

Heegner points, Stark–Heegner points, and values of L -series

Henri Darmon*

Abstract. Elliptic curves over \mathbb{Q} are equipped with a systematic collection of *Heegner points* arising from the theory of complex multiplication and defined over abelian extensions of imaginary quadratic fields. These points are the key to the most decisive progress in the last decades on the Birch and Swinnerton-Dyer conjecture: an essentially complete proof for elliptic curves over \mathbb{Q} of analytic rank ≤ 1 , arising from the work of Gross–Zagier and Kolyvagin. In [Da2], it is suggested that Heegner points admit a host of conjectural generalisations, referred to as *Stark–Heegner points* because they occupy relative to their classical counterparts a position somewhat analogous to Stark units relative to elliptic or circular units. A better understanding of Stark–Heegner points would lead to progress on two related arithmetic questions: the explicit construction of global points on elliptic curves (a key issue arising in the Birch and Swinnerton-Dyer conjecture) and the analytic construction of class fields sought for in Kronecker’s Jugendtraum and Hilbert’s twelfth problem. The goal of this article is to survey Heegner points, Stark–Heegner points, their arithmetic applications and their relations (both proved, and conjectured) with special values of L -series attached to modular forms.

Mathematics Subject Classification (2000). Primary 11G05; Secondary 11G15.

Keywords. Elliptic curves, modular forms, L -series, Heegner points, Stark–Heegner points.

1. Introduction

Elliptic curves are distinguished among projective algebraic curves by the fact that they alone are endowed with the structure of a (commutative) algebraic group. The *affine* curves with this property are the additive group \mathbb{G}_a and the multiplicative group \mathbb{G}_m . The integral points on \mathbb{G}_a (taken, say, over an algebraic number field F) is a finitely generated \mathbb{Z} -module. The same is true for the integral points on \mathbb{G}_m : these are the units of F , whose structure is well understood thanks to Dirichlet’s unit theorem. The close parallel between units and rational points on elliptic curves is frequently illuminating. In both cases, it is the natural group law on the underlying curve which lends the associated Diophantine theory its structure and richness.

An elliptic curve E over F can be described concretely as a Weierstrass equation

*The author is grateful to NSERC, CICMA, McGill University and the Centre de Recherches Mathématiques in Montreal for their support during the writing of this paper.

in projective space

$$y^2z = x^3 + axz^2 + bz^3, \quad a, b \in F, \text{ where } \Delta := 4a^3 - 27b^2 \neq 0.$$

The group $E(F)$ of F -rational (or equivalently: integral) solutions to this equation is in bijection with the F -rational solutions of the corresponding affine equation

$$y^2 = x^3 + ax + b,$$

together with an extra “point at infinity” corresponding to $(x, y, z) = (0, 1, 0)$.

The most basic result on the structure of $E(F)$ is the *Mordell–Weil Theorem* which asserts that $E(F)$ is a finitely generated abelian group, so that there is an isomorphism of abstract groups

$$E(F) \simeq T \oplus \mathbb{Z}^r,$$

where T is the finite torsion subgroup of $E(F)$. The integer $r \geq 0$ is called the *rank* of E over F . Many questions about T are well-understood, for example:

1. There is an efficient algorithm for computing T , given E and F .
2. A deep result of Mazur [Ma] describes the possible structure of T when $F = \mathbb{Q}$ and E is allowed to vary over all elliptic curves. The size of T is bounded uniformly, by 14. Mazur’s result has been generalised by Kamienny and Merel [Mer], yielding a uniform bound on the size of T when F is fixed – a bound which depends only on the degree of F over \mathbb{Q} .

In contrast, much about the rank remains mysterious. For example, can r become arbitrarily large, when F is fixed but E is allowed to vary? The answer is believed to be yes, but no proof is known for $F = \mathbb{Q}$ or for any other number field F .

An even more fundamental problem resides in the absence of effectivity in the proof of the Mordell–Weil theorem. Specifically, the answer to the following question is not known.

Question 1.1. Is there an algorithm which, given E , calculates the rank r of $E(F)$, and a system P_1, \dots, P_r of generators for this group modulo torsion?

A candidate for such an algorithm is Fermat’s method of infinite descent, but this method is not guaranteed to terminate in a finite amount of time – it would, if the so-called *Shafarevich–Tate group* $\text{III}(E/\mathbb{Q})$ of E is finite, as is predicted to be the case.

Question 1.1 is also connected with the *Birch and Swinnerton-Dyer conjecture*. This conjecture relates Diophantine invariants attached to E , such as r , to the Hasse–Weil L -series $L(E, s)$ of E , a function of the complex variable s which is defined in terms of an Euler product taken over the non-archimedean places v of F . To describe this Euler product precisely, let $\mathbb{F}_v = \mathcal{O}_F/v$ denote the residue field of F at v , and write $|v| := \#\mathbb{F}_v$ for the norm of v . The elliptic curve E is said to have *good reduction*

at v if it can be described by an equation which continues to describe a smooth curve over \mathbb{F}_v after reducing its coefficients modulo v . Set $\delta_v = 1$ if E has good reduction at v , and $\delta_v = 0$ otherwise. Finally, define integers a_v indexed by the places v of good reduction for E by setting

$$a_v := |v| + 1 - \#E(\mathbb{F}_v).$$

This definition is extended to the finite set of places of bad reduction for E , according to a recipe in which $a_v \in \{0, 1, -1\}$, the precise value depending on the type of bad reduction of E in an explicit way.

The L -series of E is given in terms of these invariants by

$$L(E, s) = \prod_v (1 - a_v |v|^{-s} + \delta_v |v|^{1-2s})^{-1} = \sum_{\mathfrak{n}} a_E(\mathfrak{n}) |\mathfrak{n}|^{-s},$$

where the product is taken over all the non-archimedean places v of F , and the sum over the integral ideals \mathfrak{n} of F . The Euler product converges absolutely for $\operatorname{Re}(s) > 3/2$, but $L(E, s)$ is expected to admit an analytic continuation to the entire complex plane. Some reasons for this expectation, and a statement of the Birch and Swinnerton-Dyer conjecture, are given in Section 2.6.

2. Elliptic curves over \mathbb{Q}

It is useful to first discuss elliptic curves over \mathbb{Q} , a setting in which a number of results currently admit more definitive formulations.

Given an elliptic curve E/\mathbb{Q} , let N denote its *conductor*. This positive integer, which measures the arithmetic complexity of E , is divisible by exactly the same primes as those dividing the minimal discriminant of E (the minimum being taken over all possible plane cubic equations describing E). Denote by a_n the coefficient of n^{-s} in the Hasse–Weil L -series of E :

$$L(E, s) = \prod_p (1 - a_p p^{-s} + \delta_p p^{1-2s})^{-1} = \sum_{n=1}^{\infty} a_n n^{-s}.$$

2.1. Modular parametrisations. Little can be asserted about the effective determination of $E(\mathbb{Q})$, or about the analytic behaviour of $L(E, s)$, without the knowledge that E is *modular*. Wiles’s far-reaching program for proving the modularity of elliptic curves (and more general Galois representations) has been completely carried out in [BCDT] when $F = \mathbb{Q}$. One way of formulating the modularity of E is to state that the generating series

$$f_E(z) := \sum_{n=1}^{\infty} a_n e^{2\pi i n z} \tag{1}$$

is a *modular form of weight 2* for the Hecke congruence group

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \text{ such that } N|c \right\}.$$

This means that $f(z)$ is a holomorphic function on the Poincaré upper half-plane

$$\mathcal{H} := \{z = x + iy, y > 0\} \subset \mathbb{C},$$

satisfying

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N), \quad (2)$$

together with suitable growth properties around the fixed points of parabolic elements of $\Gamma_0(N)$. These fixed points belong to $\mathbb{P}_1(\mathbb{Q})$, and it is useful to replace \mathcal{H} by the completed upper half-plane $\mathcal{H}^* := \mathcal{H} \cup \mathbb{P}_1(\mathbb{Q})$. After suitably defining the topology and complex structure on the quotient $\Gamma_0(N)\backslash\mathcal{H}^*$, thus making it into a compact Riemann surface, the differential form $\omega_f := 2\pi i f(z)dz$ is required to extend to a holomorphic differential on this surface.

The quotient $\Gamma_0(N)\backslash\mathcal{H}^*$ can even be identified with the set of complex points of an algebraic curve defined over \mathbb{Q} , denoted by $X_0(N)$. This algebraic curve structure arises from the interpretation of $\Gamma_0(N)\backslash\mathcal{H}$ as classifying isomorphism classes of elliptic curves with a distinguished cyclic subgroup of order N , in which the orbit $\Gamma_0(N)\tau \in \Gamma_0(N)\backslash\mathcal{H}$ is identified with the pair $(\mathbb{C}/\langle 1, \tau \rangle, \langle \frac{1}{N} \rangle)$. A (highly singular, in general) equation for $X_0(N)$ as a plane curve over \mathbb{Q} is given by the polynomial $G_N(x, y)$ of bidegree $\#\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$, where

$$G_N(x, y) \in \mathbb{Q}[x, y] \quad \text{satisfies} \quad G_N(j(\tau), j(N\tau)) = 0, \quad (3)$$

and j is the classical modular function of level 1.

An equivalent formulation of the modularity property is that there exists a non-constant map of algebraic curves defined over \mathbb{Q} ,

$$\Phi_E: X_0(N) \longrightarrow E, \quad (4)$$

referred to as the *modular parametrisation* attached to E . One of the attractive features of this modular parametrisation is that it can be computed by analytic means, without the explicit knowledge of an equation for $X_0(N)$ as an algebraic curve over \mathbb{Q} . (Such an equation, as in (3), tends to be complicated and difficult to work with numerically for all but very small values of N .)

To describe Φ_E analytically, i.e., as a map

$$\Phi_E^\infty: X_0(N)(\mathbb{C}) = \Gamma_0(N)\backslash\mathcal{H} \longrightarrow E(\mathbb{C}), \quad (5)$$

let $\Lambda_f \subset \mathbb{C}$ be the set of complex numbers of the form

$$\int_\tau^{\gamma\tau} \omega_f, \quad \text{for } \gamma \in \Gamma.$$

It can be shown that Λ_f is a lattice, and that the quotient \mathbb{C}/Λ_f is isomorphic to an elliptic curve E_f which is defined over \mathbb{Q} and is \mathbb{Q} -isogenous to E . (The curve E_f is sometimes called the *strong Weil curve* attached to f .) The modular parametrisation to E_f , denoted by Φ_f , is defined analytically by the rule

$$\Phi_f(\tau) = \int_{i\infty}^{\tau} 2\pi i f(z) dz = \sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n \tau} \pmod{\Lambda_f}, \tag{6}$$

for all $\tau \in \Gamma_0(N)\backslash\mathcal{H} \subset X_0(N)(\mathbb{C})$. The resulting value is viewed as an element of $E_f(\mathbb{C})$ via the identification $\mathbb{C}/\Lambda_f = E_f(\mathbb{C})$.

After choosing an isogeny $\alpha : E_f \rightarrow E$ defined over \mathbb{Q} , the parametrisation Φ_E is defined by setting $\Phi_E^\infty = \alpha \Phi_f$. In practice it is preferable to start with $E = E_f$, at the cost of replacing E by a curve which is isogenous to it, so that α can be chosen to be the identity. The map Φ_E^∞ is then given directly by (6).

