f)$ is finite can be replaced with $Ш(A_f)[p^\infty]$ finite for some suitable prime p. It is even possible to formulate conditions on the λ-Selmer group of A_f/K that ensure that hypothesis (e) of Theorem B holds, from which one can deduce, for example,

Theorem C. *Let E be a semistable curve over $\mathbb{Q}$ such that there is at least one odd prime at which E has nonsplit multiplicative reduction or at least two odd primes at which E has split multiplicative reduction. Suppose there is a prime $p \geq 5$ at which E has good, ordinary reduction and such that*

(a) *$E[p]$ is an irreducible $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-representation;*
(b) $\mathrm{Sel}_p(E) \cong \mathbb{Z}/p\mathbb{Z}$;
(c) *the image of the restriction map $\mathrm{Sel}_p(E) \to E(\mathbb{Q}_p)/pE(\mathbb{Q}_p)$ does not lie in the image of $E(\mathbb{Q}_p)[p]$.*

Then $\mathrm{ord}_{s=1} L(E, s) = 1 = \mathrm{rank}_{\mathbb{Z}} E(\mathbb{Q})$ *and* $\#Ш(E) < \infty$.

It is through similar variations that Theorem B plays a crucial role in a recent proof that

Theorem D (Bhargava–Skinner [3]). *When ordered by height, a positive proportion of elliptic curves has both algebraic and analytic rank one.*

To explain how to pass from the hypotheses of Theorem B to its conclusion, we begin by recalling the Gross–Zagier formula. It follows from hypothesis (e) and the parity conjecture for Selmer groups of p-ordinary modular forms

[18] that the sign of the functional equation of $L(f,K,s)$ is -1. It then follows from the general Gross–Zagier formula of X. Yuan, S. Zhang, and W. Zhang [27] that A_f is a quotient of the Jacobian of some Shimura curve (possibly a modular curve) such that the Néron–Tate height of the image $P_K(f) \in A_f(K) \otimes M_f$ of a certain 0-cycle on the curve (essentially a Heegner cycle) is related to $L'(f,K,1)$ by

$$\langle P_K(f), P_K(f) \rangle \doteq L'(f,K,1),$$

where $\langle \cdot, \cdot \rangle$ is the Néron–Tate height-pairing (relative to some symmetric ample line bundle) and "$\doteq$" denotes equality up to a non-zero constant (which depends on f and the line bundle). So if we expect to prove that $\mathrm{ord}_{s=1} L(f,K,1) = 1$, then we should expect to prove that the height of $P_K(f)$ is non-zero or, equivalently, that $P_K(f) \neq 0$. If $P_K(f) \neq 0$, then (e) together with the general Gross–Zagier formula and the work of Kolyvagin implies the conclusions of Theorem B. We establish Theorem B by proving that $P_K(f) \neq 0$.

To show that $P_K(f) \neq 0$ we do not directly show that its height is non-zero. Instead we show that its formal logarithm at a prime of K above p does not vanish, which is sufficient for our purposes. To do this we make use of p-adic analogs of the Gross–Zagier formula, proved by Bertolini, Darmon, and Prasanna and Brooks, which are analogs of a formula proved by Rubin [22] in the CM case. Recall that $p = \mathfrak{p}\bar{\mathfrak{p}}$ splits in K and that D is odd. As explained in [2] and [6] there is a p-adic L-function $L_{\mathfrak{p}}^S(f,\chi)$, a function of certain p-adic anti-cyclotomic Hecke characters χ of K, such that

$$L_{\mathfrak{p}}^S(f,1) \doteq (\log_\omega P_K(f))^2,$$

where $\log_\omega : A_f(K_{\mathfrak{p}}) \otimes L \to \overline{K}_{\mathfrak{p}}$ is the formal logarithm, determined by a certain 1-form $\omega \in \Omega^1(A_f) \otimes M_f$, and "$\doteq$" again denotes equality up to a non-zero constant. Our aim then is to show that $L_{\mathfrak{p}}^S(f,1) \neq 0$ under the hypotheses of Theorem B. Our method for doing so is via Iwasawa theory.

Iwasawa theory conjecturally relates the p-adic L-function $L_{\mathfrak{p}}^S(f,\chi)$ to the characteristic ideal of a certain p-adic Selmer group. One consequence of such a relation would be the implication

$$L_{\mathfrak{p}}^S(f,1) = 0 \implies H_{\mathfrak{p}}^1(K,V) \neq 0,$$

where $H_{\mathfrak{p}}^1(K,V) \subset H^1(K,V)$ is the subspace of classes that vanish in $H^1(K_w,V)$ for all places $w \neq \bar{\mathfrak{p}}$. However, hypothesis (e) of Theorem B ensures that $H_{\mathfrak{p}}^1(K,V) = 0$, so it would follow from this implication that $L_{\mathfrak{p}}^S(f,1) \neq 0$ and hence that $P_K(f)$ is non-torsion. Our strategy for proving Theorem B ultimately reduces to the above implication. The desired result from Iwasawa theory is part of recent work of Wan [26], following the methods of [24], under certain hypotheses on f, p, and K. The conditions (a)–(d) of Theorem B ensure that these hypotheses hold.

Following the proof of Theorem B, we include remarks emphasizing where in the arguments the various hypotheses intervene, with an eye toward future developments that should remove many of them. We then elaborate on the deduction of Theorem A from Theorem B and explain how similar arguments can be applied to the $r = 0$ case, giving an alternate proof of a special case of the results in [24] and [25].

After the first version of this paper was completed, Wei Zhang released a preprint (since published as [28]) in which he proves many cases of a conjecture of Kolyvagin, showing that the p-adic Selmer group of A_f is often spanned by classes derived from Heegner points. As a consequence Zhang obtains a theorem similar to Theorem B. This theorem does not require the restriction map at primes above p be injective (the second half of hypothesis (e) of Theorem B) but crucially requires that the Tamagawa factors at the primes that split in K or are congruent to ± 1 modulo p be indivisible by p. Theorem B imposes no hypotheses on Tamagawa factors. While there is substantial overlap in the cases covered by Theorem B and the results of [28], neither subsumes the other. The proof of the main result in [28] also relies on Iwasawa theory, in this case on consequences of the Main Conjecture for GL_2 proved in [24].

Acknowledgements. The author thanks Kartik Prasanna and Xin Wan for helpful conversations. This work was partially supported by grants from the National Science Foundation, including DMS-0701231 and DMS-0758379. The first version of this paper was written while the author was a visitor in the School of Mathematics at the Institute for Advanced Study.

2. **The proof of Theorem B**

Let $\overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$ and $K/\mathbb{Q}$ an imaginary quadratic field in $\overline{\mathbb{Q}}$. Fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$. This determines a complex conjugation $c \in G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, which induces the non-trivial automorphism on K. For each prime $\mathfrak{l}$ of K, let $\overline{K}_{\mathfrak{l}}$ be an algebraic closure of $K_{\mathfrak{l}}$ and fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{K}_{\mathfrak{l}}$; the latter realizes $G_{K_{\mathfrak{l}}} = \mathrm{Gal}(\overline{K}_{\mathfrak{l}}/K_{\mathfrak{l}})$ as a decomposition subgroup for $\mathfrak{l}$ in $G_K = \mathrm{Gal}(\overline{\mathbb{Q}}/K)$. Let $I_{\mathfrak{l}} \subset G_{K_{\mathfrak{l}}}$ be the inertia subgroup. Let $\mathbb{F}_{\mathfrak{l}}$ be the residue field of $K_{\mathfrak{l}}$ and $\overline{\mathbb{F}}_{\mathfrak{l}}$ the residue field of $\overline{K}_{\mathfrak{l}}$ (so $\overline{\mathbb{F}}_{\mathfrak{l}}$ is an algebraic closure of $\mathbb{F}_{\mathfrak{l}}$); there is then a canonical isomorphism $G_{K_{\mathfrak{l}}}/I_{\mathfrak{l}} \xrightarrow{\sim} G_{\mathbb{F}_{\mathfrak{l}}} = \mathrm{Gal}(\overline{\mathbb{F}}_{\mathfrak{l}}/\mathbb{F}_{\mathfrak{l}})$.

Let p be an odd prime.

2.1. *Modular forms and abelian varieties.* Let $f \in S_2(\Gamma_0(N))$ be a newform with trivial Nebentypus. Let M_f be the subfield of $\mathbb{C}$ generated by the Hecke eigenvalues of f (equivalently, the Fourier coefficients of the q-expansion of f at the cusp ∞); this is a totally real number field, and the fixed embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ identifies M_f with a subfield of $\overline{\mathbb{Q}}$.

A construction of Eichler and Shimura associates with f an abelian variety A_f over $\mathbb{Q}$ of dimension $[M_f : \mathbb{Q}]$ and such that $\mathrm{End}^0_{\mathbb{Q}}(A_f)$ is naturally identified with M_f and characterized (up to isogeny) by

$$L(A_f, s) = \prod_{\sigma: M_f \hookrightarrow \mathbb{C}} L(f^\sigma, s),$$

where f^σ is the conjugate of f; that is, the newform in $S_2(\Gamma_0(N))$ whose q-expansion at ∞ has coefficients obtained by applying σ to those of f.

Let T_pA_f be the p-adic Tate-module of A_f, and let $V_pA_f = T_pA_f \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$. Let λ be a prime of M_f above p, let $M_{f,p} = M_f \otimes \mathbb{Q}_p$, and let L be a finite (field) extension of $M_{f,\lambda}$. Then

$$V = V_pA_f \otimes_{M_{f,p}} L$$

is a two-dimensional L-space with a continuous, L-linear $G_{\mathbb{Q}}$-action, which we denote by $\rho_{f,\lambda}$. It is potentially semistable at p, unramified at $\ell \nmid Np$, and such that $V^\vee \cong V(-1)$ (the -1-Tate twist of V). Furthermore, if we fix an embedding $L \hookrightarrow \mathbb{C}$ that agrees with the inclusion $M_f \hookrightarrow \mathbb{C}$, then[1] $L(V^\vee, s) = L(f, s)$.

Recall that f is ordinary with respect to λ if the eigenvalue $a_p(f)$ of the action on f of the Hecke operator T_p, or U_p if $p \mid N$ (equivalently, the pth Fourier coefficient of the q-expansion at ∞), is a unit at λ — that is, if $a_p(f)$ is a unit in the ring of integers of L.