2.2. Heegner points. Let $K \subset \mathbb{C}$ be a quadratic imaginary field, and denote by K^{ab} its maximal abelian extension, equipped with an embedding into \mathbb{C} compatible with the complex embedding of K . The following theorem, a consequence of the theory of complex multiplication, is one of the important applications of the modular parametrisation Φ_E of (5):

Theorem 2.1. *If τ belongs to $K \cap \mathcal{H}$, then $\Phi_E^\infty(\tau)$ belongs to $E(K^{\text{ab}})$.*

Theorem 2.1 also admits a more precise formulation which describes the field of definition of $\Phi_E^\infty(\tau)$. Let $M_0(N) \subset M_2(\mathbb{Z})$ denote the ring of 2×2 matrices with integer entries which are upper triangular modulo N . Given $\tau \in \mathcal{H}$, the *associated order* of τ is the set of matrices in $M_0(N)$ which preserve τ under Möbius transformations, together with the zero matrix, i.e.,

$$\mathcal{O}_\tau := \left\{ \gamma \in M_0(N) \text{ such that } \gamma \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \lambda_\gamma \begin{pmatrix} \tau \\ 1 \end{pmatrix}, \text{ for some } \lambda_\gamma \in \mathbb{C} \right\}.$$

The assignment $\gamma \mapsto \lambda_\gamma$ identifies \mathcal{O}_τ with a discrete subring of \mathbb{C} . Such rings are isomorphic either to \mathbb{Z} , or to an order in a quadratic imaginary field, the latter case occurring precisely when τ generates a quadratic (imaginary) extension of \mathbb{Q} . In that case \mathcal{O}_τ is an order in the quadratic field $K = \mathbb{Q}(\tau)$.

Orders in quadratic fields have the peculiarity that they are completely determined by their discriminants. Write D for the discriminant of the order $\mathcal{O} = \mathcal{O}_\tau$, and let $G_D := \text{Pic}(\mathcal{O})$ denote the class group of this order, consisting of isomorphism classes of projective modules of rank one over \mathcal{O} equipped with the group law arising from the tensor product. A standard description identifies G_D with a quotient of the idèle class group of K :

$$G_D = \mathbb{A}_K^\times / \left(K^\times \mathbb{C}^\times \mathbb{A}_\mathbb{Q}^\times \prod_\ell \mathcal{O}_\ell^\times \right). \tag{7}$$

Here \mathbb{A}_K^\times denotes the group of idèles of K , the product is taken over the rational primes ℓ , and $\mathcal{O}_\ell := \mathcal{O} \otimes \mathbb{Z}_\ell$. The group G_D also admits a more classical description which is well adapted to explicit computations, as the set of equivalence classes of primitive binary quadratic forms of discriminant D equipped with the classical Gaussian composition law. (For more details on this classical point of view, see Bhargava's lecture in these proceedings.)

If D and N are relatively prime, and $\mathcal{O}_\tau = \mathcal{O}_D$, there is a primitive integral binary quadratic form $F_\tau(x, y) = A_\tau x^2 + B_\tau xy + C_\tau y^2$ satisfying

$$F_\tau(\tau, 1) = 0, \quad B_\tau^2 - 4A_\tau C_\tau = D, \quad N \text{ divides } A_\tau.$$

In particular,

$$\text{all the primes } \ell|N \text{ are split in } K/\mathbb{Q}, \quad (8)$$

and therefore the equation

$$x^2 = D \pmod{N}$$

has a solution (namely, B_τ). Fix a square root δ of D modulo N , and define

$$\mathcal{H}^D := \{\tau \in \mathcal{H} \text{ such that } \mathcal{O}_\tau = \mathcal{O}_D \text{ and } B_\tau \equiv \delta \pmod{N}\}.$$

The function which to $\tau \in \Gamma_0(N) \backslash \mathcal{H}^D$ associates the $\mathrm{SL}_2(\mathbb{Z})$ -equivalence class of the binary quadratic form F_τ is a bijection. (Cf., for example, Section I.1 of [GKZ].) Through this bijection, $\Gamma_0(N) \backslash \mathcal{H}^D$ inherits a natural action of G_D via the Gaussian composition law. Denote this action by $(\sigma, \tau) \mapsto \tau^\sigma$, for $\sigma \in G_D$ and $\tau \in \Gamma_0(N) \backslash \mathcal{H}^D$.

Class field theory identifies G_D with the Galois group of an abelian extension of K , as is most readily apparent, to modern eyes, from (7). This abelian extension, denoted by H_D , is called the *ring class field* attached to \mathcal{O} , or to the discriminant D . When D is a fundamental discriminant, H_D is Hilbert class field of K , i.e., the maximal unramified abelian extension of K . Let

$$\mathrm{rec}: G_D \longrightarrow \mathrm{Gal}(H_D/K) \quad (9)$$

denote the reciprocity law map of global class field theory.

A more precise form of Theorem 2.1 is given by

Theorem 2.2. *If τ belongs to $\Gamma_0(N) \backslash \mathcal{H}^D$, then $\Phi_E^\infty(\tau)$ belongs to $E(H_D)$, and*

$$\Phi_E^\infty(\tau^\sigma) = \mathrm{rec}(\sigma)^{-1} \Phi_E^\infty(\tau), \quad \text{for all } \sigma \in G_D.$$

The fact that Φ_E^∞ intertwines the explicit action of G_D on $\Gamma_0(N) \backslash \mathcal{H}^D$ arising from Gaussian composition with the natural action of $\mathrm{Gal}(H_D/K)$ on $E(H_D)$ gives a concrete realisation of the reciprocity map (9) of class field theory. It is a special case of the *Shimura reciprocity law*.

The points $\Phi_E^\infty(\tau)$, as τ ranges over $\mathcal{H} \cap K$ are called *Heegner points* attached to K . (Sometimes, this appellation is confined to the case where the discriminant of \mathcal{O}_τ is relatively prime to N .) Theorems 2.1 and 2.2 are of interest for the following reasons, which are discussed at greater length in Sections 2.3, 2.4, and 2.5 respectively.

1. They provide a simple, computationally efficient construction of rational and algebraic points on E .
2. They are a manifestation of the fact that we dispose of an *explicit class field theory* for imaginary quadratic fields, allowing the construction of abelian extensions of such fields from values of modular functions evaluated at quadratic imaginary arguments of the upper half-plane.
3. There are deep connections between the points $\Phi_E^\infty(\tau)$, for $\tau \in \mathcal{H}^D$, and the first derivative at $s = 1$ of the Hasse–Weil L -series $L(E/K, s)$ and of related partial L -series associated to ideal classes of K . These connections lead to new insights into the behaviour of these L -series and the Birch and Swinnerton-Dyer conjecture.

2.3. The efficient calculation of global points. The fact that the theory of complex multiplication, combined with modularity, can be used to construct rational and algebraic points on E is of interest in its own right. This was noticed and exploited by Heegner, and taken up systematically by Birch in the late 60s and early 70s [BS], [Bi].

Given any (not necessarily fundamental) discriminant D for which $\mathcal{H}^D \neq \emptyset$, let $K = \mathbb{Q}(\sqrt{D})$ and set

$$P_D := \text{trace}_{H_D/\mathbb{Q}}(\Phi_E^\infty(\tau)), \quad \text{for any } \tau \in \mathcal{H}^D,$$

$$P_K := \text{trace}_{H/K}(\Phi_E^\infty(\tau)), \quad \text{for any } \tau \in \mathcal{H}^D, \quad D = \text{Disc}(K).$$

When are the points P_D and P_K of *infinite order* (in $E(\mathbb{Q})$ and $E(K)$ respectively)? This question is part of the larger problem of efficiently constructing rational or algebraic points of infinite order on elliptic curves. It is instructive to consider this problem from the point of view of its *computational complexity*.

From the outset, one is stymied by the fact that an answer to Question 1.1 is not known. Complexity issues are therefore better dealt with by focussing on the following more special problem, which depends on the curve E and a positive real parameter h . To state this problem precisely, define the *height* of a rational number $r = a/b$ (represented, of course, in lowest terms) to be

$$\text{height}(r) = \log(|ab| + 1).$$

Thus, the height of r is roughly proportional to the number of digits needed to write r down. The height of an equation is the sum of the heights of its coefficients. The height of a solution (x_1, \dots, x_n) to such an equation is taken to be the sum of the height of the x_j . (In the case of an elliptic curve, one might prefer a coordinate-free definition by taking the height of E to be the height of the minimal discriminant of E .)

It is expected that, for infinitely many E , the smallest height of a point of infinite order in $E(\mathbb{Q})$ can be at least as large as an exponential function of the height of E . In this respect, the behaviour of elliptic curves is not unlike that of Pell's equation,

where a fundamental solution to $x^2 - Dy^2 = 1$ has height roughly $O(\sqrt{D})$ if $\mathbb{Q}(\sqrt{D})$ has class number one. Of greatest relevance for complexity questions are the “worst-case” elliptic curves E for which the point of infinite order P_{\min} of smallest height in $E(\mathbb{Q})$ has height which is *large* relative to the height of E , i.e., for which

$$\text{height}(P_{\min}) \gg \exp(\text{height}(E)).$$

In order to focus on these curves, and avoid technical side-issues associated with elliptic curves having non-torsion points of small height, we formulate the following problem:

Problem 2.3. Given an elliptic curve E , and a real number

$$h > \exp(\text{height}(E)), \tag{10}$$

find a point P of infinite order on E with $\text{height}(P) < h$, if it exists, or assert that no such point exists, otherwise.

Denote by $P(E, h)$ the instance of this problem associated to E and the parameter h . In light of (10), this parameter can be chosen as a natural measure of the size of the problem.

Note that $P(E, h)$ continues to make sense for *any* Diophantine equation. Even in such great generality, problem $P(E, h)$ has the virtue of possessing an algorithmic solution: a brute force search over all possible points (in the projective space in which E is embedded) of height less than h , say. Such an exhaustive search requires $O(\exp(h))$ operations to solve an instance of $P(E, h)$. The exponential complexity of the brute force approach provides a crude benchmark against which to measure other approaches, and leads naturally to the following definition.

Definition 2.4. A class \mathcal{C} of Diophantine equations is said to be *solvable in polynomial time* if there exists $n \in \mathbb{N}$ and an algorithm that solves $P(E, h)$, with $E \in \mathcal{C}$, in at most $O(h^n)$ operations.

The property that \mathcal{C} is solvable in polynomial time can be expressed informally by stating that the time required to find a *large* solution to any $E \in \mathcal{C}$ is *not much worse* than the time it takes to write that solution down. Thus, an (infinite) class \mathcal{C} of equations being solvable in polynomial time indicates that there is a method for “zeroing in” on a solution (x_0, y_0) to any equation in \mathcal{C} in a way that is qualitatively more efficient than running through all candidates of smaller height.

The prototype for a class of equations that possess a polynomial time solution in the sense of Definition 2.4 is Pell’s equation. A polynomial time algorithm for finding a fundamental solution to $x^2 - Dy^2 = 1$ is given by the continued fraction method that was known to the Indian mathematicians of the 10th century (although Fermat seems to be the first to have shown its effectivity.) See [Le] for a more thorough discussion of Pell’s equation from the point of view of its computational complexity.

The strong analogy that exists between Pell’s equation and elliptic curves suggests that the class **ELL** of all elliptic curves over \mathbb{Q} might also be solvable in polynomial time. Indeed, Fermat’s method of infinite descent (applied, say, to a rational 2-isogeny η , if it exists) reduces $P(E, h)$ to d_E instances of $P(C_1, h/2), \dots, P(C_{d_E}, h/2)$ where the C_j are principal homogeneous spaces for E , and the number d_E is related to the cardinality of the Selmer group attached to η . Applying this remark iteratively suggests that the complexity for solving $P(E, h)$ might be a polynomial of degree related to d_E . The analysis required to make this discussion precise does not appear in the literature, and it would be interesting to determine whether the method of infinite descent can be used to determine to what extent **ELL** is solvable in polynomial time (assuming, eventually, the finiteness of the Shafarevich–Tate group of an elliptic curve).

It should be stressed that the method of descent is often complicated in practice because of the mounting complexity of the principal homogeneous spaces that arise in the procedure. On the other hand, the Heegner point construction, *when it produces a point of infinite order in $E(\mathbb{Q})$* , can be used to solve $P(E, h)$ by a method that is also extremely efficient in practice. See [E12] for a discussion of this application of the Heegner point construction.

For example, let

$$E : y^2 + y = x^3 - x^2 - 10x - 20$$

be the strong Weil curve of conductor 11. (This is the elliptic curve over \mathbb{Q} of smallest conductor.) The following table lists a few values of the x -coordinate of P_K for some more or less randomly chosen K . It takes a desktop computer a fraction of a second to find these x -coordinates, far less than would be required to find points of comparable height on the corresponding quadratic twist of E by a naive search.