By the K-twist of A_f we mean the abelian variety A_f^D over $\mathbb{Q}$ obtained by twisting by the cocycle in $H^1(\mathbb{Q}, \mathrm{Aut}_{\mathbb{Q}}A_f)$ defined by the quadratic character $\chi_D : G_{\mathbb{Q}} \to \{\pm 1\} \subset \mathrm{Aut}_{\mathbb{Q}}A_f$ associated with K (so G_K is the kernel of χ_D). Then

$$V_D = V_pA_f^D \otimes_{M_{f,p}} L \cong V \otimes \chi_D$$

as continuous L-linear representations of $G_{\mathbb{Q}}$ and[2]

$$L(A_f^D, s) = \prod_{\sigma: M_f \hookrightarrow \mathbb{C}} L(f^\sigma \otimes \chi_D, s).$$

The natural map $A_f \times A_f^D \to \mathrm{Res}_{K/\mathbb{Q}}A_f$ is a $\mathbb{Q}$-isogeny with kernel and cokernel annihilated by 2.

Let $\epsilon(f) \in \{\pm 1\}$ be the sign of the functional equation of $L(f, s)$. Let

$$L(f, K, s) = L(f, s)L(f \otimes \chi_D, s).$$

The sign of the functional equation of $L(f, K, s)$ is then $\epsilon(f, K) = \epsilon(f)\epsilon(f \otimes \chi_D)$.

[1]Our conventions for L-functions of potentially semistable Galois representations of $G_{\mathbb{Q}}$ or G_K are geometric: the local Euler factors are defined using the characteristic polynomials of geometric Frobenius elements.

[2]For convenience we will identify the Galois character χ_D with the quadratic Dirichlet character of the same conductor.

Let $\pi = \otimes \pi_v$ be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A})$ such that $L(\pi, s-1/2) = L(f,s)$. Then $L(f,K,s) = L(BC_K(\pi), s-1/2)$, where $BC_K(\pi)$ is the base change of π to an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_K)$. For a prime $\mathfrak{l} \mid \ell$ of K over ℓ, we also write $BC_{K_\mathfrak{l}}(\pi_\ell)$ for the base change of π_ℓ to an admissible representation of $\mathrm{GL}_2(K_\mathfrak{l})$; so $BC_{K_\mathfrak{l}}(\pi_\ell)$ is the $\mathfrak{l}$-constituent of $BC_K(\pi)$.

Let $\epsilon(\pi, K) = \epsilon(BC_K(\pi), 1/2)$ be the global root number of $BC_K(\pi)$. Similarly, for a prime ℓ, let $\epsilon_\ell(\pi,K) = \prod_{\mathfrak{l}|\ell} \epsilon(BC_{K_\mathfrak{l}}(\pi_\ell), 1/2)$ be the product of the local root numbers. Then

$$\epsilon(f,K) = \epsilon(\pi,K) = -\prod_{\mathfrak{l}} \epsilon(BC_{K_\mathfrak{l}}(\pi_\ell), 1/2) = -\prod_{\ell} \epsilon_\ell(\pi,K).$$

If ℓ splits in K, then $\epsilon_\ell(\pi,K) = \epsilon(\pi_\ell,1/2)^2 = +1$ since $BC_{K_\mathfrak{l}}(\pi_\ell) \cong \pi_\ell$, so the local contribution to the global sign $\epsilon(f,K)$ comes only from primes that are inert or ramified in K. Furthermore, if π_ℓ is the special representation σ_ℓ, then $\epsilon_\ell(\pi,K) = -1$ if ℓ is inert or ramified as there is then only one prime $\mathfrak{l}$ of K above ℓ and $BC_{K_\mathfrak{l}}(\sigma_\ell)$ is the special representation. And if π_ℓ is the twist $\sigma_\ell \otimes \xi_\ell$ of the special representation by the unique unramified quadratic character ξ_ℓ, then $\epsilon_\ell(\pi,K) = -1$ if ℓ is inert, as $BC_{K_\mathfrak{l}}(\pi_\ell)$ is then the special representation, and $\epsilon_\ell(\pi,K) = +1$ if ℓ is ramified, as $BC_{K_\mathfrak{l}}(\pi_\ell)$ is then the twist of the special representation by the unique unramified quadratic character of $K_\mathfrak{l}$. Here we have used that the root number of the special representation is -1 and the root number of the twist of the special representation by the unramified quadratic extension is $+1$; [17] and [13, Props. 3.5 and 3.6] are useful references for these and other facts about epsilon factors and root numbers.

2.2. *Selmer groups.* Bloch and Kato [4] (see also [8]) defined Selmer groups for geometric p-adic Galois representations. For the representation V, this Selmer group is

$$H^1_f(K,V) = \ker\{H^1(K,V) \stackrel{\text{res}}{\to} \prod_{\mathfrak{l}} H^1(K_\mathfrak{l},V)/H^1_f(K_\mathfrak{l},V)\},$$

where $H^1_f(K_\mathfrak{l},V) = \ker\{H^1(K_\mathfrak{l},V) \to H^1(K_\mathfrak{l}, B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} V)\}$, with B_{cris} the ring of crystalline periods, if $\mathfrak{l} \mid p$, and $H^1_f(K_\mathfrak{l},V) = H^1(\mathbb{F}_\mathfrak{l}, V^{I_\mathfrak{l}})$ if $\mathfrak{l} \nmid p$. By Tate's local Euler characteristic formula and local duality, if $\mathfrak{l} \nmid p$, then

$$\dim_L H^1(K_\mathfrak{l},V) = \dim_L H^0(K_\mathfrak{l},V) + \dim_L H^2(K_\mathfrak{l},V) = 2\dim_L H^0(K_\mathfrak{l},V),$$

where we have used $V \cong V^\vee(1))$ in the second equality. If follows from the local-global compatibility of V with π_ℓ that $H^0(K_\mathfrak{l},V) = V^{G_{K_\mathfrak{l}}} = 0$ (cf. [19, Lemma 3.1.3]), whence $H^1_f(K_\mathfrak{l},V) = H^1(K_\mathfrak{l},V) = 0$.

Let S be any finite set of primes containing those at which V is ramified (so those dividing pN), and let $G_{K,S}$ be the Galois group over K of the maximal extension of K in $\overline{\mathbb{Q}}$ that is unramified outside the prime ideals dividing those in S. Since $H^1(K_{\mathfrak{l}}, V) = 0$ if $\mathfrak{l} \nmid p$,

$$H^1_f(K, V) = \ker\{H^1(G_{K,S}, V) \xrightarrow{\text{res}} \prod_{\mathfrak{l}|p} H^1(K_{\mathfrak{l}}, V)/H^1_f(K_{\mathfrak{l}}, V)\}.$$

The same definitions can, of course, be made with V replaced by V_D as well as with K replaced by $\mathbb{Q}$ and the primes $\mathfrak{l}$ replaced with rational primes ℓ. Then the restriction map from $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ to G_K induces identifications

$$H^1_f(\mathbb{Q}, V) \oplus H^1_f(\mathbb{Q}, V_D) \xrightarrow{\sim} H^1_f(K, V).$$

The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ acts on the right-hand side above, and the left-hand side is identified with the decomposition into the sum of the subgroups on which $\mathrm{Gal}(K/\mathbb{Q})$ acts trivially and non-trivially, respectively.

Lemma 2.2.1. *Suppose $p \nmid ND$ and f is ordinary with respect to $\lambda \mid p$. If $\dim_L H^1_f(K, V)$ is odd, then $\epsilon(f, K)$ is -1.*

Proof. Since

$$\dim_L H^1_f(K, V) = \dim_L H^1_f(\mathbb{Q}, V) + \dim_L H^1_f(\mathbb{Q}, V_D)$$

is odd, one of $\dim_L H^1_f(\mathbb{Q}, V)$ and $\dim_L H^1_f(\mathbb{Q}, V_D)$ is odd and the other is even. It then follows from the parity conjecture for the Selmer groups of modular forms that are ordinary at λ, proved by Nekovář [18, Th. 12.2.3], that one of the signs $\epsilon(f)$ and $\epsilon(f \otimes \chi_D)$ is -1 and the other is $+1$. Then $\epsilon(f, K) = \epsilon(f)\epsilon(f \otimes \chi_D) = -1$. □

Connections with the Selmer group of A_f. Recall that the p^∞-Selmer group of A_f/K is

$$\mathrm{Sel}_{p^\infty}(A_f/K) = \ker\{H^1(K, A_f[p^\infty]) \xrightarrow{\text{res}} \prod_{\mathfrak{l}} H^1(K_{\mathfrak{l}}, A_f(\overline{K}_{\mathfrak{l}}))\}$$

and the p-primary part of the Tate–Shafarevich group of A_f/K is

$$Ш(A_f/K)[p^\infty] = \ker\{H^1(K, A_f(\overline{\mathbb{Q}}))[p^\infty] \xrightarrow{\text{res}} \prod_{\mathfrak{l}} H^1(K_{\mathfrak{l}}, A_f(\overline{K}_{\mathfrak{l}}))\}$$

and that these sit in the fundamental exact sequence:

$$0 \to A_f(K) \otimes \mathbb{Q}_p/\mathbb{Z}_p \to \mathrm{Sel}_{p^\infty}(A_f/K) \to Ш(A_f/K)[p^\infty] \to 0.$$

Let $W = V_pA_f/T_pA_f = A_f[p^\infty]$ (the last identification being $(x_n) \otimes \frac{1}{p^m} \mapsto x_m$). Then $\mathrm{Sel}_{p^\infty}(A_f/K)$ consists of those classes with restriction at each prime $\mathfrak{l}$ in the image $H^1_f(K_{\mathfrak{l}}, W)$ of the Kummer map $A_f(K_{\mathfrak{l}}) \otimes \mathbb{Q}_p/\mathbb{Z}_p \hookrightarrow H^1(K_{\mathfrak{l}}, W)$. Bloch and Kato proved that this subgroup is just the image of $H^1_f(K_{\mathfrak{l}}, V_pA_f)$ in $H^1(K_{\mathfrak{l}}, W)$, where $H^1_f(K_{\mathfrak{l}}, V_pA_f)$ is defined just as $H^1_f(K_{\mathfrak{l}}, V)$. (Note that $H^1(K_{\mathfrak{l}}, V_pA_f)$, $H^1_f(K_{\mathfrak{l}}, V_pA_f)$, and $H^1_f(K_{\mathfrak{l}}, W)$ are all 0 if $\mathfrak{l} \nmid p$.) In particular, if S is a finite set of primes of K containing all those that divide pN, then $\mathrm{Sel}_{p^\infty}(A_f/K) \subset H^1(G_{K,S}, W)$ consists of those classes with restriction to $H^1(K_{\mathfrak{l}}, W)$ belonging to $H^1_f(K_{\mathfrak{l}}, W)$ for all $\mathfrak{l} \in S$. As the image of $H^1(G_{K,S}, V_pA_f)$ in $H^1(G_{K,S}, W)$ is the maximal divisible subgroup (and has finite index), it follows that the maximal divisible subgroup of $\mathrm{Sel}_{p^\infty}(A_f/K)$ is the image in $H^1(K, W)$ of the characteristic zero Bloch–Kato Selmer group $H^1_f(K, V_pA_f)$.

The projective limit of the Kummer maps for the multiplication by p^n-maps yields an injection

$$A_f(K) \otimes \mathbb{Q}_p \hookrightarrow H^1_f(K, V_pA_f)$$

that is compatible with the fundamental exact sequence, so the cokernel has $\mathbb{Q}_p$-dimension equal to the corank of $\text{Ш}(A_f/K)[p^\infty]$.

The connection with $H^1_f(K, V)$ is just

$$H^1_f(K_{\mathfrak{l}}, V) = H^1_f(K_{\mathfrak{l}}, V_pA_f) \otimes_{M_{f,p}} L \quad \text{and} \quad H^1_f(K, V) = H^1_f(K, V_pA_f) \otimes_{M_{f,p}} L,$$

from which, together with the fact that $A_f(K) \otimes \mathbb{Q}_p$ is a free $M_{f,p}$-module, we deduce

LEMMA 2.2.2. *If* $\mathrm{rank}_{\mathbb{Z}} A_f(K) = [M_f : \mathbb{Q}]$ *and* $\#\text{Ш}(A_f/K)[p^\infty] < \infty$, *then* $\dim_L H^1_f(K, V) = 1$ *and the restriction map* $H^1_f(K, V) \overset{\mathrm{res}}{\to} H^1(K_{\mathfrak{l}}, V)$ *is an injection for each* $\mathfrak{l} \mid p$.

2.3. *More Galois cohomology.* Let S be any finite set of primes containing those dividing pN, and let

$$H^i(K_p, V) = \prod_{\mathfrak{l} | p} H^i(K_{\mathfrak{l}}, V)$$

and

$$H^1_{\mathrm{str}}(K, V) = \ker\{H^1(G_{K,S}, V) \overset{\mathrm{res}}{\to} H^1(K_p, V)\}.$$

Note that $H^1_{\mathrm{str}}(K, V)$ (often called the "strict" Selmer group of V) is independent of S.

LEMMA 2.3.1. $\dim_L \mathrm{im}\{H^1(G_{K,S}, V) \overset{\mathrm{res}}{\to} H^1(K_p, V)\} = 2.$

Proof. By Tate global duality, $H^1_{\mathrm{str}}(K, V)$ is dual to $H^2(G_{K,S}, V)$ (here we are using that $H^1(K_{\mathfrak{l}}, V) = 0$ if $\mathfrak{l} \nmid p$ and $H^2(K_{\mathfrak{l}}, V) = 0$ for all $\mathfrak{l}$), so

$$\begin{aligned} \dim_L H^1(G_{K,S}, V) - {} & \dim_L H^1_{\mathrm{str}}(K, V) \\ &= \dim_L H^1(G_{K,S}, V) - \dim_L H^2(G_{K,S}, V) \\ &= 2, \end{aligned}$$

the last equality following from $H^0(G_{K,S}, V) = 0$ and Tate's formula for the global Euler characteristic. □

Suppose that p splits in K:

$$p = \mathfrak{p}\bar{\mathfrak{p}}.$$

Let

$$H^1_{\mathfrak{p}}(K,V) = \ker\{H^1(G_{K,S},V) \stackrel{\text{res}}{\to} H^1(K_{\mathfrak{p}},V)\},$$

and let $H^1_{\bar{\mathfrak{p}}}(K,V)$ be defined similarly; these are independent of the finite set S.

Lemma 2.3.2. *If* $\dim_L \operatorname{im}\{H^1_f(K,V) \stackrel{\text{res}}{\to} H^1_f(K_p,V)\} = 1$, *then*

$$H^1_{\mathfrak{p}}(K,V) = H^1_{\text{str}}(K,V) = H^1_{\bar{\mathfrak{p}}}(K,V).$$

If also $\dim_L H^1_f(K,V) = 1$, *then* $H^1_{\mathfrak{p}}(K,V) = H^1_{\text{str}}(K,V) = H^1_{\bar{\mathfrak{p}}}(K,V) = 0$.

Proof. By Lemma 2.3.1, the image X of $H^1(G_{K,S},V)$ in $H^1(K_p,V)$ is two-dimensional over L. Let $X_{\mathfrak{p}}$ and $X_{\bar{\mathfrak{p}}}$ be the respective images of $H^1_{\mathfrak{p}}(K,V)$ and $H^1_{\bar{\mathfrak{p}}}(K,V)$ in $H^1(K_p,V)$; the action of the non-trivial automorphism c of K swaps $X_{\mathfrak{p}}$ and $X_{\bar{\mathfrak{p}}}$. Let X_f be the image of $H^1_f(K,V)$ in $H^1(K_p,V)$; this is stable under c and one-dimensional by hypothesis.

Suppose $X_{\mathfrak{p}} \neq 0$. Then $X_{\bar{\mathfrak{p}}} \neq 0$ and $X = X_{\mathfrak{p}} \oplus X_{\bar{\mathfrak{p}}}$, from which it follows that $X = X^+ \oplus X^-$ with $X^{\pm} = \{x \pm c(x) \ : \ x \in X_{\mathfrak{p}}\}$, and $X^{\pm}$ is one-dimensional. Then X_f equals X^+ or X^-. But it then follows that $X_{\mathfrak{p}} \subset H^1_f(K_{\bar{\mathfrak{p}}},V)$ and $X_{\bar{\mathfrak{p}}} \subset H^1_f(K_{\mathfrak{p}},V)$ and hence that X_f is two-dimensional, a contradiction. Therefore, $X_{\mathfrak{p}} = 0 = X_{\bar{\mathfrak{p}}}$.

If also $\dim_L H^1_f(K,V) = 1$, then $H^1_{\text{str}}(K,V) = 0$, whence the final conclusion. □

2.4. *The Heegner points* $P_K(f)$. We consider two cases:

I. Every prime $\ell \mid N$ either splits or ramifies in K.

II. $N = N^-N^+$ with N^- a nontrivial product of an even number of primes that are inert in K and N^+ is the product of primes that split in K (in particular, $(D,N) = 1$).

Suppose first that f and K are as in Case I. Let $\mathbb{T}$ be the Hecke algebra generated over $\mathbb{Z}$ by the usual Hecke operators T_ℓ, for primes $\ell \nmid N$, acting on the space of cuspforms $S_2(\Gamma_0(N))$. Let $\mathbb{T}_{M_f} = \mathbb{T} \otimes M_f$, and let $\varepsilon_f \in \mathbb{T}_{M_f}$ be the idempotent corresponding to the projection $\mathbb{T}_{M_f} \twoheadrightarrow M_f$ sending T_ℓ to the eigenvalue $a_\ell(f)$ of its action on the newform f. Let $\mathfrak{p}_f = \ker \varepsilon_f|_{\mathbb{T}}$.

Let $X = X_0(N)$ be the modular curve over $\mathbb{Q}$. Then f determines a differential $\omega_f \in \Omega^1(X) \otimes \mathbb{C} = \Omega^1(X(\mathbb{C}))$: the pullback of ω_f to the upper half-plane via the usual complex uniformization of $X(\mathbb{C})$ is $2\pi i f(\tau)d\tau$. The operator T_ℓ can also be viewed as acting via a correspondence on X such that $T_\ell \cdot \omega_f = \omega_{T_\ell \cdot f} = a_\ell(f) \cdot \omega_f$. The induced action of the T_ℓ's on the Jacobian $J(X)$ of X realizes $\mathbb{T}$ as a subring of $\operatorname{End}_{\mathbb{Q}}(J(X))$, and ω_f is a basis for the one-dimensional M_f space $\varepsilon_f(\Omega^1(J(X)) \otimes M_f)$ (where we have identified

$\Omega^1(X) = \Omega^1(J(X))$ in the usual way). The abelian variety A_f is just the quotient $A_f = J(X)/\mathfrak{p}_f J(X)$, and we let $\phi : J(X) \to A_f$ be the quotient map. We let $\omega \in \Omega^1(A_f) \otimes M_f$ be the 1-form such that $\phi^*\omega = \omega_f$.

Let $\mathcal{O}_K$ be the integer ring of K and $\mathfrak{n} \subset \mathcal{O}_K$ an ideal of norm N; this is possible by the hypothesis that each prime $\ell \mid N$ either splits or ramifies in K. The degree N isogeny $\mathbb{C}/\mathcal{O}_K \to \mathbb{C}/\mathfrak{n}^{-1}$ (the canonical projection) of CM elliptic curves is cyclic since N is square free (in particular, $\ell^2 \nmid N$ if ℓ ramifies in K) and defines a point $P \in X(H)$ on X over the Hilbert class field H of K. Let $D_K = \sum_{\sigma \in \mathrm{Gal}(H/K)} P^\sigma \in \mathrm{Div}(X)$. For $\ell \nmid N$, let $D_{K,\ell} = (T_\ell - 1 - \ell) \cdot D_K \in \mathrm{Div}^0(X)$ (the degree of the correspondence T_ℓ is $\ell + 1$). Let $Q_K(f) = \frac{1}{a_\ell(f) - 1 - \ell} \varepsilon_f \cdot [D_{K,\ell}] \in J(X)(K) \otimes M_f$. This is independent of ℓ since $\varepsilon_f \cdot T_\ell = a_\ell(f) \cdot \varepsilon_f$ in $\mathbb{T}_{M_f}$. Put

$$P_K(f) = \phi(Q_K(f)) \in A_f(K) \otimes M_f.$$

For the purposes of comparison with other constructions, we also consider $D_K^0 = D_K - \#\mathrm{Gal}(H/K) \cdot \infty \in \mathrm{Div}^0(X)$, where $\infty \in X(\mathbb{Q})$ is the usual cusp at infinity, and $Q_K^0 = [D_K^0] = \sum_{\sigma \in \mathrm{Gal}(H/K)} [P - \infty]^\sigma \in J(X)(K)$. As $T_\ell \cdot \infty = (1 + \ell) \cdot \infty$ (that is, the cusps are Eisenstein), $\varepsilon_f \cdot Q_K^0 = Q_K(f)$. Similarly, if $\xi \in \mathrm{Div}(X) \otimes \mathbb{Q}$ is the normalized, degree one Hodge divisor defined in [27, §§1.2.2, 3.1.3], we let $D_K^\xi = D_K - \#\mathrm{Gal}(H/K) \cdot \xi \in \mathrm{Div}^0(X) \otimes \mathbb{Q}$ and $Q_K^\xi = [D_K^\xi] = \sum_{\sigma \in \mathrm{Gal}(H/K)} [P - \xi]^\sigma \in J(X)(K) \otimes \mathbb{Q}$. It follows directly from the expression in [27, §3.1.3] for the Hodge divisor in terms of the canonical divisor that ξ is also Eisenstein, so $\varepsilon_f \cdot Q_K^\xi = Q_K(f)$.