Disc(K)	$x(P_K)$
−139	$\frac{-208838\sqrt{-139}-3182352}{1957201}$
−211	$\frac{-11055756376\sqrt{-211}-36342577392}{29444844025}$
−259	$\frac{64238721198\sqrt{-259}-2458030017103}{992886694969}$
−1003	$\frac{-24209041615561516569638\sqrt{-1003}-1053181310754386354274847}{219167070502034515453609}$

2.4. Explicit Class Field Theory. The Heegner point construction is a manifestation of an explicit class field theory for imaginary quadratic fields. Normally, this is stated in terms of the elliptic modular function j . The field

$$K^? := \bigcup_{\alpha \in \mathbb{Q}, \tau \in K \cap \mathcal{H}} K(e^{2\pi i \alpha}, j(\tau))$$

obtained by adjoining to the imaginary quadratic field K all the roots of unity, as well as the values $j(\tau)$ for $\tau \in K \cap \mathcal{H}$, is almost equal to the maximal abelian extension K^{ab} of K . More precisely, $K^{\text{ab}}/K^?$ is an extension whose Galois group, although infinite, has exponent two. (See [Se].)

Given a negative (not necessarily fundamental) discriminant D , let τ_1, \dots, τ_h be representatives for \mathcal{H}^D (with $N = 1$) modulo the action of $\text{SL}_2(\mathbb{Z})$. Then the so-called *modular polynomial*

$$Z_D(z) := \prod_{i=1}^h (z - j(\tau_i)) \tag{11}$$

has rational coefficients and its splitting field is the ring class field attached to the discriminant D . One might also fix an elliptic curve E and consider the function $j_E(\tau)$ of $\tau \in \Gamma_0(N) \backslash \mathcal{H}^D$ defined as the x -coordinate of the point $\Phi_E^\infty(\tau)$, where the x coordinate refers, say, to a minimal Weierstrass equation for E . Let Z_D^E denote the polynomial defined as in (11) with j replaced by j_E .

For example, consider the discriminants $D = -83, -47$, and -71 of class number 3, 5 and 7 respectively. The polynomials Z_D attached to the first two of these discriminants are given by:

$$\begin{aligned} &x^3 + 2691907584000x^2 - 41490055168000000x + 54975581388800000000 \\ &x^5 + 2257834125x^4 - 9987963828125x^3 + 5115161850595703125x^2 \\ &\quad - 14982472850828613281250x + 16042929600623870849609375. \end{aligned}$$

(The degree seven polynomial Z_{-71} has been omitted to save space, its coefficients being integers of roughly 30 digits.) The following table gives the values of the polynomials $Z_D^E(z)$ for a few elliptic curves (labelled according to the widely used conventions of the tables of Cremona [Cr2]) whose conductor is a prime that splits in $\mathbb{Q}(\sqrt{D})$, for these three discriminants.

E	$Z_{-83}^E(x)$	$Z_{-47}^E(x)$
37A	$x^3 + 5x^2 + 10x + 4$	$x^5 - x^4 + x^3 + x^2 - 2x + 1$
61A	$x^3 - 2x^2 + 2x + 1$	$x^5 - x^3 + 2x^2 - 2x + 1$
79A		$x^5 + 4x^4 + 3x^3 - 3x^2 - x + 1$

E	$Z_{-71}^E(x)$
37A	$x^7 - 2x^6 + 9x^5 - 10x^4 - x^3 + 8x^2 - 5x + 1$
43A	$x^7 + 2x^6 + 2x^5 + x^3 + 3x^2 + x + 1$
79A	$x^7 + 4x^6 + 5x^5 + x^4 - 3x^3 - 2x^2 + 1$

This data illustrates the well-known fact that in computing class fields one is often better off working with modular functions other than j (such as modular units for instance). The above data suggests (at least anecdotally) that the functions j_E can be excellent choices in certain cases. For a systematic discussion of the heights of Heegner points and of the polynomials $Z_D^E(x)$ as D varies, see [RV].

2.5. Relation with L -series. The following result of Gross and Zagier [GZ] provides a connection between Heegner points and the L -series of E over K .

Theorem 2.5. *The height of P_K is equal to an explicit non-zero multiple of $L'(E/K, 1)$.*

In particular, the point P_K is of infinite order if and only if $L'(E/K, 1) \neq 0$. This result can be exploited in two ways.

Firstly, since Heegner points are so readily computable, specific instances where the point P_K is of finite order yield non-trivial examples where $L'(E/K, 1) = 0$. The *vanishing* of the leading term in an L -series is notoriously difficult to prove numerically. The Gross–Zagier theorem makes it possible to produce elliptic curves for which, provably, $L(E, 1) = L'(E, 1) = 0$. Considerations involving the sign in the functional equation for $L(E, s)$ may even force this function to vanish to odd order, and therefore to order at least 3, at $s = 1$. (The smallest elliptic curve of prime conductor with this property has conductor 5077.) The existence of elliptic curves and modular forms whose L -series has a triple zero at $s = 1$ was exploited to great effect by Goldfeld [Go] in his effective solution of the analytic class number problem of Gauss.

Secondly, and more germane to the theme of this survey, the Gross–Zagier theorem gives a criterion for the “Heegner point method” to produce a point of infinite order on $E(K)$ or on $E(\mathbb{Q})$. This provides a neat characterization of the elliptic curves for which Heegner points lead to an efficient solution of problem $P(E, h)$.

When $\text{ord}_{s=1}(L(E, s)) \geq 2$, constructing the Mordell–Weil group $E(\mathbb{Q})$ is more elusive. It is an apparent paradox of the subject that we are the least well-equipped to produce global points on elliptic curves in precisely those cases when these points are expected to be more plentiful! (On the other hand, this reflects a common occurrence in mathematics, where an object that is uniquely defined is easier to produce explicitly.)

2.6. The Birch and Swinnerton-Dyer conjecture. The Birch and Swinnerton-Dyer conjecture relates the behaviour of $L(E, s)$ at $s = 1$ to arithmetic invariants of E over \mathbb{Q} , such as its rank. To facilitate the subsequent exposition, we state it in a form that involves an integer parameter $r \geq 0$.

Conjecture 2.6 (BSD $_r$). If $\text{ord}_{s=1} L(E, s) = r$, then the rank of $E(\mathbb{Q})$ is equal to r , and the Shafarevich–Tate group $\text{III}(E/\mathbb{Q})$ of E is finite.

The Birch and Swinnerton-Dyer conjecture predicts that $E(\mathbb{Q})$ should be infinite precisely when $L(E, 1) = 0$. (The latter condition can be easily ascertained computationally in examples, because $L(E, 1)$ is known *a priori* to belong to a specific sublattice of \mathbb{R} .)

Remark 2.7. The Birch and Swinnerton-Dyer conjecture (suitably generalised) is consistent with the presence of a systematic supply of algebraic points defined over certain ring class fields of imaginary quadratic fields. To elucidate this remark, we begin by noting that the Birch and Swinnerton-Dyer conjecture generalises to elliptic

curves over number fields, where it predicts that the rank of $E(F)$ is equal to the order of vanishing of $L(E/F, s)$ at $s = 1$. This L -series (and its twists $L(E/F, \chi, s)$ by abelian characters of $\text{Gal}(\bar{F}/F)$) admits a functional equation relating $L(E/F, \chi, s)$ to $L(E/F, \bar{\chi}, 2 - s)$. Suppose that E is defined over \mathbb{Q} , that $F = K$ is a quadratic extension of \mathbb{Q} , and that $\chi : \text{Gal}(H/K) \rightarrow \mathbb{C}^\times$ factors through the Galois group of a ring class field H of K . Then the definition of $L(E/K, \chi, s)$ as an Euler product shows that

$$L(E/K, \chi, s) = L(E/K, \bar{\chi}, s).$$

The *sign* that appears in the functional equation of the L -series $L(E/K, \chi, s)$, denoted by $\text{sign}(E/K, \chi) \in \{-1, 1\}$, therefore determines the parity of its order of vanishing $\text{ord}_{s=1}(L(E/K, \chi, s))$.

When (E, K) satisfies the *Heegner hypothesis* of equation (8), it can be shown that $\text{sign}(E, K) = -1$ so that $L(E/K, 1) = 0$. Moreover, the same is true of $\text{sign}(E/K, \chi)$ when χ is *any* ring class character of conductor prime to N_E , so that $L(E/K, \chi, 1) = 0$ for such ring class characters. In particular, if H is a ring class field of K of discriminant prime to N_E , we find

$$\text{ord}_{s=1} L(E/H, s) = \text{ord}_{s=1} \left(\prod_{\chi \in \widehat{\text{Gal}(H/K)}} L(E/K, \chi, s) \right) \geq [H : K], \quad (12)$$

so that the Birch and Swinnerton-Dyer conjecture predicts the inequality:

$$\text{rank}(E(H)) \stackrel{?}{\geq} [H : K]. \quad (13)$$

The Gross–Zagier formula (Theorem 2.5), suitably generalised to the L -series $L(E/K, \chi, s)$ with character, as in the work of Zhang discussed in Section 3.4, makes it possible to bound the rank of $E(H)$ from below by establishing the non-triviality of certain Heegner points, and yields

Corollary 2.8. *If the inequality in (12) is an equality, then the inequality (13) holds.*

A short time after the proof of the Gross–Zagier formula, Kolyvagin discovered a general method for using Heegner points to bound the ranks of Mordell–Weil groups *from above*.

Theorem 2.9 (Kolyvagin). *If P_K is of infinite order, then $E(K)$ has rank one and $\text{III}(E/K)$ is finite.*

Crucial to Kolyvagin’s proof is the fact that the Heegner point P_K does not come alone, but is part of an infinite collection of algebraic points

$$\{\Phi_E^\infty(\tau)\}_{\tau \in \mathcal{H}^D}$$

as D ranges over all discriminants of orders in K . These points are defined over abelian extensions of K and obey precise compatibility relations under the norm

maps. They are used to construct a supply of cohomology classes that can be used, under the non-triviality assumption on P_K , to bound $E(K)$ and $\text{III}(E/K)$, showing that the former has rank one and the latter is finite. See [Ko] (or the expositions given in [Gr3] or Chapter X of [Da2]) for the details of the argument.

In relation with Corollary 2.8 we note the following consequence of Theorem 2.9 (suitably adapted to the problem of bounding Mordell–Weil groups over ring class fields in terms of Heegner points, as in [BD1] for example)

Corollary 2.10. *If the inequality in (12) is an equality, then the inequality predicted in (13) is an equality.*

Theorem 2.9 completes Theorem 2.5 by relating the system of Heegner points attached to E/K to the arithmetic of E over K . When combined with Theorem 2.5, it yields the following striking evidence for the Birch and Swinnerton-Dyer conjecture.

Theorem 2.11. *Conjectures BSD_0 and BSD_1 are true for all elliptic curves over \mathbb{Q} .*

Sketch of proof. If $\text{ord}_{s=1} L(E, s) \leq 1$, one can choose an auxiliary quadratic imaginary field K in which all the primes dividing N are split, and for which

$$\text{ord}_{s=1} L(E/K, s) = 1.$$

The existence of such a K is a consequence of non-vanishing results for special values and derivatives of twisted L -series. (See the book [MM], for example, for an attractive exposition of these results.) After choosing such a K , Theorem 2.5 implies that P_K is of infinite order, since $L'(E/K, 1) \neq 0$. Theorem 2.9 then implies that P_K generates a finite index subgroup of $E(K)$, and that $\text{III}(E/K)$ is finite. Explicit complementary information on the action of $\text{Gal}(K/\mathbb{Q})$ on the point P_K implies that the rank of $E(\mathbb{Q})$ is at most one, with equality occurring precisely when $L(E, 1) = 0$. The finiteness of $\text{III}(E/K)$ directly implies the finiteness of $\text{III}(E/\mathbb{Q})$ since the restriction map $\text{III}(E/\mathbb{Q}) \rightarrow \text{III}(E/K)$ has finite kernel. \square

Theorem 2.11 is the best evidence at present for Conjecture 2.6. We remark that almost nothing is known about this conjecture when $r > 1$.