In Case II we let X be the Shimura curve over $\mathbb{Q}$ associated with the indefinite quaternion algebra B of discriminant N^- and an Eichler order $\mathcal{O}_{B,N^+}$ of level N^+ in a maximal order $\mathcal{O}_B$ of B. Let $K \hookrightarrow B$ be an embedding such that $K \cap \mathcal{O}_{B,N^+} = \mathcal{O}_K$. Replacing f with its Jacquet-Langlands transfer f_B to the space of weight 2 cuspforms for the subgroup determined by O_{B,N^+} (and normalized as in [6, §2.8] to be defined over M_f) and $\mathfrak{n}$ by an integral ideal $\mathfrak{n}^+$ of K with norm N^+, there are constructions analogous to those yielding $Q_K(f)$ and $P_K(f)$ in Case I that in this case yield $Q_K(f) = Q_K(f_B) \in J(X)(K) \otimes M_f$ and $P_K(f) = P_K(f_B) \in A_f(K) \otimes M_f$. We can also define Q_K^ξ in this case, and, just as in Case I, $\varepsilon_f \cdot Q_K^\xi = Q_K(f)$.

2.5. *The Gross–Zagier theorem.* We recall a consequence of a special case of the general Gross–Zagier formula of Yuan, Zhang, and Zhang [27].

Consider the following hypotheses:

(sqf) $\quad N$ is squarefree,

(sgn) $\quad \epsilon(f, K) = -1.$

As the central character of π is trivial, the first of these hypotheses implies that for each $\ell \mid N$, π_ℓ is either σ_ℓ or $\sigma_\ell \otimes \xi_\ell$, where σ_ℓ is the special representation and ξ_ℓ is the unramified quadratic character of $\mathbb{Q}_\ell^\times$. The second implies that that the number of primes $\ell \mid N$ for which $\epsilon_\ell(\pi, K) = -1$ is even. In particular, if $(D, N) = 1$ then, since $\epsilon_\ell(\pi, K) = +1$ if ℓ splits in K and $\epsilon_\ell(\pi, K) = -1$ if $\ell \mid N$ is inert in K, the number of prime divisors of N that are inert in K is even.

Consider the additional hypothesis:

(ram) If $(D, N) \neq 1$, then $\pi_\ell \cong \sigma_\ell \otimes \xi_\ell$ for each $\ell \mid (D, N)$, and each $\ell \mid \frac{N}{(D,N)}$ splits in K.

If (sqf), (sgn), and (ram) hold, then f and K are as in either Case I or Case II of Section 2.4.

Proposition 2.5.1 ([27]). *Suppose* (sqf), (sgn), *and* (ram) *hold. If* $P_K(f) \neq 0$, *then* $\mathrm{ord}_{s=1} L(f, K, s) = 1$.

We explain how Proposition 2.5.1 follows from the main result of [27]. Let $\iota : M_f \hookrightarrow \mathbb{C}$ be the identity map. We first note that $P_K(f) = h_K \cdot P(f_1)^\iota$, where $h_K = \#\mathrm{Gal}(H/K)$ and $P(f_1) \in A_f(K)$ is associated to the homomorphism $f_1 = \phi : J(X) \to A_f$ and the point $P \in X(H)$ as in [27, §3.2.5]. Let $\mathcal{L}$ be a symmetric, ample line bundle on A_f and $\lambda : A_f \to A_f^\vee$ the associated polarization. Then $\lambda(P_K(f)) \in A_f^\vee(K) \otimes M_f$ is just the point $h_K \cdot P(f_2)^\iota$ as in loc. cit. with $f_2 = \lambda \circ \phi : J(X) \to A_f^\vee$. Let $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ be the Néron–Tate height-pairing on $A_f(K)$ associated with the line bundle $\mathcal{L}$, and let $\langle \cdot, \cdot \rangle_{NT}$ be the canonical height pairing on $A_f(\overline{K}) \otimes \mathbb{R} \times A_f^\vee(\overline{K}) \otimes \mathbb{R}$. Then

$$\begin{aligned}\langle P_K(f), P_K(f) \rangle_{\mathcal{L}} = \langle P_K(f), \lambda(P_K(f)) \rangle_{NT} &= h_K^2 \langle P(f_1)^\iota, P(f_2)^\iota \rangle_{NT} \\ &= \frac{h_K^2 \zeta(2) L'(f, K, 1) \cdot (f_1 \circ f_2^\vee)^\iota}{4L(\chi_D, 1)^2 L(\mathrm{Sym}^2 f, 2) \mathrm{vol}(X)},\end{aligned}$$

where $f_1 \circ f_2^\vee \in \mathrm{End}^0_{\mathbb{Q}}(A_f) = M_f$. The last equality is just [27, Th. 3.13]. As $P_K(f) \neq 0$ if and only if $\langle P_K(f), P_K(f) \rangle_{\mathcal{L}} \neq 0$, the proposition follows.

If $P_K(f) = \sum P \otimes r_P$ with P running over a basis of $A_f(K) \otimes \mathbb{Q}$ and $r_P \in M_f$, then for any $\sigma : M_f \hookrightarrow \mathbb{R}$, $P_K(f^\sigma) = \sum P \otimes \sigma(r_P)$. So $P_K(f) \neq 0$ if and only if $P_K(f^\sigma) \neq 0$ for all σ. Appealing to the above proposition for all the Galois conjugates f^σ of f we deduce

Corollary 2.5.2. *If* $P_K(f) \neq 0$, *then* $\mathrm{ord}_{s=1} L(A_f/K, s) = [M_f : \mathbb{Q}]$.

2.6. *A p-adic L-function and a formal log of $P_K(f)$.* As described in the introduction, our proof that $P_f(K) \neq 0$, and hence of Theorem B, hinges on an identity expressing the value of a certain p-adic L-function as a non-zero multiple of the square of a formal logarithm of $P_K(f)$. We now recall this L-function and identity.

Let K_∞/K be the anticyclotomic $\mathbb{Z}_p$-extension of K; so $\Gamma = \mathrm{Gal}(K_\infty/K) \cong \mathbb{Z}_p$ and conjugation by c sends $\gamma \in \Gamma$ to γ^{-1}. Let $\Psi : G_K \twoheadrightarrow \Gamma$ be the canonical projection. We continue to assume that $p = \mathfrak{p}\bar{\mathfrak{p}}$ splits in K.

We also assume that λ is the prime of M_f determined by the fixed embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{K}_\mathfrak{p}$. And for the purposes of p-adic interpolation and comparisons with complex values, we fix an embedding $\overline{K}_\mathfrak{p} \hookrightarrow \mathbb{C}$ such that induced complex embedding of $\overline{\mathbb{Q}}$ is just the fixed one.

Given a continuous character $\psi : \Gamma \to \mathbb{Q}_p^\times$ we consider ψ to be a Galois character via composition with the projection Ψ. We say that such a ψ is Hodge–Tate if ψ is Hodge–Tate as a representation of both $G_{K_\mathfrak{p}}$ and $G_{K_{\bar{\mathfrak{p}}}}$. As ψ is anticyclotomic (that is, $\psi(c^{-1}gc) = \psi(g)^{-1}$), ψ is Hodge–Tate if and only if it is Hodge–Tate as a representation of one of $G_{K_\mathfrak{p}}$ and $G_{K_{\bar{\mathfrak{p}}}}$; the respective Hodge–Tate weights must be $-n$ and n for some integer n, and in this case we say that ψ is Hodge–Tate of weight $(-n, n)$. The Galois character ψ is the p-adic avatar of a unitary algebraic Hecke character ψ^{alg} of K with infinity type $z^n\bar{z}^{-n}$. The character $\psi^{\mathrm{alg}} : \mathbb{A}_K^\times \to \mathbb{C}^\times$ is given by $\psi^{\mathrm{alg}}((x_v)) = x_\mathfrak{p}^{-n} x_{\bar{\mathfrak{p}}}^n x_\infty^n \bar{x}_\infty^{-n} \cdot \psi \circ \mathrm{rec}_K((x_v))$, where $\mathrm{rec}_K : K^\times\backslash\mathbb{A}_K^\times \to G_K^{ab}$ is the reciprocity map of class field theory, which we normalize so that uniformizers correspond to geometric Frobenius elements. In particular, there is an equality of L-functions $L(\psi^{\mathrm{alg}}, s) = L(\psi, s)$. The characters ψ and ψ^{alg} are unramified at all places not dividing p. Let $\Sigma_\mathfrak{p}^c$ be the set of crystalline characters $\psi : \Gamma \to \mathbb{Q}_p^\times$ of weight $(-n, n)$ with $n > 0$ and $n \equiv 0 \bmod p - 1$; the crystalline condition is equivalent to ψ^{alg} being unramified at $\mathfrak{p}$ and $\bar{\mathfrak{p}}$.

Suppose in addition to (sqf), (sgn), and (ram) that

(flt) $$p \nmid N,$$

(odd) $$D \text{ is odd},$$

(L-lrg) L contains a large enough[3] finite extension of $\mathbb{Q}_p$.

Let $S = \{\ell \mid pND\}$. Let $\mathcal{O}$ be the ring of integers of L, and let $\mathcal{O}^{\mathrm{ur}}$ be the ring of integers of the completion of the maximal unramified extension L^{ur} of L. There is an anticyclotomic p-adic L-function $L_\mathfrak{p}^S(f) \in \mathcal{O}^{\mathrm{ur}}[\![\Gamma]\!]$ such that for $\psi \in \Sigma_\mathfrak{p}^c$,

$$\psi(L_\mathfrak{p}^S(f)) = C(f,K)w(f,\psi)e_\infty(f,\psi)e_p(f,\psi)\Omega_p(\psi)\frac{L^S(f,\psi^{\mathrm{alg}},1)}{\Omega(\psi^{\mathrm{alg}})},$$

[3]It is enough that L contain the image of the Hilbert class field of K, though this is not important here.

where

- $L^S(f, \psi^{\mathrm{alg}}, s) = L^S(V^\vee \otimes \psi, s)$ is the L-function with the Euler factors at the primes of K dividing the primes in S omitted;
- $\Omega(\psi^{\mathrm{alg}}) = (2\pi i)^{-2n-1}(2\pi)^{1-2n}\Omega^{4n}D^{n-1/2}$, with Ω a period of an elliptic curve E_0 with CM by the ring of integers of K;
- $\Omega_p(\psi) = \Omega_p^{4n}$ with Ω_p a p-adic period for the elliptic curve E_0;
- $e_\infty(f, \psi) = 4(2\pi)^{-2n-1}\Gamma(n)\Gamma(n+1)$;
- $e_p(f, \psi) = \frac{(1-a_p(f)\psi^{\mathrm{alg}}(\bar{\mathfrak{p}})p^{-1}+\psi^{\mathrm{alg}}(\bar{\mathfrak{p}})^2p^{-1})}{(1-a_p(f)\psi^{\mathrm{alg}}(\mathfrak{p})p^{-1}+\psi^{\mathrm{alg}}(\mathfrak{p})^2p^{-1})}$, where $a_p(f)$ is the eigenvalue for the action of the Hecke operator T_p on f;
- $w(f, \psi) = \psi(W)$ for a unit $W \in \mathcal{O}^{\mathrm{ur}}[\![\Gamma]\!]^\times$ and $w(f, 1) = w_K$, the number of roots of unity in K;
- $C(f, K)$ is some non-zero constant depending on f and K.

For any ψ, we write $L_{\mathfrak{p}}^S(f, \psi)$ for $\psi(L_{\mathfrak{p}}^S(f))$.

If there exists a prime $\ell \mid N$ that is inert or ramified in K, then this p-adic L-function is constructed in [7] and in [26]. It has been constructed more generally in [2], [6], and [5]. It can be related to a specialization of a three-variable L-function constructed by Hida [12]. (This is done in [26].) To be precise, the functions constructed in [2] and [6] are a priori only continuous on Γ. That they belong to the Iwasawa algebra (which follows from the constructions in [7] and [26]) requires additional argument, essentially extending the formulas to characters ramified at primes above p as is done in [5].

For $\psi \in \Sigma_{\mathfrak{p}}^c$, let χ be the Hecke character of K such that $\chi^{-1} = \psi^{\mathrm{alg}}|\mathrm{Nm}(\cdot)|_{\mathbb{Q}}$, where Nm is the norm map from $\mathbb{A}_K$ to $\mathbb{A}_{\mathbb{Q}}$ and $|\cdot|_{\mathbb{Q}}$ is the usual absolute value on $\mathbb{A}_{\mathbb{Q}}$. Recall that under the assumptions (sqf), (sgn), and (ram), f and K are as in Cases I or II of Section 2.4. Then $L_{\mathfrak{p}}^S(f)$ is the imprimitive variant of the p-adic L-function denoted $L_p(f, \chi)$ in [2] (in Case I) and in [6] (in Case II). By "imprimitive variant" we mean that the Euler factors at the primes in S not dividing p have been removed. The set of such χ for $\psi \in \Sigma_{\mathfrak{p}}^c$ is denoted $\Sigma_{cc}^{(2)}(\mathfrak{n})$ and $\Sigma_{cc}^{(2)}(\mathfrak{n}^+)$, respectively, in [2] and [6], and the interpolated values are given in terms of the values $L(f, \chi^{-1}, 0)$, which is just $L(f, \psi^{\mathrm{alg}}, 1)$. Also, still in the notation of [2] and [6], $e_\infty(f, \psi)w_K(2\pi i)^{1+2n} = C(f, \chi, 1)$, and $w(f, \psi) = w_K w(f, \chi)^{-1}$. The constant we have denoted $C(f, K)$ is denoted $\alpha(f, f_{\mathrm{GL}_2})^{-1}$ in [6] (wherein f denotes a form on an indefinite quaternion algebra and f_{GL_2} is a suitably normalized Jacquet-Langlands lift of f to GL_2; the quaternion algebra depends on K).

Among the important results in [1] and [6] is the following expression for the value of $L_{\mathfrak{p}}^S(f, \psi)$ at the trivial character $\psi = 1$, which we call the BDP point, in terms of a formal log of the Heegner point $P_K(f)$. Note that the BDP point is not in $\Sigma_{\mathfrak{p}}^c$.

Proposition 2.6.1 ([1], [6]). *Suppose* (sqf), (sgn), (ram), (flt), *and* (odd) *hold. Then*

$$L_{\mathfrak{p}}^S(f,1) \doteq (\log_\omega P_K(f))^2.$$

Recall that "$\doteq$" means equality up to a non-zero constant (which here also depends on the eigenvalues at the primes in S), and $\omega \in \Omega^1(A_f)\otimes M_f$ is a 1-form such that $\phi^*\omega = \omega_f$ (so the action of M_f on ω through its action on $\Omega^1(A_f)$ agrees with its scalar action), and $\log_\omega : A_f(K_{\mathfrak{p}})\otimes L \to \overline{K}_{\mathfrak{p}}$ is the formal logarithm determined by ω.

In Case I, Proposition 2.6.1 is just [1, Th. 3.12]. The formula in loc. cit. has $\log_\omega P_K(f)$ replaced with $\log_\omega P_f$, where $P_f = \phi(Q_K^0)$. But $\log_\omega P_K = \log_{\omega_f} Q_K^0$, and, since $\varepsilon_f\cdot\omega_f = \omega_f$, $\log_{\omega_f} Q_K^0 = \log_{\omega_f} \varepsilon_f\cdot Q_K^0 = \log_{\omega_f} Q_K(f) = \log_\omega P_K(f)$, whence the formula in this case. In Case II, the proposition similarly follows from [6, Prop. 8.13].

To prove that $P_K(f)\neq 0$, and hence that $A_f(K)$ has positive rank, it suffices to show that $\log_\omega P_K(f)\neq 0$ and, therefore, to show that $L_{\mathfrak{p}}^S(f,1)\neq 0$.

Corollary 2.6.2. $P_K(f)\neq 0$ *if and only if* $L_{\mathfrak{p}}^S(f,1)\neq 0$.

In the following section we explain some other consequences of $L_{\mathfrak{p}}^S(f,1)=0$.

2.7. *Some Iwasawa theory for f and K.* We continue to assume $p=\mathfrak{p}\bar{\mathfrak{p}}$ splits in K.

Let $\mathcal{O}$ be the ring of integers of L, and let $T\subset V$ be a $G_{\mathbb{Q}}$-stable $\mathcal{O}$-lattice. Let $\Lambda = \mathcal{O}[\![\Gamma]\!]$, and let $\Lambda^{\mathrm{ur}} = \mathcal{O}^{\mathrm{ur}}[\![\Gamma]\!]$, where $\mathcal{O}^{\mathrm{ur}}$ is the ring of integers of the completion of the maximal unramified extension of L. We view the projection $\Psi : G_K \twoheadrightarrow \Gamma$ as a continuous $\Lambda^\times$-valued character. Let $\Lambda^* = \mathrm{Hom}_{cts}(\Lambda, \mathbb{Q}_p/\mathbb{Z}_p)$ be the Pontryagin dual of Λ, and let

$$M = T\otimes_{\mathcal{O}} \Lambda^*(\Psi^{-1}),$$

that is, the discrete Λ-module $T\otimes_{\mathcal{O}}\Lambda^*$ with continuous G_K action $\rho_{f,\lambda}\otimes\Psi^{-1}$. Let $S=\{\ell \mid pND\}$. Let

$$\mathrm{Sel}_\infty(f,K,S) = \ker\left\{H^1(G_{K,S},M) \overset{\mathrm{res}}{\to} H^1(I_{\mathfrak{p}},M)\right\}.$$

This is a discrete Λ-module, and its Pontryagin dual

$$X_\infty(f,K,S) = \mathrm{Hom}_\Lambda(\mathrm{Sel}_\infty(f,K,S),\Lambda^*)$$

is a finitely generated Λ-module. Let $\mathrm{Ch}_\Lambda(f,K,S)$ be its characteristic ideal over Λ; this is non-zero if and only if $X_\infty(f,K,S)$ is a torsion Λ-module.

The Selmer group $\mathrm{Sel}_\infty(f,K,S)$ is essentially an imprimitive version of one of Greenberg's Selmer groups for the "big" Galois module M, as we now explain. Let $\psi\in\Sigma_{\mathfrak{p}}^c$ with Hodge–Tate weights $(-n,n)$; recall $n>0$. Then the induced representation $\mathrm{Ind}_{G_K}^{G_{\mathbb{Q}}}(V \otimes \psi^{-1})$ satisfies the Panchishkin condition as in [10, §3]: its restriction to $G_{\mathbb{Q}_p}$ is just $(V \otimes \psi^{-1}) \oplus (V \otimes \psi^{-c})$, where the first summand is identified with $(V \otimes \psi^{-1})|_{G_{K_{\mathfrak{p}}}}$ and has non-negative Hodge–Tate weights n and $n-1$ while the second summand is identified with $(V \otimes \psi^{-1})|_{G_{K_{\bar{\mathfrak{p}}}}}$ and has negative Hodge–Tate weights $-n$ and $-n-1$; the subspace with negative[4] Hodge–Tate weights has dimension equal to the dimension of the $+1$-eigenspace for complex conjugation on the induced representation. Furthermore, if $\gamma \in \Gamma$ is a topological generator, then $M[\gamma - \psi(\gamma)] \cong (T \otimes \psi^{-1}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p/\mathbb{Z}_p$. That is, M is the discrete Galois module associated with the deformation $T \otimes \Lambda(\Psi^{-1})$ of T over Λ, which contains a Zariski-dense set of specializations that satisfy the Panchishkin condition (namely, the specializations under the maps $\gamma \mapsto \psi(\gamma)$ for $\psi \in \Sigma_{\mathfrak{p}}^c$). This is a simple generalization to G_K-representations of the situation considered in [10, §4], and the Selmer group $\mathrm{Sel}_\infty(f, K, S)$ is the corresponding generalization of the Selmer groups defined in *loc. cit.* Analogously to [10, Conj. 4.1], assuming also

(irr$_K$) $\qquad\qquad \bar{\rho}_{f,\lambda}|_{G_K}$ is irreducible,

one then conjectures[5]

Conjecture 2.7.1. *$\mathrm{Ch}_\Lambda(f, K, S)$ and $L_{\mathfrak{p}}^S(f)$ generate the same ideal of $\Lambda^{\mathrm{ur}} \otimes_{\mathcal{O}} L$.*

This is essentially the Iwasawa–Greenberg Main Conjecture for M.