3. Elliptic curves over totally real fields

Summarising the discussion of the previous chapter, the Heegner point construction (attached to an elliptic curve over \mathbb{Q} , and a quadratic imaginary field K) is appealing because it provides an elegant and efficient method for calculating global points on elliptic curves as well as class fields of imaginary quadratic fields. It also leads to a proof of Conjecture BSD_r for $r = 0$ and 1.

It is therefore worthwhile to investigate whether elliptic curves defined over a number field F other than \mathbb{Q} are equipped with a similar collection of algebraic

points. The modularity property so crucial in defining Heegner points does have an analogue for elliptic curves defined over F , which is most conveniently couched in the language of automorphic representations: an elliptic curve E/F should correspond to an automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_F)$, the correspondence being expressed in terms of an equality of associated L -series:

$$L(E, s) = L(\pi, s).$$

(For an explanation of these concepts, see for example [Ge] or [BCdeSGKK].)

When $F = \mathbb{Q}$, the automorphic form attached to E corresponds to a differential on a modular curve, and leads to the modular parametrisation Φ_E^∞ of (4). Unfortunately, such a geometric formulation of modularity is not always available; therefore the Heegner point construction does not carry over to other number fields without further ideas.

The number fields for which Heegner points are best understood are the *totally real fields*. Let F be such a field, of degree ν , and fix an ordering v_1, \dots, v_ν on the real embeddings of F . For $x \in F$, write $x_j := v_j(x)$ ($1 \leq j \leq \nu$). The v_j determine an embedding of F into \mathbb{R}^ν and an embedding of $\mathrm{SL}_2(\mathcal{O}_F)$ as a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})^\nu$ with finite covolume. Given any ideal \mathcal{N} of \mathcal{O}_F , denote by $\Gamma_0(\mathcal{N})$ the subgroup of $\mathrm{SL}_2(\mathcal{O}_F)$ consisting of matrices which are upper-triangular modulo \mathcal{N} .

Assume now for simplicity that F has *narrow class number one*. (The definitions to be made below need to be modified in the general case, by adopting adèlic notation which is better suited to working in greater generality but might also obscure the analogy with the classical case that we wish to draw.) A *Hilbert modular form* of parallel weight 2 and level \mathcal{N} is a holomorphic function $f(z_1, \dots, z_\nu)$ on \mathcal{H}^ν satisfying the transformation rule analogous to (2), for all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(\mathcal{N})$:

$$f\left(\frac{a_1 z_1 + b_1}{c_1 z_1 + d_1}, \dots, \frac{a_\nu z_\nu + b_\nu}{c_\nu z_\nu + d_\nu}\right) = (c_1 z_1 + d_1)^2 \dots (c_\nu z_\nu + d_\nu)^2 f(z_1, \dots, z_\nu), \quad (14)$$

together with suitable growth properties around the fixed points of parabolic elements of $\Gamma_0(\mathcal{N})$, which imply in particular that f admits a Fourier expansion “near infinity”

$$f(z_1, \dots, z_\nu) := a(0) + \sum_{\mathfrak{n} \gg 0} a(\mathfrak{n}) e(\delta^{-1} \mathfrak{n} \cdot z),$$

in which the sum is taken over all totally positive $\mathfrak{n} \in \mathcal{O}_F$, δ is a totally positive generator of the different ideal of F , and

$$e(\mathfrak{n} \cdot z) := \exp(2\pi i (n_1 z_1 + \dots + n_\nu z_\nu)).$$

Let $N \in F$ be a totally positive generator of the conductor ideal of E over F , and let $a_E(\mathfrak{n})$ denote the coefficients in the Hasse–Weil L -series of this elliptic curve. The following conjecture is a generalisation of the Shimura–Taniyama–Weil conjecture for totally real fields (of narrow class number one)

Conjecture 3.1. The generating series analogous to (1)

$$f_E(z_1, \dots, z_v) := \sum_{\mathfrak{n} \gg 0} a_E(\mathfrak{n}) e(\delta^{-1} \mathfrak{n} \cdot z)$$

is a modular form of parallel weight 2 and level \mathcal{N} .

The methods of Wiles have successfully been extended to prove many instances of Conjecture 3.1, under a number of technical hypotheses. (See [SW] [Fu] for example.) In the sequel, it will always be assumed that any elliptic curve E/F satisfies the conclusion of Conjecture 3.1, to avoid having to worry about the precise technical conditions under which this is known unconditionally. (These conditions are fluid and ever-changing, and one might hope that they will eventually be completely dispensed with. This hope is bolstered by the wealth of new ideas – which the reader can appreciate, for instance, by consulting [BCDT], [SW], or [Ki], to cite just three in a roster that is too long and rapidly evolving to give anything like a complete list – emerging from the branch of number theory devoted to generalising and extending the scope of Wiles’s methods.)

The differential form $\omega_f := f(z_1, \dots, z_v) dz_1 \dots dz_v$ defines a $\Gamma_0(\mathcal{N})$ -invariant holomorphic differential on $\Gamma_0(\mathcal{N}) \backslash \mathcal{H}^v$, but these objects do not give rise to a modular parametrisation. (Indeed, the natural generalisation of modular curves are Hilbert modular varieties, which are of dimension $[F : \mathbb{Q}]$ and probably do not admit any non-constant maps to E when $F \neq \mathbb{Q}$.) To define Heegner points on $E(F)$, it becomes crucial to consider *Shimura curve* parametrisations arising from automorphic forms on certain quaternion algebras.

3.1. Shimura curve parametrisations. Let S be a set of places of odd cardinality, containing all the archimedean places of F . Associated to S there is a *Shimura curve* denoted by X_S . This curve has a canonical model over F arising from a connection between it and the solution to a moduli problem classifying abelian varieties with “quaternionic endomorphisms”. (Cf. Section 1.1 of [Zh1], for example, where it is called M_K .)

For each place $v \in S$, the curve X_S also admits an explicit v -adic analytic description. Since this description is useful for doing concrete calculations with X_S , we now describe it in some detail, following a presentation that the author learned from Gross. (Cf. [Gr4].)

If $v \in S$ is an archimedean (and hence, real) place, denote by \mathcal{H}_v the Poincaré upper half-plane. If v is non-archimedean, let \mathbb{C}_v denote the completion of the algebraic closure of F_v , and let $\mathcal{H}_v := \mathbb{P}_1(\mathbb{C}_v) - \mathbb{P}_1(F_v)$ denote the v -adic upper half-plane. It is equipped with a natural structure as a v -adic analytic space which plays the role of the complex structure on \mathcal{H} in the non-archimedean case. (See for instance Chapter IV of [Da2] for a description of this structure.)

Let B denote the quaternion algebra over F which is ramified precisely at the places of $S - \{v\}$. (Since this set of places has even cardinality, such a quaternion

algebra exists; it is unique up to isomorphism.) Identifying v with the corresponding embedding $F \rightarrow F_v$ of F into its completion at v , there is an F_v -algebra isomorphism

$$\iota_v : B \otimes_v F_v \longrightarrow M_2(F_v).$$

Let R denote a maximal \mathcal{O}_F -order of B if v is archimedean, and a maximal $\mathcal{O}_F[1/v]$ -order of B if v is non-archimedean, and write R_1^\times for the group of elements of R of reduced norm 1. Then $\Gamma_v := \iota_v(R_1^\times)$ is a discrete and finite covolume (and compact, if $(F, S) \neq (\mathbb{Q}, \infty)$) subgroup of $\mathrm{SL}_2(F_v)$. The quotient $\Gamma \backslash \mathcal{H}_v$ is naturally equipped with the structure of a complex curve (if v is real) or of a rigid analytic curve over \mathbb{C}_v (if v is non-archimedean).

Theorem 3.2. *The quotient $\Gamma \backslash \mathcal{H}_v$ is analytically isomorphic to $X_S(\mathbb{C}_v)$.*

The complex uniformisation of $X_S(\mathbb{C})$ at the real places of F follows directly from the description of X_S in terms of the solution to a moduli problem. The non-archimedean uniformisation follows from the theory of Cerednik and Drinfeld. For more details on Drinfeld's proof of Theorem 3.2 for v non-archimedean see [BC].

If \mathcal{N}^+ is any ideal (or totally positive element) of F prime to the places of S , one can also define a Shimura curve $X_S(\mathcal{N}^+)$ by adding ‘‘auxiliary level structure’’ of level \mathcal{N}^+ .

Denote by J_S and $J_S(\mathcal{N}^+)$ the jacobian varieties of X_S and $X_S(\mathcal{N}^+)$ respectively. The relevance of these jacobians is that they are expected to parametrise certain elliptic curves over F in the same way that jacobians of modular curves uniformise elliptic curves over \mathbb{Q} .

More precisely, a (modular, in the sense of Conjecture 3.1) elliptic curve E over F is said to be *arithmetically uniformisable* if there exists a Shimura curve $X_S(\mathcal{M})$ and a non-constant map of abelian varieties over F , generalising (4)

$$\Phi_{S,E} : J_S(\mathcal{M}) \longrightarrow E. \tag{15}$$

Conjecture 3.1 leads one to expect that many (*but not all*, in general!) elliptic curves over F are arithmetically uniformisable. More precisely,

Theorem 3.3. *A modular elliptic curve E over F is arithmetically uniformisable if and only if at least one of the following conditions holds.*

1. *The degree of F over \mathbb{Q} is odd;*
2. *There is a place v of F for which $\mathrm{ord}_v(\mathcal{N})$ is odd.*

When condition 1 is satisfied, a Shimura curve uniformising E can be taken to be of the form $X_S(\mathcal{N}_E)$, where $S = S_\infty$ is the set of archimedean places of F . If condition 1 is not satisfied, but 2 is, one can consider a Shimura curve associated to $S = \{v\} \cup S_\infty$ with a suitable choice of level structure. See [Zh1] for more details on Shimura curves and their associated modular parametrisations.

3.2. Heegner points. From now on we assume that E/F is *semistable*, and that there is a factorisation $\mathcal{N} = \mathcal{N}^+ \mathcal{N}^-$ of the conductor into a product of ideals with the property that the set of places of F

$$S := \{v \text{ divides } \infty \text{ or } \mathcal{N}^-\}$$

has odd cardinality. (Placing oneself in this special situation facilitates the exposition, and does not obscure any of the essential features we wish to discuss.) This assumption implies that E is arithmetically uniformisable and occurs as a quotient of the Jacobian $J_S(\mathcal{N}^+)$ of the Shimura curve $X_S(\mathcal{N}^+)$ of the previous section. Let

$$\Phi_{S,E}^{\mathcal{N}^+}: \text{Div}^0(X_S(\mathcal{N}^+)) \longrightarrow E \tag{16}$$

denote the Shimura curve parametrisation attached to this data.

Just like classical modular curves, the curve $X_S(\mathcal{N}^+)$ is also equipped with a collection of CM points attached to certain CM extensions of F . More precisely, let K be a quadratic extension of F satisfying:

1. For all places $v \in S$, the F_v -algebra $K \otimes_v F_v$ is a field.
2. For all places $v \nmid \mathcal{N}^+$, the F_v -algebra $K \otimes_v F_v$ is isomorphic to $F_v \oplus F_v$.

Note that condition 1 implies in particular that K is a CM extension of F , since S contains all the archimedean places of F .

Fix an \mathcal{O}_F -order \mathcal{O} of K , and let H denote the associated ring class field of K . There is a canonical collection $CM(\mathcal{O}) \subset X_S(\mathcal{N}^+)(H)$ associated, in essence, to solutions to the moduli problem related to $X_S(\mathcal{N}^+)$ which have “extra endomorphisms by \mathcal{O} .” This fact allows an extension of the theory of Heegner points to the context of totally real fields.