Significant progress toward this conjecture has been made by X. Wan [26], following the methods of [24]. Suppose in addition to (sqf), (flt), (L-lrg), and (irr$_K$) that (sgn) and (odd) hold and that

(big) $\qquad\qquad p \geq 5,$

(ord) $\qquad\qquad f$ is ordinary with respect to λ,

(spl) $\qquad\qquad$ both 2 and p split in K,

(res) $\bar{\rho}_{f,\lambda}$ is ramified at some odd prime $\ell | N$ that is inert or ramified in K.

Then it is proved in [26] that

Proposition 2.7.2 ([26]). *Under the above assumptions, the ideal of $\Lambda^{\mathrm{ur}} \otimes_{\mathcal{O}} L$ generated by $\mathrm{Ch}_\Lambda(f, K, S)$ is contained in the ideal generated by $L_{\mathfrak{p}}^S(f)$.*

[4] Our conventions for Hodge–Tate weights are the negative of those in [10].

[5] The order of the Selmer group for $V \otimes \psi^{-1}$ is expected to be related to the L-value $L(V^\vee \otimes \psi, 1)$. This dictates which p-adic L-function should be identified with the characteristic ideal of $\mathrm{Sel}_\infty(f, K, S)$, namely, $L_{\mathfrak{p}}^S(f)$.

This proposition is not explicitly given in [26], however it follows easily from [26, Th. 1.2], as we explain.

Let π be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ such that $L(\pi, s-1/2) = L(f,s)$, and let f_0 be the ordinary stabilization of the newform f. We fix a character $\psi_0 \in \Sigma_{\mathfrak{p}}^c$ with Hodge–Tate weights $(-n_0, n_0)$ with $n_0 > 6$. Let ψ^{alg} be the associated algebraic Hecke character, and let $\xi = \psi^{\mathrm{alg}}\omega$, where ω is the finite order Hecke character of K associated with the Teichmüller character (the lift of the mod p reduction of the p-adic cyclotomic character). Then π, f_0, and ξ satisfy the hypotheses of [26, Th. 1.2]. (In the notation of *loc. cit.*, $\Gamma_{\mathcal{K}}$ is the Galois group of the composite of all $\mathbb{Z}_p$-extensions of $\mathcal{K} = K$, so $\Gamma_{\mathcal{K}} \cong \mathbb{Z}_p^2$ and Γ is a natural quotient of $\Gamma_{\mathcal{K}}$.) It remains to explain how the conclusion of that theorem (or really its proof) implies the above proposition. The p-adic L-function $\mathcal{L}_{f_0,\mathcal{K},\xi}^{\Sigma} \in \mathcal{O}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]$ with $\Sigma = S$ from [26, §7.5] is a two-variable extension of the p-adic L-function considered herein: under the composition

$$\mathcal{O}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!] \overset{\Gamma_{\mathcal{K}}\twoheadrightarrow\Gamma}{\longrightarrow} \Lambda^{\mathrm{ur}} \overset{\gamma\mapsto\psi_0(\gamma)^{-1}\gamma}{\longrightarrow} \Lambda^{\mathrm{ur}},$$

$\mathcal{L}_{f_0,\mathcal{K},\xi}^{\Sigma}$ maps to $L_{\mathfrak{p}}^S(f)$. Similarly, under the base change from $\mathcal{O}[\![\Gamma_{\mathcal{K}}]\!]$ to Λ given by this map (that is, tensoring with Λ over $\mathcal{O}[\![\Gamma_{\mathcal{K}}]\!]$), the Selmer group denoted $\mathrm{Sel}_{f_0,\mathcal{K},\xi}^{\Sigma}$ in [26, §2.2] becomes $\mathrm{Sel}_\infty(f,K,S)$ and $X_{f_0,\mathcal{K},\xi}^{\Sigma}$ becomes $X_\infty(f,K,S)$. The proof of [26, Th. 1.2] involves showing that for $\Sigma = S$, the ideal of $\mathcal{O}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!] \otimes_{\mathcal{O}} L$ generated by the characteristic ideal $\mathrm{char}_{\mathcal{O}[\![\Gamma_{\mathcal{K}}]\!]} X_{f_0,\mathcal{K},\xi}^{\Sigma}$ is contained in the ideal generated by $\mathcal{L}_{f_0,\mathcal{K},\xi}^{\Sigma}$. The corresponding inclusion of the ideal of $\Lambda^{\mathrm{ur}} \otimes_{\mathcal{O}} L$ generated by $\mathrm{Ch}_\Lambda(f,K,S)$ in the ideal generated by $L_{\mathfrak{p}}^S(f)$ then follows easily, using that the rings $\mathcal{O}^{\mathrm{ur}}[\![\Gamma_{\mathcal{K}}]\!]$ and Λ^{ur} are unique factorization ideals and the various characteristic ideals are principal (cf. [24, Cor. 3.8]).

We now explain a simple consequence of $L_{\mathfrak{p}}^S(f,1) = 0$. Let $W = M[\gamma-1] = T \otimes_{\mathbb{Z}_p} \mathbb{Q}_p/\mathbb{Z}_p$, and let

$$\mathrm{Sel}_{\mathfrak{p}}(f,K,S) = \ker\left\{H^1(G_{K,S},W) \overset{\mathrm{res}}{\to} H^1(I_{\mathfrak{p}},W)\right\}.$$

Let also

$$X_{\mathfrak{p}}(f,K,S) = \mathrm{Hom}_{\mathbb{Z}_p}(\mathrm{Sel}_{\mathfrak{p}}(f,K,S), \mathbb{Q}_p/\mathbb{Z}_p).$$

Assuming (irr_K),

$$H^1(G_{K,S},W) = H^1(G_{K,S},M[\gamma-1]) = H^1(G_{K,S},M)[\gamma-1].$$

It then follows from the exactness of the bottom row of the commutative diagram

$$\begin{array}{ccccc}
 & & H^1(G_{K,S},W) & = & H^1(G_{K,S},M)[\gamma-1] \\
 & & \downarrow{\scriptstyle \mathrm{res}} & & \downarrow{\scriptstyle \mathrm{res}} \\
(M^{I_{\mathfrak{p}}}/(\gamma-1)M^{I_{\mathfrak{p}}})^{G_{K_{\mathfrak{p}}}} & \longrightarrow & H^1(I_{\mathfrak{p}},W)^{G_{K_{\mathfrak{p}}}} & \longrightarrow & H^1(I_{\mathfrak{p}},M)[\gamma-1]^{G_{K_{\mathfrak{p}}}}
\end{array}$$

(the bottom row comes from the long exact cohomology sequence associated with the short exact sequence $0 \to W \to M \stackrel{\times(\gamma-1)}{\to} M \to 0$) that $\mathrm{Sel}_{\mathfrak{p}}(f,K,S)$ is contained in $\mathrm{Sel}_{\infty}(f,K,S)[\gamma-1]$ with finite index, and therefore

$$(2.7.1)\quad \#X_{\mathfrak{p}}(f,K,S)=\infty \iff \#(X_{\infty}(f,K,S)/(\gamma-1)X_{\infty}(f,K,S))=\infty.$$

Suppose $L_{\mathfrak{p}}^{S}(f,1)=0$. This means $L_{\mathfrak{p}}^{S}(f)\in(\gamma-1)$, so if the assumptions of Proposition 2.7.2 also hold, then $\mathrm{Ch}_{\Lambda}(f,K,S)\subset(\gamma-1)$. By basic properties of characteristic ideals, this last inclusion implies

$$\#(X_{\infty}(f,K,S)/(\gamma-1)X_{\infty}(f,K,S))=\infty.$$

Combining this with (2.7.1) we conclude

$$(2.7.2)\qquad L_{\mathfrak{p}}^{S}(f,1)=0 \implies \#X_{\mathfrak{p}}(f,K,S)=\infty.$$

As $H^1(K_{\mathfrak{p}},V)\hookrightarrow H^1(I_{\mathfrak{p}},V)$, we have

$$H^1_{\mathfrak{p}}(K,V)=\ker\{H^1(G_{K,S},V)\to H^1(I_{\mathfrak{p}},V)\},$$

from which it follows easily that the image of $H^1_{\mathfrak{p}}(K,V)$ in $H^1(G_{K,S},W)$ is the maximal divisible subgroup of $\mathrm{Sel}_{\mathfrak{p}}(f,K,S)$. Combining this observation with (2.7.1) and (2.7.2), we conclude that

Proposition 2.7.3. *If the hypotheses of Proposition 2.7.2 hold, then*

$$L_{\mathfrak{p}}^{S}(f,1)=0 \implies H^1_{\mathfrak{p}}(K,V)\neq 0.$$

2.8. *An observation about hypothesis* (irr_K). We include a simple lemma on the irreducibility of $\bar\rho_{f,\lambda}|_{G_K}$. Consider the hypothesis

(irr) $\qquad \bar\rho_{f,\lambda}$ is an irreducible $G_{\mathbb{Q}}$-representation.

Lemma 2.8.1. *If* (irr) *and* (res) *hold, then so does* (irr_K).

Proof. Suppose (irr) and (res) hold. Let $\ell\mid N$ be a prime at which $\bar\rho_{f,\lambda}$ is ramified. As $\ell\,||\,N$, the action of I_ℓ on V is unipotent and factors through tame inertia. In particular, if τ_ℓ is a topological generator of tame inertia at q, then $\rho_{f,\lambda}(\tau_\ell)$ is unipotent,[6] hence so, too, is $\bar\rho_{f,\lambda}(\tau_\ell)$; the latter is a unipotent element of order a power of p. As $\tau_\ell^2\in G_K$, it follows that the image of $\bar\rho_{f,\lambda}|_{G_K}$ contains a unipotent element of order a power of p. Let k be the residue field of L. Suppose now that $\bar\rho_{f,\lambda}$ is reducible over $\bar k$. Then the image of $\bar\rho_{f,\lambda}|_{G_K}$ is contained in either a torus (split or non-split) or a Borel of $\mathrm{GL}_2(k)$. The first possibility is ruled out as the image contains a unipotent element of order a power of p; the image of $\bar\rho_{f,\lambda}|_{G_K}$ is therefore contained in a Borel. But as the image of G_K is normalized by the image of $G_{\mathbb{Q}}$, it follows easily that the image of $G_{\mathbb{Q}}$ is also contained in the Borel, contradicting (irr). ∎

[6]This follows from the local-global compatibility satisfied by $\rho_{f,\lambda}|_{G_{\mathbb{Q}_\ell}}$.