3.3. The efficient calculation of global points. Assume for notational simplicity that $\mathcal{N}^+ = 1$. From a computational perspective, it would be useful to have efficient numerical recipes for computing the points of $CM(\mathcal{O})$ and their images in $E(H)$ under the parametrisation $\Phi_{S,E}$ of (16). Difficulties arise in calculating Heegner points arising from Shimura curve parametrisations, largely because the absence of Fourier expansions for modular forms on $\Gamma \backslash \mathcal{H}_v$ prevents one from writing down an explicit analytic formula for $\Phi_{S,E}$ analogous to (6).

The article [E11] proposes to work with Shimura curves by computing algebraic equations for them. This approach can be carried out when the group Γ arising in an archimedean uniformisation of $X_S(\mathbb{C})$ following Theorem 3.2 is contained with small index in a Hecke triangle group. Adapting the ideas of [KM] to the context of Shimura curves might also yield a more systematic approach to these types of questions. Nonetheless, it appears that an approach relying on an explicit global equation for the Shimura curve may become cumbersome, since such an algebraic equation is expected to be quite complicated for even modest values of F and S .

Alternately, one may try to exploit the non-archimedean uniformisations of $X_S(\mathbb{C}_v)$ given by Theorem 3.2. Given a non-archimedean place $v \in S$, let B and R be the quaternion algebra and Eichler order associated to S and v as in the statement of this theorem. An F -algebra embedding

$$\Psi: K \longrightarrow B$$

is said to be *optimal* relative to \mathcal{O} if $\Psi(K) \cap R = \Psi(\mathcal{O})$. It can be shown that the number of distinct optimal embeddings of K into B , up to conjugation by the normaliser of R^\times in B^\times , is equal to the class number of \mathcal{O} . Let h denote this class number and let Ψ_1, \dots, Ψ_h be representatives for the distinct conjugacy classes of optimal embeddings of \mathcal{O} into R . Let τ_j and $\bar{\tau}_j$ denote the fixed points for $\Psi_j(K^\times)$ acting on \mathcal{H}_v . Then the points in $CM(\mathcal{O})$ are identified with the points $\tau_j, \bar{\tau}_j$ under the identification of Theorem 3.2.

In his thesis [Gre], Matthew Greenberg exploits this explicit v -adic description of the points in $CM(\mathcal{O})$ and computes their images in $E(\mathbb{C}_v)$ analytically. The absence of cusps on X_S and of the attendant Fourier expansion of modular forms is remedied in part by an alternate combinatorial structure on $X_S(\mathbb{C}_v)$ which allows explicit v -adic analytic calculations with cusp forms on X_S . This combinatorial structure arises from the *reduction map*

$$r: \mathcal{H}_v \longrightarrow \mathcal{T}$$

on \mathcal{H}_v , where \mathcal{T} is the *Bruhat–Tits tree* of $\mathrm{PGL}_2(F_v)$, a homogeneous tree with valency $|v| + 1$. Thanks to this structure, rigid analytic modular forms of weight two on $\Gamma \backslash \mathcal{H}_v$ admit a simple description as functions on the edges of the quotient graph $\Gamma \backslash \mathcal{T}$ satisfying a suitable harmonicity property. (For a more detailed discussion of the description of rigid analytic modular forms on $\Gamma \backslash \mathcal{H}_v$ in terms of an associated Hecke eigenfunction on the edges of the Bruhat–Tits tree, see Chapters 5 and 6 of [Da2] for example.)

Greenberg explains how the knowledge of the eigenfunction on $\Gamma \backslash \mathcal{T}$ associated to E can be parlayed into an efficient algorithm for computing the Shimura curve parametrisation $\Phi_{S,E}$ of (16), viewed as a v -adic analytic map

$$\Phi_{S,E}^v: \mathrm{Div}^0(\Gamma \backslash \mathcal{H}_v) \longrightarrow E(\mathbb{C}_v).$$

The main ingredient in Greenberg’s approach is the theory of “overconvergent modular symbols” developed in [PS], adapted to the context of automorphic forms on definite quaternion algebras.

For example, setting $\omega = \frac{1+\sqrt{5}}{2}$, Greenberg considers the elliptic curve

$$E: y^2 + xy + \omega y = x^3 + (-\omega - 1)x^2 + (-30\omega - 45)x + (-111\omega - 117)$$

defined over $F = \mathbb{Q}(\sqrt{5})$. This curve has conductor $\mathcal{N} = v = (3 - 5\omega)$, a prime ideal above 31. Consider the CM extension $K = F(\sqrt{-\omega - 5})$ of F . It has class number two, and its Hilbert class field is equal to $H = K(i)$ (where $i = \sqrt{-1}$) by genus

theory. Letting $\tau \in \mathcal{H}_v$ be an element of $CM(\mathcal{O}_K)$, and τ' its translate by the element of order 2 in the class group of K , Greenberg computes the image of $\Phi_{S,E}^v((\tau) - (\tau'))$ in $E(K_v)$ to a v -adic accuracy of 31^{-30} , obtaining a point that agrees with the global point

$$P = \left(\frac{578\omega - 1}{90}, -\frac{27178\omega + 9701}{2700}i - \frac{668\omega - 1}{180} \right)$$

to that degree of accuracy.

The calculations of [Gre] convincingly demonstrate that Heegner points arising from Shimura curve parametrisations can be computed fairly systematically in significant examples using the Cerednik–Drinfeld theory. It would be interesting to understand whether the archimedean uniformisations described in Theorem 3.2 can be similarly exploited.

3.4. Relation with L -series. Retaining the notations of the previous section, let P be any point of $CM(\mathcal{O}) \subset X_S(H)$, and let χ be a character of $G = \text{Gal}(H/K)$. Suppose for simplicity that this character is non-trivial, so that

$$D_\chi := \sum_{\sigma \in G} \chi(\sigma) P^\sigma \text{ belongs to } \text{Div}^0(X_S(H)) \otimes \mathbb{C}.$$

Let P_χ denote the image of D_χ ,

$$P_\chi := \Phi_{S,E}(D_\chi).$$

The Heegner point P_χ enjoys the following property analogous to the formula of Gross and Zagier.

Theorem 3.4 (Zhang). *The height of P_χ is equal to an explicit non-zero multiple of $L'(E/K, \chi, 1)$.*

The proof of Theorem 3.4, which is explained in [Zh1], [Zh2], and [Zh3], proceeds along general lines that are similar to those of [GZ] needed to handle the case $F = \mathbb{Q}$, although significant new difficulties have to be overcome in handling Shimura curve parametrisations. Note that, even when $F = \mathbb{Q}$, Zhang’s theorem asserts something new since an elliptic curve over \mathbb{Q} may possess, along with the usual modular curve parametrisation, a number of Shimura curve parametrisations.

3.5. The Birch and Swinnerton-Dyer conjecture. Zhang’s formula has applications to the arithmetic of elliptic curves defined over totally real fields that are analogous to those of the original Gross–Zagier formula.

Theorem 3.5. *Suppose that E is arithmetically uniformisable. Then conjectures BSD_0 and BSD_1 are true for E .*

Sketch of proof. Since E is arithmetically uniformisable, there is a Shimura curve $X_S(\mathcal{M})$ parametrising E , for an appropriate $\mathcal{M}|\mathcal{N}_E$. If $\text{ord}_{s=1} L(E, s) \leq 1$, one can choose as in the proof of Theorem 2.11 an auxiliary quadratic CM extension K of F in which all the primes of S are inert, those dividing \mathcal{M} are split, and for which

$$\text{ord}_{s=1} L(E/K, s) = 1.$$

After choosing such a K , the Heegner point P_K attached to K and the parametrisation (15) is of infinite order by Theorem 3.4. A natural extension of Kolyvagin's Theorem 2.9 to the context of totally real fields has been proved by Kolyvagin and Logachev [KL]. Their result implies that P_K generates a subgroup of $E(K)$ of finite index, and that $\text{III}(E/K)$ is finite. Theorem 3.5 now follows much as in the proof of Theorem 2.11. \square

The proof of Theorem 3.5 sketched above breaks down for elliptic curves that are not arithmetically uniformisable in the sense of Theorem 3.3. This is the case for the elliptic curve

$$E : y^2 + xy + \varepsilon^2 y = x^3, \quad \varepsilon = \frac{5 + \sqrt{29}}{2} \in \mathcal{O}_F^\times. \quad (17)$$

defined over the real quadratic field $F = \mathbb{Q}(\sqrt{29})$ and having everywhere good reduction over F .

Remark 3.6. It should be noted however that the curve E of (17) is isogenous to a quotient of the modular Jacobian $J_1(29)$, this circumstance arising from the fact that E is a \mathbb{Q} -curve, i.e., is isogenous to its Galois conjugate. Hence a variant of the Heegner point construction exploiting CM points on $X_1(29)$ might provide some information on the arithmetic of E .

In light of this remark, an even more puzzling example is given by the following elliptic curve discovered by R. Pinch,

$$y^2 - xy - \omega y = x^3 + (2+2\omega)x^2 + (162+3\omega)x + (71+34\omega), \quad \omega = \frac{1 + \sqrt{509}}{2}, \quad (18)$$

which has everywhere good reduction over $F = \mathbb{Q}(\sqrt{509})$, and is *not* isogenous to its Galois conjugate. The curve given by (18), and any of its quadratic twists over F , are elliptic curves for which no variant of the Heegner point construction relying on CM points is known. For such elliptic curves, the strategy of proof of Theorem 3.5 runs across a fundamental barrier.

In spite of this the following theorem has been proved independently in [Lo1], [Lo2] and [TZ].

Theorem 3.7 (Longo, Tian-Zhang). *Suppose that E is any (modular) elliptic curve over a totally real field F . Then conjecture BSD_0 is true for E .*

Sketch of proof. We indicate the idea of the proof in the simplest case where E has everywhere good reduction over a real quadratic field F . Let K be any CM extension of F , and fix a rational prime p . The key fact is that, even though E is *not* arithmetically uniformisable, it is still possible to produce a sequence X_1, \dots, X_n, \dots of Shimura curves in such a way that the Galois module given by the p^n -torsion $E[p^n]$ of E appears as a Jordan–Hölder constituent of $J_n[p^n]$, where J_n denotes the Jacobian of X_n . The Shimura curve X_n is associated to the set $S_n := \{\ell_n, \infty_1, \infty_2\}$ of places of F , for a judiciously chosen (non-archimedean) place ℓ_n of F . The existence of X_n follows from the theory of congruences between modular forms and the Jacquet–Langlands correspondence. The Heegner point attached to K and X_n can then be used to produce, following a variant of Kolyvagin’s original recipe, a global cohomology class in $H^1(K, J_n[p^n])$, and, from this, a class $\kappa_n \in H^1(K, E[p^n])$. A key formula, whose proof exploits the Cerednik–Drinfeld theory of ℓ_n -adic uniformisation of X_n , relates the restriction of κ_n in the local cohomology group $H^1(K_{\ell_n}, E[p^n])$ to the special value of $L(E/K, 1)$. (More precisely, to a suitable *algebraic part*, taken modulo p^n .) In particular, if this special value is non-zero, then the class κ_n is non-trivial for n sufficiently large. (In fact, this is even so locally at ℓ_n .) This local control of the classes κ_n is enough to prove (following the lines of Kolyvagin’s original argument) that the p^n -Selmer group of E over K has cardinality bounded independently of n , and therefore that $E(K)$ and the p -primary component of $\text{III}(E/K)$ are both finite. The same finiteness results hold *a fortiori* with K replaced by F . It is in ensuring the existence of a suitable auxiliary CM field K for which $L(E/K, 1) \neq 0$ that the non-vanishing hypothesis on $L(E/F, 1)$ made in the statement of Conjecture BSD₀ is used in a crucial way. \square

A similar approach to bounding the Selmer group of E relying on congruences between modular forms was first exploited in [BD3] where it was used to prove part of the “main conjecture” of Iwasawa Theory attached to an elliptic curve E/\mathbb{Q} and the anticyclotomic \mathbb{Z}_p -extension of an imaginary quadratic field K .

Theorem 3.7 notwithstanding, the following question retains an alluring aura of mystery.

Question 3.8. Prove Conjecture BSD₁ for elliptic curves over totally real fields that are *not* arithmetically uniformisable.

For example, let E_0 be an elliptic curve with everywhere good reduction over a real quadratic field F such as the curve given in equations (17) and (18). Let K be a quadratic extension of F which is neither totally real nor complex, i.e., an extension with one complex and two real places. Let E denote the twist of E_0 by K . It can be shown that $\text{sign}(E, F) = -1$, so that $L(E/F, s)$ vanishes to odd order. Can one show that $E(F)$ is infinite, if $L'(E/F, 1) \neq 0$? This would follow from a suitable variant of Theorem 2.5 or 3.4, but it is unclear how such a variant could be proved – or even formulated precisely! – in the absence of a known Heegner point construction for E .

4. Stark–Heegner points

Question 3.8 points out one among many instances where Heegner points are not sufficient to produce algebraic points on elliptic curves, even when the presence of such points is predicted by the Birch and Swinnerton-Dyer conjecture.

The notion of *Stark–Heegner point* is meant to provide a *conjectural* remedy by proposing constructions in a number of situations lying ostensibly outside the scope of the theory of complex multiplication.

4.1. ATR extensions of totally real fields. Let F be a totally real field of narrow class number 1, as in Section 3. A quadratic extension K of F is said to be *almost totally real* (or “ATR” for short) if it has exactly one complex place, so that the remaining real places split in K/F . The field K can be viewed as a subfield of \mathbb{C} via its unique complex embedding. A point in the complex upper half-plane is called an *ATR point* if it generates an ATR extension of F . Let \mathcal{H}' denote the set of all ATR points on \mathcal{H} , relative to a fixed real place v of F . Note that \mathcal{H}' is preserved under the action of the Hecke congruence group $\Gamma_0(\mathcal{N}) \subset \mathrm{SL}_2(\mathcal{O}_F)$, although, because the action of this group is not discrete, the quotient $\Gamma_0(\mathcal{N}) \backslash \mathcal{H}'$ inherits no obvious topology (other than the discrete one). Let f_E denote the Hilbert modular form of level \mathcal{N} associated to E in Conjecture 3.1, and write

$$\omega_f := f_E(z_1, \dots, z_v) dz_1 \dots dz_v$$

for the corresponding $\Gamma_0(\mathcal{N})$ -invariant differential form on \mathcal{H}^v . The article [DL] describes a kind of *natural substitute* of the modular parametrisation attached to E , denoted

$$\Phi_E^v : \Gamma_0(\mathcal{N}) \backslash \mathcal{H}' \longrightarrow E(\mathbb{C}). \quad (19)$$

A precise description of this map is given in Chapter VIII of [Da2] as well as in [DL]. We will not recount the details of this construction here, mentioning only that Φ_E^v is defined in terms of the periods of ω_f . It is in that sense that it can be viewed as purely analytic, even though Φ_E^v does not extend to a holomorphic or even continuous map on \mathcal{H} (as is apparent from the fact that $\Gamma_0(\mathcal{N})$ acts on \mathcal{H} with dense orbits). We note that the definition of Φ_E^v is quite concrete and lends itself well to computer calculations. In fact, working with the Hilbert modular form attached to E has the added computational advantage that the Fourier expansion of ω_f is available as an aid to computing its periods numerically.

The main conjecture that is spelled out precisely in [DL] is that the points $\{\Phi_E^v(\tau)\}_{\tau \in \mathcal{H}' \cap K}$ belong to ring class fields of the ATR extension K of F , and that they enjoy all the properties (Shimura reciprocity law, norm compatibility relations) of classical Heegner points. This conjecture is also tested numerically and used to produce global points on the elliptic curve of equation (17) in terms of periods of the associated Hilbert modular form over $\mathbb{Q}(\sqrt{29})$. A proof of the conjectures of [DL] (an admittedly tall order, at present) would presumably lead to a solution to Question 3.8 proceeding along much the same lines as the proof of Theorems 2.11 and 3.5.

4.2. Ring class fields of real quadratic fields. We return now to the setting where E is an elliptic curve over \mathbb{Q} . Little changes in the analysis of Remark 2.7 when the imaginary quadratic field is replaced by a *real* quadratic field. Hence, if K is such a field and $\text{sign}(E, K) = -1$, one expects the presence of a systematic collection of points defined over various ring class fields of K . This is intriguing, since the theory of complex multiplication gives no means of producing these points.

Suppose now that the conductor of E is the form $N = pM$, where p is a prime that does not divide M , so that E has multiplicative reduction at p . Let K be a *real* quadratic field satisfying the following “modified” Heegner hypothesis:

1. All the primes dividing M are split in K ;
2. The prime p is inert in K .

These conditions are analogous to the ones that are imposed in the setting of classical Heegner points, with the prime p now playing the role of ∞ . It can be shown that $\text{sign}(E, K) = -1$, and the same holds for all twists of $L(E/K, s)$ by ring class characters of conductor prime to N . The analysis carried out in Remark 2.7 therefore shows that if H is any ring class field of K of discriminant prime to N , one has the same inequality as in (12):

$$\text{ord}_{s=1}(L(E/H, s)) \geq [H : K]. \tag{20}$$

The article [Da1] describes a conjectural recipe for constructing certain canonical points in $E(H)$, which is expected to yield a subgroup of finite index in $E(H)$ whenever (20) is an equality.

The idea behind the construction of [Da1] is to attach p -adic periods to f in a way which formally suggests viewing f as a “mock Hilbert modular form” on $\Gamma \backslash (\mathcal{H}_p \times \mathcal{H})$, where $\Gamma \subset \text{SL}_2(\mathbb{Z}[1/p])$ is the subgroup of matrices which are upper triangular modulo M . The construction of these p -adic periods, which is described in [Da1], is essentially elementary. The main ingredient that enters in their definition is the theory of *modular symbols* associated to f , which states that the period integral

$$I_f\{r \rightarrow s\} := \frac{1}{\Omega^+} \text{Re} \left(\int_r^s 2\pi i f(z) dz \right), \quad r, s \in \mathbb{P}_1(\mathbb{Q}) \tag{21}$$

takes *integer values* for a suitable choice of “real period” $\Omega^+ \in \mathbb{R}$, which is, up to a non-zero rational multiple, the real period in the Néron lattice of E .

Further pursuing the analogy with the setting of Section 4.1, the counterpart of the set \mathcal{H}' of ATR points in \mathcal{H} (associated to the real quadratic base field F and a choice of real embedding) is the collection \mathcal{H}'_p of elements of \mathcal{H}_p which generate a *real quadratic* extension of \mathbb{Q} . In particular, after fixing a p -adic embedding $K \subset \mathbb{C}_p$, the set $\mathcal{H}'_p \cap K$ is *non-empty*. Mimicking the formal aspects of the definition of the map (19) of Section 4.1, with the complex periods attached to a Hilbert modular form replaced by the (p -adic) periods on $\mathcal{H}_p \times \mathcal{H}$ attached to f , leads to the definition of a “modular parametrisation” analogous to (19)

$$\Phi_E^p : \Gamma \backslash \mathcal{H}'_p \longrightarrow E(\mathbb{C}_p). \tag{22}$$

Let D be the discriminant of K , and choose a $\delta \in \mathbb{Z}[1/p]$ satisfying

$$\delta^2 \equiv D \pmod{M}.$$

Let \mathcal{F}^D be the set of primitive binary quadratic forms $Ax^2 + Bxy + Cy^2$ with coefficients in $\mathbb{Z}[1/p]$, satisfying

$$B^2 - 4AC = D, \quad M|A, \quad B \equiv \delta \pmod{M}.$$

(A quadratic form is said to be *primitive* in this context if the ideal of $\mathbb{Z}[1/p]$ generated by (A, B, C) is equal to $\mathbb{Z}[1/p]$.) The group Γ acts naturally on \mathcal{F}^D by “change of variables”, and the quotient $\Gamma \backslash \mathcal{F}^D$ is equipped with a natural simply transitive action of the class group G_D of K arising from the Gaussian composition law. (Or rather, the Picard group of $\mathcal{O}_K[1/p]$, but these coincide since p is inert in K .) This action is completely analogous to the action of G_D on $\Gamma_0(N0 \backslash \mathcal{H}^D$ (for D a negative discriminant) that is described in Section 2.2. Choose an embedding of K into \mathbb{C}_p , and for each quadratic form $F = [A, B, C] \in \mathcal{F}^D$, let

$$\tau = \frac{-B + \sqrt{D}}{2A} \in \mathcal{H}_p \tag{23}$$

be the corresponding element of \mathcal{H}_p satisfying $F(\tau, 1) = 0$. The set \mathcal{H}_p^D of all τ that arise in this way is preserved under the action of Γ , and the natural assignment given by (23) induces a bijection

$$\Gamma \backslash \mathcal{F}^D \longrightarrow \Gamma \backslash \mathcal{H}_p^D.$$

Hence the target of this bijection inherits a simply transitive action of G_D . Denote this action by $(\sigma, \tau) \mapsto \tau^\sigma$, for all $\sigma \in G_D$ and $\tau \in \mathcal{H}_p^D$. Conjectures 5.9 and 5.15 of [Da1] predict that

Conjecture 4.1. 1. For all $\tau \in \mathcal{H}_p^D$, the point $\Phi_E^p(\tau)$ is defined over H .

2. If $\chi : G_D \longrightarrow \mathbb{C}^\times$ is a complex character, then the expression

$$\sum_{\sigma \in G_D} \chi(\sigma) \Phi_E^p(\tau^\sigma) \in E(H) \otimes \mathbb{C}$$

is non-zero if and only if $L'(E/K, \chi, 1) \neq 0$. In particular, the subgroup of $E(H)$ generated by the Stark–Heegner points $\Phi_E^p(\tau)$, as $\tau \in \Gamma \backslash \mathcal{H}_p^D$, has rank $h = [H : K]$ if and only if $\text{ord}_{s=1} L(E/H, s) = h$.

A proof of Conjecture 4.1 would not yield any new information about Conjecture BSD_r for $r \leq 1$, since this conjecture is already known for elliptic curves over \mathbb{Q} . It would, however, give a proof of some new cases of the Birch and Swinnerton-Dyer conjecture for Mordell–Weil groups over ring class fields of real quadratic fields, following a simple extension of Kolyvagin’s arguments which is explained in [BD1] and in Chapter X of [Da2]. See also [BDD] for other ways in which a strengthening of

Conjecture 4.1 to modular forms with non-rational Fourier coefficients would imply new cases of Conjecture BSD_0 , by adapting the ideas that are used in the proof of Theorem 3.7.

Conjecture 4.1 has been extensively tested numerically in [DG]. A significant improvement of the algorithms of [DG], based on ideas of Pollack and Stevens which grew out of their theory of overconvergent modular symbols, as mentioned in Section 3.3, is described in [DP1]. These improvements make it possible to find global points of large height on E rather efficiently. For example, the Stark–Heegner point on the elliptic curve E of conductor 11 given by the equation

$$y^2 + y = x^3 - x^2 - 10x - 20$$

attached to the field $K = \mathbb{Q}(\sqrt{101})$ can be computed to an 11-adic accuracy of 11^{-100} in a few seconds on a standard computer. It can then be “recognized” as the global point in $E(\mathbb{Q}(\sqrt{101}))$ with x -coordinate equal to

$$x = \frac{1081624136644692539667084685116849}{246846541822770321447579971520100}.$$

Of course, the calculation of Stark–Heegner points also has applications to explicit class field theory for real quadratic fields analogous to those described in Section 2.4 for imaginary quadratic fields. For example, let E be the unique elliptic curve over \mathbb{Q} of conductor $p = 79$, defined by the Weierstrass equation

$$y^2 + xy + y = x^3 + x^2 - 2x.$$

The prime p is inert in the real quadratic field $K = \mathbb{Q}(\sqrt{401})$, which has class number five. The 5 distinct representatives in \mathcal{H}_79^{401} can be chosen to be

$$\tau = \frac{-19 + \sqrt{401}}{2}, \quad \frac{19 - \sqrt{401}}{4}, \quad \frac{-15 + \sqrt{401}}{8}, \quad \frac{17 - \sqrt{401}}{8}, \quad \frac{-17 + \sqrt{401}}{4}.$$

The x -coordinates of the corresponding Stark–Heegner points $\Phi_E^{79}(\tau)$ (computed modulo 79^{20}) appear to satisfy the polynomial

$$x^5 - 20x^4 + 47x^3 - 31x^2 + x + 3$$

whose splitting field is indeed the Hilbert class field H of K . In fact this calculation leads to the discovery of points in $E(H)$. The analogous polynomial, for the real quadratic field $\mathbb{Q}(\sqrt{577})$ of class number seven, is

$$x^7 - 22x^6 + 74x^5 - 51x^4 - 40x^3 + 32x^2 + 2x - 1.$$

These examples are chosen at random among the hundreds of calculations that were performed in [DP1] to test the conjectures of [Da1] numerically. (More such calculations could be performed by the interested reader using the publicly available software [DP2] for calculating Stark–Heegner points, written in Magma, which is documented in [DP1].)