2.9. *Finishing the proof of Theorem B*. Let f, p, and K be as in Theorem B. We begin by noting that (sqf), (odd), and (big) hold by hypothesis. Hypothesis (a) of Theorem B ensures that (flt) and (ord) hold. Hypothesis (b) is just (irr) and (res) and so (irr$_K$) holds by Lemma 2.8.1, and hypothesis (c) is just (spl). Hypothesis (e) then implies, by Lemma 2.2.1, that (sgn) holds. So Propositions 2.7.2 and 2.7.3 apply that if $H^1_{\mathfrak{p}}(K,V)=0$, then $L^S_{\mathfrak{p}}(f,1)\neq 0$. Hypothesis (e) also implies, by Lemma 2.3.2, that $H^1_{\mathfrak{p}}(K,V)=0$. Since hypothesis (d) is just (ram), it follows from Corollary 2.6.2 that $P_K(f)\neq 0$.

The rank of $A_f(K)$ is equal to $[M_f:\mathbb{Q}]$ times the $M_{f,p}$-rank of the free $M_{f,p}$-space $A_f(K)\otimes\mathbb{Q}_p$ which, by the injection $A_f(K)\otimes\mathbb{Q}_p\hookrightarrow H^1_f(K,V_pA_f)$, is at most the L-dimension of $H^1(K,V_pA_f)\otimes_{M_{f,p}}L=H^1_f(K,V)$. The latter has L-dimension 1 by hypothesis (e). Since $P_K(f)\neq 0$, so $A_f(K)\otimes\mathbb{Q}\neq 0$, it follows that $\mathrm{rank}_{\mathbb{Z}}A_f(K)=[M_f:\mathbb{Q}]$. It also follows, by Proposition 2.5.1 and Corollary 2.5.2, that $\mathrm{ord}_{s=1}L(f,K,s)=1$ and $\mathrm{ord}_{s=1}L(A_f/K,s)=[M_f:\mathbb{Q}]$.

To conclude that $Ш(A_f/K)$ is finite, we first observe that it is enough to show that both $Ш(A_f)$ and $Ш(A_f^D)$ are finite. To show these are finite we begin by noting that since $L(f,K,s)=L(f,s)L(f\otimes\chi_D,s)$ has order 1 at $s=1$, one of $L(f,s)$ and $L(f\otimes\chi_D,s)$ has order 1 at $s=1$ and the other has order 0. The finiteness of the Tate–Shafarevich groups then just follows from the work of Gross, Zagier, and Kolyvagin, as cited in the introduction.

This completes the proof of Theorem B. □

Remark 2.9.1. We indicate how the various hypotheses of Theorem B intervene in its proof and make some additional remarks on the theorem and its proof.

(i) The requirement that N be squarefree is made in [26] and in [6]. (That N be squarefree at those primes dividing (D,N) is also required in [2] and [1].)

(ii) The hypothesis that $p\geq 5$ comes from [26], where it is imposed for convenience.

(iii) The hypothesis that D be odd is made in [2] and [6] as well as [26] (in which 2 is also required to split in K) and stems from some gaps in our knowledge of the theta correspondence for local fields of residue characteristic 2.

(iv) The assumption that $p\nmid N$ intervenes most crucially in [2], [1], and [6]. We have also used it to simplify our use of the parity conjecture [18] for Selmer groups of modular forms (to verify (sgn)).

(v) The hypothesis that f is ordinary at some $\lambda\mid p$ is only needed to use the results of [26] and, again, in our appeal to [18]. In particular, f being ordinary is not crucial for the methods employed herein: if a version of

Proposition 2.7.2 were available[7] for forms with finite non-critical slope, for example, then all the results of this paper would hold in that case, provided the corresponding Selmer parity result was also known. (This is known for elliptic curves with supersingular reduction.)

(vi) The hypothesis that $\bar{\rho}_{f,\lambda}$ is irreducible is required in [26], as is the hypothesis that $\bar{\rho}_{f,\lambda}$ is ramified at an odd prime $\ell \neq p$ that is either inert or ramified in K. The latter ensures, among other things, that π has a transfer to a definite unitary group $U(2)$ (that is ramified at ℓ) defined using K; the primary results in [26] relate the p-adic L-function $L_{\mathfrak{p}}^S(f)$ to the index of an Eisenstein ideal on $U(3,1)$ coming from an Eisenstein series induced from this cuspform on $U(2)$.

(vii) The requirement that p split in K is needed to use the results in [2], [6], and [26]. It also comes into the Galois arguments, especially the proof of Lemma 2.3.2. It is likely one of the most difficult hypotheses to relax in the methods employed in this paper.

(viii) The hypotheses in (d) for when $(D,N) \neq 1$ are needed to appeal to the results of [2], which requires that $\epsilon_\ell(f,K) = +1$ for all primes $\ell \mid (N,D)$.

(ix) The hypothesis that $H^1_f(K,V)$ is one-dimensional is used to know beforehand that the root number $\epsilon(f,K)$ is -1; that is, (sgn) holds. This is required for the results in [27], [2], [1], and [6].

(x) The injectivity of the restriction map $H^1_f(K,V) \stackrel{\text{res}}{\to} \prod_{\mathfrak{l}|p} H^1(K_{\mathfrak{l}},V)$ is needed to ensure that $H^1_{\mathfrak{p}}(K,V) = 0$. Conjecturally, it should be enough that $H^1_f(K,V)$ is one-dimensional, and then the injectivity would follow from the conclusion that $A_f(K) \neq 0$. In [28] Wei Zhang obtains a version of Theorem B without requiring this injectivity, but at the expense of requiring certain Tamagawa numbers be indivisible by p.

(xi) Many, if not all, of the local hypotheses on π and K can likely be relaxed. For example, recent work of Y. Liu, S. Zhang, and W. Zhang essentially establishes the identity in Proposition 2.6.1 in the general Gross–Zagier set-up of [27] (including over a totally real field).

(xii) As recalled in the introduction, the analog of Theorem A for CM elliptic curves is explained in [22, Ths. 8.1,8.2] as a consequence of Rubin's proof of the main conjecture for CM curves, Perrin-Riou's p-adic Gross–Zagier formula, and Bertand's proof of the non-degeneracy of the relevant p-adic height pairing. It is also possible to give a proof for the CM case along the lines of the proof of Theorem B by using the Main Conjecture for CM

[7] Such a result has been announced in a preprint of Wan.

forms and the analog of Proposition 2.6.1 for the CM case, which is just [22, Th. 9.5] or [1, Th. 2].

(xiii) The methods employed to prove Theorem B in this paper can be adapted to provide an alternative proof of the base case of the induction argument in [28] that avoids appealing to [24] and so should also work for supersingular primes (see remark (v)). This is part of forthcoming work.

3. Theorem A follows from Theorem B

Let f and A_f be as in Theorem A. In particular, f is not a CM form.

Consider the set of primes $p \nmid N$ that are greater than 4 and unramified in M_f. Suppose that f is not ordinary for all $\lambda \mid p$ for some such p. Then the norm of $a_p(f)$, which has absolute value at most $(2p^{1/2})^{[M_f:\mathbb{Q}]}$ by the Ramanujan bounds, is an integer divisible by $p^{[M_f:\mathbb{Q}]}$, so $a_p(f) = 0$. But, as f is not a CM form, the set of primes with $a_p(f) = 0$ has density zero [23, §7.2, Cor. 2]. Thus the set of primes $p \nmid N$ such that f is ordinary with respect to some $\lambda \mid p$ of M_f has density 1.

If p is sufficiently large, then $\bar{\rho}_{f,\lambda}$ is irreducible for all $\lambda \mid p$ [20, Th. 2.1]. If for some $\ell \mid N$ there were arbitrarily large primes p and primes $\lambda \mid p$ of M_f such that $\bar{\rho}_{f,\lambda}$ were unramified at ℓ, then, by the finiteness of the number of newforms of weight 2 and level dividing N and by the main result of [21], there would be a newform g of weight 2 and level prime to ℓ such that f and g would be congruent modulo primes of arbitrarily large characteristic p, in the sense that their prime-to-Np coefficients would be congruent. It would then follow that the prime-to-Np coefficients of f and g would be the same and hence, by multiplicity one, that $f = g$, a contradiction. Thus, for sufficiently large p, $\bar{\rho}_{f,\lambda}$ is ramified at all primes that divide N.

By the preceding observations, we may fix a $p \geq 5$, $p \nmid N$, and a $\lambda \mid p$ such that A_f is ordinary with respect to λ and $\bar{\rho}_{f,\lambda}$ is irreducible and ramified at all primes that divide N. The hypotheses that $\mathrm{rank}_{\mathbb{Z}} A_f(\mathbb{Q}) = [M_f : \mathbb{Q}]$ and $Ш(A_f)[p^\infty]$ are finite imply, by the obvious analog of Lemma 2.2.2 with K replaced by $\mathbb{Q}$, that $\dim_L H^1_f(\mathbb{Q}, V) = 1$. It then follows from Nekovář's work on the parity conjecture for Selmer groups of modular forms [18, Th. 12.2.3] that $\epsilon(f) = -1$.

We choose an imaginary quadratic field $K/\mathbb{Q}$ of discriminant D such that

(i) 2 and p split in K (so D is odd and hypothesis (c) of Theorem B holds);

(ii) if for some odd prime ℓ the local representation π_ℓ is the twist of the special representation by the unique unramified quadratic character, then K is ramified at ℓ but all other prime divisors of N split in K;

(iii) if there is no π_ℓ as in (ii) but there are two odd primes ℓ_1 and ℓ_2 such that π_{ℓ_1} and π_{ℓ_2} are special, then ℓ_1 and ℓ_2 are inert in K and all other prime divisors of N split in K;

(iv) $L(f\otimes\chi_D,1)\neq 0$.

If (i) and (ii) hold, then $\epsilon(f,K)=-\epsilon_\ell(\pi,K)=-1$ (as $BC_{K_\mathfrak{l}}(\pi_\ell)$ is again the twist of the special representation by the unique unramified quadratic character and so has root number $+1$). If (i) and (iii) hold, then $\epsilon(f,K)=-\epsilon_{\ell_1}(\pi,K)\epsilon_{\ell_2}(\pi,K)=-1$ (as $BC_{K_{\ell_i}}(\pi_{\ell_i})$ is again the special representation and so has root number -1). In particular, for a K satisfying (i), (ii), and (iii) we always have $\epsilon(f)\epsilon(f\otimes\chi_D)=\epsilon(f,K)=-1$, so $\epsilon(f\otimes\chi_D)=-\epsilon(f)=+1$. It then follows from [9, Th. B] that K can also be chosen to satisfy (iv). Conditions (i), (ii) and (iii) imply that hypotheses (b), (c), and (d) of Theorem B hold. By the work of Gross, Zagier, and Kolyvagin cited in the introduction (alternatively, one could appeal to results of Kato [14]) condition (iv) implies that both $A_f^D(\mathbb{Q})$ and $Ш(A_f^D)$ are finite, from which it then follows that $\mathrm{rank}_\mathbb{Z} A_f(K)=[M_f:\mathbb{Q}]$ and $Ш(A_f/K)[p^\infty]$ is finite. By Lemma 2.2.2, hypothesis (e) of Theorem B then also holds. As hypothesis (a) holds by the choice of p and λ, we conclude from Theorem B that $\mathrm{ord}_{s=1}L(f,K,s)=1$. As $L(f,K,s)=L(f,s)L(f\otimes\chi_D,s)$ and $L(f\otimes\chi_D,1)\neq 0$, it follows that $\mathrm{ord}_{s=1}L(f,s)=1$.