4.3. Beyond totally real fields? The assumption that E is defined over a totally real field F , although it arises naturally in considering automorphic forms and their associated Shimura varieties, is not a natural one from the point of view of the Diophantine study of elliptic curves. It would be just as desirable to understand elliptic curves defined over general number fields, and to have the means of tackling conjectures BSD_0 and BSD_1 for such curves.

The simplest case arises when E is an elliptic curve defined over an *imaginary* quadratic field, denoted by F (and not K as in Section 2, since now it plays the role of the “ground field” over which E is defined). Assume for simplicity that F has class number one, and let \mathcal{N} denote the conductor of E .

As in Section 3, the Shimura–Taniyama conjecture predicts that E corresponds to an automorphic form f on $\text{GL}_2(F)$, which gives rise, following the description given in [Cr1], to a differential form ω_f on the *hyperbolic upper half space*

$$\mathcal{H}^{(3)} := \mathbb{C} \times \mathbb{R}^{>0}.$$

This three-dimensional real manifold is equipped with a hyperbolic metric and an action of $\text{SL}_2(\mathbb{C})$ by isometries, and the differential ω_f is *invariant* under the resulting action of the subgroup $\Gamma_0(\mathcal{N}) \subset \text{SL}_2(\mathcal{O}_F)$ consisting of matrices which are upper triangular modulo \mathcal{N} . Congruence subgroups of $\text{SL}_2(\mathcal{O}_F)$ are examples of so-called *Bianchi groups*; for information about their structure and properties and further references, see [EGM] for example.

The modular form ω_f does not give rise to a modular parametrisation of E analogous to (4). In fact, the symmetric space $\mathcal{H}^{(3)}$ is not even endowed with a natural complex structure (since it has real dimension 3); therefore the quotient $\Gamma_0(\mathcal{N}) \backslash \mathcal{H}^{(3)}$ cannot be viewed as the points of a complex analytic variety, much less an algebraic one. The absence of a Shimura variety attached to f implies that one has less control on the arithmetic of this modular form. For certain f , Taylor [Ta] has been able to construct the Galois representations which the Langlands conjectures attach to f by exploiting congruences with modular forms on $\text{GSp}(4)$ whose associated Galois representations can be found in the cohomology of the appropriate Shimura varieties. Unfortunately, global points on elliptic curves or abelian varieties, unlike p -adic Galois representations (as in the work of Taylor) or Galois cohomology classes attached to rational points (as in the proof of Theorem 3.7), do not readily lend themselves to constructions based on congruences between modular forms.

Nonetheless, the differential form ω_f comes with an attendant notion of *modular symbol* which enjoys the same integrality properties as in the classical case. (For a discussion of modular symbols attached to forms on $\text{GL}_2(F)$, and their computational applications, see [Cr1], [CW].) Trifkovic [Tr] exploits this modular symbol to transpose to f the definition of the p -adic periods on $\mathcal{H}_p \times \mathcal{H}$ alluded to in Section 4.2. In this way he associates to ω_f a “modular form on $\Gamma \backslash (\mathcal{H}_p \times \mathcal{H}^{(3)})$ ”, where

1. p is a prime of K dividing $\mathcal{N} = \mathcal{M}p$ exactly;

2. $\mathcal{H}_{\mathfrak{p}} = \mathbb{P}_1(\mathbb{C}_{\mathfrak{p}}) - \mathbb{P}_1(F_{\mathfrak{p}})$ is the \mathfrak{p} -adic upper half plane (defined after choosing an embedding $F_{\mathfrak{p}} \subset \mathbb{C}_{\mathfrak{p}}$);
3. $\Gamma \subset \mathrm{SL}_2(\mathcal{O}_F[1/\mathfrak{p}])$ is the subgroup consisting of matrices which are upper triangular modulo \mathcal{M} .

The set $\mathcal{H}'_{\mathfrak{p}}$ is simply the set of $\tau \in \mathcal{H}_{\mathfrak{p}}$ which generate a quadratic extension of $F \subset F_{\mathfrak{p}}$. Trifkovic uses his \mathfrak{p} -adic periods to define an explicit, numerically computable map

$$\Phi_E^{\mathfrak{p}} : \Gamma \backslash \mathcal{H}'_{\mathfrak{p}} \longrightarrow E(\mathbb{C}_{\mathfrak{p}}),$$

and formulates an analogue of Conjecture 4.1 for this map, predicting that $\Phi_E^{\mathfrak{p}}(\tau)$ is defined over a specific ring class field of $K = F(\tau)$ for all $\tau \in \mathcal{H}'_{\mathfrak{p}}$.

Trifkovic has been able to gather extensive numerical evidence for his “Stark–Heegner conjectures” in this setting. Here is just one example taken among the many calculations that are reported on in [Tr]. Let E be the elliptic curve over $F = \mathbb{Q}(\sqrt{-11})$ given by the Weierstrass equation

$$y^2 + y = x^3 + \left(\frac{1 - \sqrt{-11}}{2}\right)x^2 - x.$$

Its conductor is the prime $\mathfrak{p} = 6 + \sqrt{-11}$ of F of norm 47. Note that E is not isogenous to its Galois conjugate, since $\mathfrak{p} \neq \bar{\mathfrak{p}}$. The quadratic extension $K = F(\sqrt{29})$ has class number 5. Trifkovic computes the five distinct Stark–Heegner points attached to the maximal order of K , as elements of $E(\mathbb{Q}_{47})$ (using the isomorphism $K_{\mathfrak{p}} = \mathbb{Q}_{47}$) with an accuracy of 20 significant 47-adic digits. This allows him to “guess” that the x coordinates of these Stark–Heegner points satisfy the degree 5 polynomial

$$\begin{aligned} x^5 - & \left(\frac{80299 + 139763\sqrt{-11}}{149058}\right)x^4 + \left(\frac{-558203 + 71567\sqrt{-11}}{149058}\right)x^3 \\ & + \left(\frac{141709 + 45575\sqrt{-11}}{74529}\right)x^2 + \left(\frac{8372 - 7727\sqrt{-11}}{24843}\right)x + \left(\frac{-473 + 35\sqrt{-11}}{2366}\right) \end{aligned}$$

whose splitting field can then be checked to be the Hilbert class field of K .

For many more calculations of this type, and a precise statement of the conjecture on which they rest, the reader is invited to consult [Tr].

4.4. Theoretical evidence. In spite of the convincing numerical evidence that has been gathered in their support, the conjectures on Stark–Heegner points suffer from the same paucity of theoretical evidence as in the setting of Stark’s original conjectures on units. What little evidence there is at present can be grouped roughly under the following two rubrics:

4.4.1. Stark–Heegner points and Stark units. Many of the basic theorems and applications of elliptic curves have counterparts for units of number fields. (For instance, the Mordell–Weil theorem is analogous to Dirichlet’s Unit Theorem; Lenstra’s factorisation algorithm based on elliptic curves, to the Pollard $p - 1$ method; to name just two examples.) The very terminology “Stark–Heegner point” is intended to convey the idea that these points are analogous to Stark units constructed from special values of L -series.

To make this sentiment precise, one can replace the cusp forms that enter into the constructions of Section 4.2 by *modular units*, or rather, their logarithmic derivatives which are Eisenstein series of weight two. Pursuing this idea, the article [DD] associates to any modular unit α on $\Gamma_0(N)\backslash\mathcal{H}$ and to $\tau \in \mathcal{H}'_p \cap K$ where K is a real quadratic field in which p is inert, an element $u(\alpha, \tau) \in K_p^\times$, which is predicted to behave like an elliptic unit defined over a ring class field of an imaginary quadratic field. More precisely, if \mathcal{O} is the order associated to τ , and H denotes the corresponding ring class field of K , it is conjectured that $u(\alpha, \tau)$ belongs to $\mathcal{O}_H[1/p]^\times$ and obeys a Shimura reciprocity law formulated exactly as in Conjecture 4.1.

Section 3.1 [DD] attaches to α and to τ a ζ -function $\zeta(\alpha, \tau, s)$ which is essentially (up to a finite collection of Euler factors depending on α) the partial zeta-function attached to K and the narrow ideal class corresponding to τ . In particular, this zeta-function has a meromorphic continuation to \mathbb{C} with at worst a simple pole at $s = 1$. Sections 4.1–4.3 of [DD] explain how a p -adic zeta function $\zeta_p(\alpha, \tau, s)$ can be defined by p -adically interpolating the special values of $\zeta(\alpha, \tau, k)$ at certain negative integers.

The main result of [DD], which is contained in Theorems 3.1 and 4.1 of [DD], is then

Theorem 4.2. For all $\tau \in \mathcal{H}'_p$,

$$\zeta(\alpha, \tau, 0) = \frac{1}{12} \operatorname{ord}_p(u(\alpha, \tau)); \quad (24)$$

$$\zeta'_p(\alpha, \tau, 0) = -\frac{1}{12} \log_p \operatorname{Norm}_{K_p/\mathbb{Q}_p}(u(\alpha, \tau)). \quad (25)$$

This theorem is consistent with Gross’s p -adic analogue of the Stark conjectures [Gr1], [Gr2], which expresses the left hand side of (25) in terms of p -adic logarithms of the norm to \mathbb{Q}_p certain global p -units in abelian extensions of K . We note that the conjecture of [DD] represents a genuine *refinement* of Gross’s conjecture in the special case of ring class fields of real quadratic fields, since it gives a formula for the Gross–Stark units *as elements of* K_p^\times , and not just for their norms to \mathbb{Q}_p^\times .

Remark 4.3. A purely archimedean analogue of the setting of Theorem 4.2 is considered in [CD], leading to the conjectural construction of units in abelian extensions of an ATR extension K of a totally real field F in terms of periods of weight two Eisenstein series on the Hilbert modular group attached to F . This construction can be viewed either as the archimedean analogue of the main construction of [DD], or

as the analogue of the main construction of [DL] in which cusp forms on the Hilbert modular group are replaced by Eisenstein series. This construction (in the setting of abelian extensions of K) goes further than the original Stark conjectures by proposing a formula for the Stark units as elements of \mathbb{C}^\times , and not just for their *lengths* which are expressed in terms of values of L -series at $s = 0$. In other words, the formulae of [CD] capture the *arguments* as well as the absolute values of these Stark units (relative to a complex embedding of the ring class field H of K extending the unique complex embedding of K .)

Remark 4.4. The proof of Theorem 4.2 brings to light the role of the Eisenstein series of weight k and their associated periods (with k a weight which can be taken to vary p -adically) in relating the invariants $u(\alpha, \tau)$ to special values of L -functions. This suggests that the Stark–Heegner points of Section 4.2 should be related to the periods of a *Hida family* interpolating the cuspidal eigenform f in weight two.

4.4.2. The rationality of Stark–Heegner points over genus fields. Returning to the setting of Section 4.2, let K be a real quadratic field of discriminant D satisfying the auxiliary hypotheses relative to N that were mentioned in Section 4.2, and let H be its Hilbert class field. Write $G_D = \text{Gal}(H/K)$ as before.