This completes the proof of Theorem A. □

The deduction of Theorem C from Theorem B is similar: Hypothesis (c) of Theorem C is easily seen to imply that $H^1_f(\mathbb{Q},V)\xrightarrow{\sim}H^1_f(\mathbb{Q}_p,V)\cong\mathbb{Q}_p$. The representation $E[p]$ must be ramified at some odd prime ℓ of bad reduction for E (again by Ribet's level-lowering results). An appropriate imaginary quadratic field K that is either inert or ramified at ℓ is then chosen, depending on whether E has split or non-split reduction at ℓ.

4. **A remark on the $r=0$ case**

The arguments used to deduce Theorem A from Theorem B can be adapted to show that if $A_f(\mathbb{Q})$ and $Ш(A_f)$ are finite then $L(f,1)\neq 0$. This gives an alternate proof of a special case of the results in [24] and [25] cited in the introduction.

THEOREM E. *Suppose N is squarefree. If there is at least one odd prime ℓ such that π_ℓ is the twist of the special representation by the unique unramified quadratic character or at least two odd primes ℓ_1 and ℓ_2 such that π_{ℓ_1} and π_{ℓ_2} are special, then*

$$\#A_f(\mathbb{Q}),\ \#Ш(A_f)<\infty \implies \mathrm{ord}_{s=1}L(f,s)=0.$$

The argument is virtually identical to the proof that Theorem B implies Theorem A. The changes involved are that the hypotheses now imply that $\epsilon(f) = +1$, and K is chosen to satisfy (i), (ii), (iii), and

(iv)$'$ $\mathrm{ord}_{s=1} L(f \otimes \chi_D, s) = 1$.

This is possible as $\epsilon(f \otimes \chi_D)$ will equal -1. We then conclude from Theorem B, much as before, that $\mathrm{ord}_{s=1} L(f, K, s) = 1$, which implies by the choice of K that $L(f, 1) \neq 0$.

References

[1] M. Bertolini, H. Darmon, and K. Prasanna, p-adic Rankin L-series and rational points on CM elliptic curves, *Pacific J. Math.* **260** no. 2 (2012), 261–303. MR 3001796. Zbl 1326.11026. https://doi.org/10.2140/pjm.2012.260.261.

[2] M. Bertolini, H. Darmon, and K. Prasanna, Generalized Heegner cycles and p-adic Rankin L-series, *Duke Math. J.* **162** no. 6 (2013), 1033–1148, With an appendix by Brian Conrad. MR 3053566. Zbl 1302.11043. https://doi.org/10.1215/00127094-2142056.

[3] M. Bhargava and C. Skinner, A positive proportion of elliptic curves over $\mathbb{Q}$ have rank one, *J. Ramanujan Math. Soc.* **29** no. 2 (2014), 221–242. MR 3237733. Zbl 1315.11045.

[4] S. Bloch and K. Kato, L-functions and Tamagawa numbers of motives, in *The Grothendieck Festschrift*, *Vol. I*, *Progr. Math.* **86**, Birkhäuser Boston, Boston, MA, 1990, pp. 333–400. MR 1086888. Zbl 0768.14001. https://doi.org/10.1007/978-0-8176-4574-8_9.

[5] M. Brakočević, Anticyclotomic p-adic L-function of central critical Rankin-Selberg L-value, *Int. Math. Res. Not. IMRN* **21** (2011), 4967–5018. MR 2852303. Zbl 1267.11069. https://doi.org/10.1093/imrn/rnq275.

[6] E. H. Brooks, Shimura curves and special values of p-adic L-functions, *Int. Math. Res. Not. IMRN* no. 12 (2015), 4177–4241. MR 3356751. Zbl 1378.11069. https://doi.org/10.1093/imrn/rnu062.

[7] E. Eischen, M. Harris, J-S. Li, and C. Skinner, p-adic L-functions for unitary groups, in preparation. arXiv 1602.01776.

[8] J.-M. Fontaine and B. Perrin-Riou, Autour des conjectures de Bloch et Kato: cohomologie galoisienne et valeurs de fonctions L, in *Motives* (Seattle, WA, 1991), *Proc. Sympos. Pure Math.* **55**, Amer. Math. Soc., Providence, RI, 1994, Part 1, pp. 599–706. MR 1265546. Zbl 0821.14013. https://doi.org/10.1090/pspum/055.1/1265546.

[9] S. Friedberg and J. Hoffstein, Nonvanishing theorems for automorphic L-functions on GL(2), *Ann. of Math.* (2) **142** no. 2 (1995), 385–423. MR 1343325. Zbl 0847.11026. https://doi.org/10.2307/2118638.

[10] R. Greenberg, Iwasawa theory and p-adic deformations of motives, in *Motives* (Seattle, WA, 1991), *Proc. Sympos. Pure Math.* **55**, Amer. Math. Soc., Providence, RI, 1994, Part 2, pp. 193–223. MR 1265554. Zbl 0819.11046. https://doi.org/10.1090/pspum/055.2/1265554.

[11] B. H. Gross and D. B. Zagier, Heegner points and derivatives of L-series, *Invent. Math.* **84** no. 2 (1986), 225–320. MR 0833192. Zbl 0608.14019. https://doi.org/10.1007/BF01388809.

[12] H. Hida, A p-adic measure attached to the zeta functions associated with two elliptic modular forms. II, *Ann. Inst. Fourier* (*Grenoble*) **38** no. 3 (1988), 1–83. MR 0976685. Zbl 0645.10028. https://doi.org/10.5802/aif.1141.

[13] H. Jacquet and R. P. Langlands, *Automorphic Forms on* GL(2), *Lecture Notes in Math* **114**, Springer-Verlag, New York, 1970. MR 0401654. Zbl 0236.12010. https://doi.org/10.1007/BFb0058988.

[14] K. Kato, p-adic Hodge theory and values of zeta functions of modular forms, in *Cohomologies p-Adiques et Applications Arithmétiques. III*, *Astérisque* **295**, Soc. Math. France, Paris, 2004, pp. ix, 117–290. MR 2104361. Zbl 1142.11336. Available at https://smf.emath.fr/publications/theorie-de-hodge-p-adique-et-fonctions-zeta-de-formes-modulaires.

[15] V. A. Kolyvagin, Euler systems, in *The Grothendieck Festschrift*, *Vol. II*, *Progr. Math.* **87**, Birkhäuser Boston, Boston, MA, 1990, pp. 435–483. MR 1106906. Zbl 0742.14017. https://doi.org/10.1007/978-0-8176-4575-5_11.

[16] V. A. Kolyvagin and D. Y. Logachëv, Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties, *Algebra i Analiz* **1** no. 5 (1989), 171–196. MR 1036843. Zbl 0728.14026. Available at http://mi.mathnet.ru/eng/aa/v1/i5/p171.

[17] K. Kramer and J. Tunnell, Elliptic curves and local ε-factors, *Compositio Math.* **46** no. 3 (1982), 307–352. MR 0664648. Zbl 0496.14030. Available at http://www.numdam.org/item?id=CM_1982__46_3_307_0.

[18] J. Nekovář, *Selmer Complexes*, *Astérisque* **310**, Soc. Math. France, Paris, 2006. MR 2333680. Zbl 1211.11120. Available at http://www.numdam.org/item/AST_2006__310__R1_0/.

[19] J. Nekovář and A. Plater, On the parity of ranks of Selmer groups, *Asian J. Math.* **4** no. 2 (2000), 437–497. MR 1797592. Zbl 0973.11066. https://doi.org/10.4310/AJM.2000.v4.n2.a11.

[20] K. A. Ribet, On l-adic representations attached to modular forms. II, *Glasgow Math. J.* **27** (1985), 185–194. MR 0819838. Zbl 0596.10027. https://doi.org/10.1017/S0017089500006170.

[21] K. A. Ribet, On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms, *Invent. Math.* **100** no. 2 (1990), 431–476. MR 1047143. Zbl 0773.11039. https://doi.org/10.1007/BF01231195.

[22] K. Rubin, p-adic L-functions and rational points on elliptic curves with complex multiplication, *Invent. Math.* **107** no. 2 (1992), 323–350. MR 1144427. Zbl 0770.11033. https://doi.org/10.1007/BF01231893.

[23] J-P. Serre, Quelques applications du théorème de densité de Chebotarev, *Inst. Hautes Études Sci. Publ. Math.* no. 54 (1981), 123–201. MR 0644559. Zbl 0496.12011. Available at http://archive.numdam.org/article/PMIHES_1981__54__123_0.pdf.

[24] C. Skinner and E. Urban, The Iwasawa Main Conjectures for GL_2, *Invent. Math.* **195** no. 1 (2014), 1–277. MR 3148103. Zbl 1301.11074. https://doi.org/10.1007/s00222-013-0448-1.

[25] X. Wan, *The Iwasawa Theory for Unitary Groups*, ProQuest LLC, Ann Arbor, MI, 2012, Thesis (Ph.D.)–Princeton University. MR 3078439. Available at http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqm&rft_dat=xri:pqdiss:3522464.

[26] X. Wan, Iwasawa main conjecture for Rankin–Selberg p-adic L-functions, 2014, to appear in *Algebra Number Theory*. arXiv 1408.4044.

[27] X. Yuan, S.-W. Zhang, and W. Zhang, *The Gross-Zagier Formula on Shimura Curves*, *Ann. of Math. Stud.* **184**, Princeton Univ. Press, Princeton, NJ, 2013. MR 3237437. Zbl 1272.11082. Available at https://www.jstor.org/stable/j.cttq94r3.

[28] W. Zhang, Selmer groups and the indivisibility of Heegner points, *Camb. J. Math.* **2** no. 2 (2014), 191–253. MR 3295917. Zbl 1390.11091. https://doi.org/10.4310/CJM.2014.v2.n2.a2.

(Received: August 1, 2013)
(Revised: December 1, 2019)

Princeton University, Princeton, NJ
E-mail: cmcls@princeton.edu