To each factorisation $D = D_1 D_2$ of D as a product of two fundamental discriminants is associated the unramified quadratic extension $L = \mathbb{Q}(\sqrt{D_1}, \sqrt{D_2}) \subset H$ of K . This field corresponds to a quadratic character

$$\chi : G_D \longrightarrow \pm 1,$$

called the *genus character* associated to the factorisation (D_1, D_2) . Let χ_1 and χ_2 denote the quadratic Dirichlet characters attached to $\mathbb{Q}(\sqrt{D_1})$ and $\mathbb{Q}(\sqrt{D_2})$ respectively. Then χ , viewed as a character of the ideals of K , is characterised on ideals prime to D by the rule

$$\chi(\mathfrak{n}) = \chi_1(\text{Norm } \mathfrak{n}) = \chi_2(\text{Norm } \mathfrak{n}).$$

The field L is also called the *genus field* of K attached to (D_1, D_2) . Let $E(L)^\chi$ denote the submodule of the Mordell–Weil group $E(L)$ on which G_D acts via the character χ .

Recall the action of G_D on $\Gamma \backslash \mathcal{H}_p^D$ arising from its identification with the class group of K . Define the point

$$P_\chi = \sum_{\sigma \in G_D} \chi(\sigma) \Phi_E^p(\tau^\sigma) \in E(K_p).$$

Conjecture 4.1 predicts that this local point belongs to $E(L)^\chi$ (after fixing an embedding $L \subset K_p$), and that it is of infinite order if and only if

$$L(E/K, \chi, s) = L(E, \chi_1, s)L(E, \chi_2, s)$$

has a simple zero at $s = 1$.

For each $m|N$ with $\gcd(m, N/m) = 1$, let w_m denote the sign of the Fricke involution at m acting on f . Note that the modified Heegner hypothesis implies that $\chi_1(-M) = \chi_2(-M)$. The main result of [BD5] is

Theorem 4.5. *Suppose that E has at least two primes of multiplicative reduction, and that $\chi_1(-M) = -w_M$.*

1. *There is a global point $P_\chi \in E(L)^\times$ and $t \in \mathbb{Q}^\times$ such that*

$$P_\chi = tP_\chi \quad \text{in } E(K_p) \otimes \mathbb{Q}. \quad (26)$$

2. *The point P_χ is of infinite order if and only if $L'(E/K, \chi, 1) \neq 0$.*

The proof of Theorem 4.5 relies on the connection between Stark–Heegner points and Shintani-type periods attached to Hida families alluded to in Remark 4.4, which grew out of the calculations of [DD]. A second key ingredient is the relation, made precise in [BD2] and [BD4], between classical Heegner points arising from certain Shimura curve parametrisations and the derivatives of associated two-variable anti-cyclotomic p -adic L -functions attached to Hida families. In a nutshell, these two ingredients are combined to express the Stark–Heegner point P_χ as a classical Heegner point, following an idea whose origins (as is explained in the introduction of [BD5]) can be traced back to Kronecker’s “solution of Pell’s equation” in terms of special values of the Dedekind eta-function.

Remark 4.6. A result of Gross–Kohnen–Zagier [GKZ] suggests that the position of the Stark–Heegner point P_χ in the Mordell–Weil group $E(L)^\times$ is controlled by the Fourier coefficients of a modular form of weight $3/2$ associated to f via the Shimura lift. See [DT] where results of this type are discussed.

References

- [BC] Boutot, J.-F., Carayol, H., Uniformisation p -adique des courbes de Shimura: les théorèmes de Čerednik et de Drinfeld. in *Courbes modulaires et courbes de Shimura* (Orsay, 1987/1988), *Astérisque* **196–197** (1991), 45–158.
- [BCDT] Breuil, C., Conrad, B., Diamond, F., Taylor, R., On the modularity of elliptic curves over \mathbb{Q} : wild 3-adic exercises. *J. Amer. Math. Soc.* **14** (2001), 843–939.
- [BCdeSGKK] Bump, D., Cogdell, J. W., de Shalit, E., Gaitsgory, D., Kowalski, E., Kudla, S. S., *An introduction to the Langlands program*. Lectures presented at the Hebrew University of Jerusalem (ed. by J. Bernstein and S. Gelbart), Birkhäuser, Boston, MA, 2003.
- [BD1] Bertolini, M., Darmon, H., Kolyvagin’s descent and Mordell–Weil groups over ring class fields. *J. Reine Angew. Math.* **412** (1990), 63–74.
- [BD2] Bertolini, M., Darmon, H., Heegner points, p -adic L -functions, and the Čerednik–Drinfeld uniformization. *Invent. Math.* **131** (3) (1998), 453–491.

- [BD3] Bertolini, M., Darmon, H., Iwasawa’s Main Conjecture for Elliptic Curves over Anticyclotomic \mathbb{Z}_p -extensions, *Ann. of Math.* **162** (2005), 1–64.
- [BD4] Bertolini, M., Darmon, H., Hida families and rational points on elliptic curves. Submitted.
- [BD5] Bertolini, M., Darmon, H., The rationality of Stark–Heegner points over genus fields of real quadratic fields. Submitted.
- [BDD] Bertolini, M., Darmon, H., Dasgupta, S., Stark–Heegner points and special values of L -series. In *Proceedings of the Durham Symposium on L -functions and Galois representations* (July 2004), to appear.
- [BDG] Bertolini, M., Darmon, H., Green, P., Periods and points attached to quadratic algebras. In *Heegner points and Rankin L -series* (ed. by H. Darmon and S. Zhang), Math. Sci. Res. Inst. Publ. 49, Cambridge University Press, Cambridge 2004, 323–367.
- [Bi] Birch, B., Heegner points: the beginnings. In *Heegner points and Rankin L -series* (ed. by H. Darmon and S. Zhang), Math. Sci. Res. Inst. Publ. 49, Cambridge University Press, Cambridge 2004, 1–10.
- [BS] Birch, B., Stephens, N., Computation of Heegner points. In *Modular forms* (Durham, 1983), Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., Horwood, Chichester 1984, 13–41.
- [CD] Charollois, P., Darmon, H., Périodes des séries d’Eisenstein et arguments des unités de Stark. In progress.
- [Cr1] Cremona, J. E., Hyperbolic tessellations, modular symbols, and elliptic curves over complex quadratic fields. *Compositio Math.* **51** (3) (1984), 275–324.
- [Cr2] Cremona, J. E., *Algorithms for modular elliptic curves*. Second edition. Cambridge University Press, Cambridge 1997.
- [CW] Cremona, J. E., Whitley, E. Periods of cusp forms and elliptic curves over imaginary quadratic fields. *Math. Comp.* **62** (205) (1994), 407–429.
- [Da1] Darmon, H., Integration on $\mathcal{H}_p \times \mathcal{H}$ and arithmetic applications. *Ann. of Math.* (2) **154** (3) (2001), 589–639.
- [Da2] Darmon, H., *Rational points on modular elliptic curves*. CBMS Reg. Conf. Ser. Math. 101, Amer. Math. Soc., Providence, RI, 2004.
- [DD] Darmon, H., Dasgupta, S., Elliptic units for real quadratic fields. *Ann. of Math.* **163** (2006), 301–346.
- [DG] Darmon, H., Green, P., Elliptic curves and class fields of real quadratic fields: algorithms and evidence. *Experiment. Math.* **11** (1) (2002), 37–55.
- [DL] Darmon, H., Logan, A., Periods of Hilbert modular forms and rational points on elliptic curves. *Internat. Math. Res. Notices* **40** (2003), 2153–2180.
- [DP1] Darmon, H., Pollack, R., Efficient calculation of Stark–Heegner points via over-convergent modular symbols. *Israel J. Math.* **153** (2006), 319–354.
- [DP2] Darmon, H., Pollack, R., The `shp` package, Software written in Magma. Downloadable from <http://www.math.mcgill.ca/darmon/programs/shp/shp.html>.
- [DT] Darmon, H., Tornaria, G., A Gross–Kohnen Zagier theorem for Stark–Heegner points. In progress.

- [E11] Elkies, N. D., Shimura curve computations. In *Algorithmic number theory* (Portland, OR, 1998), Lecture Notes in Comput. Sci. 1423, Springer-Verlag, Berlin 1998, 1–47.
- [E12] Elkies, N. D., Heegner point computations. In *Algorithmic number theory* (Ithaca, NY, 1994) Lecture Notes in Comput. Sci. 877, Springer-Verlag, Berlin 1994, 122–133.
- [EGM] Elstrodt, J., Grunewald, F., Mennicke, J., *Groups acting on hyperbolic space*. Springer Monogr. Math., Springer-Verlag, Berlin 1998.
- [Fu] Fujiwara, K., Modular varieties and Iwasawa theory. In *Algebraic number theory and related topics* (Kyoto, 1996), Sūrikaiseikikenkyūsho Kūokyūroku (RIMS, Kyoto) **998**, (1997), 1–19.
- [Ge] Gelbart, S. S., *Automorphic forms on adèle groups*. Ann. of Math. Stud. 83, Princeton University Press/University of Tokyo Press, Princeton, N.J./Tokyo 1975.
- [GKZ] Gross, B., Kohlen, W., Zagier, D., Heegner points and derivatives of L -series. II. *Math. Ann.* **278** (1–4) (1987), 497–562.
- [Go] Goldfeld, D., The Gauss class number problem for imaginary quadratic fields. In *Heegner points and Rankin L -series* (ed. by H. Darmon and S. Zhang), Math. Sci. Res. Inst. Publ. 49, Cambridge University Press, Cambridge, 2004, 25–36.
- [Gre] Greenberg, M., Heegner points and rigid analytic modular forms. PhD thesis, McGill University, 2006.
- [Gr1] Gross, B. H., p -adic L -series at $s = 0$. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **28** (3) (1981), 979–994.
- [Gr2] Gross, B. H., On the values of abelian L -functions at $s = 0$. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **35** (1) (1988), 177–197.
- [Gr3] Gross, B. H., Kolyvagin’s work on modular elliptic curves. In *L -functions and arithmetic* (Durham, 1989), London Math. Soc. Lecture Note Ser. 153, Cambridge University Press, Cambridge 1991, 235–256.
- [Gr4] Gross, B. H., Heegner points and representation theory. In *Heegner points and Rankin L -series* (ed. by H. Darmon and S. Zhang), Math. Sci. Res. Inst. Publ. 49, Cambridge University Press, Cambridge 2004, 37–65.
- [GZ] Gross, B. H., Zagier, D. B., Heegner points and derivatives of L -series. *Invent. Math.* **84** (2) (1986), 225–320.
- [Ki] Kisin, M., Geometric deformations of modular Galois representations. *Invent. Math.* **157** (2) (2004), 275–328.
- [KL] Kolyvagin, V. A., Logachev, D. Yu., Finiteness of III over totally real fields. *Izv. Akad. Nauk SSSR Ser. Mat.* **55** (4) (1991), 851–876; English transl. *Math. USSR-Izv.* **39** (1) (1992), 829–853.
- [KM] Khuri-Makdisi, K., Moduli interpretation of Eisenstein series. In progress.
- [Ko] Kolyvagin, V. A., Finiteness of $E(\mathbb{Q})$ and $III(E, \mathbb{Q})$ for a subclass of Weil curves. *Izv. Akad. Nauk SSSR Ser. Mat.* **52** (3) (1988), 522–540, 670–671; English transl. *Math. USSR-Izv.* **32** (3) (1989), 523–541.
- [Le] Lenstra, H. W., Solving the Pell Equation. *Notices Amer. Math. Soc.* **49** (2002), 182–192.

- [Lo1] Longo, M., On the Birch and Swinnerton-Dyer conjecture over totally real fields. PhD thesis, University of Padova, 2004.
- [Lo2] Longo, M., On the Birch and Swinnerton-Dyer conjecture for modular elliptic curves over totally real fields. *Ann. Inst. Fourier*, to appear.
- [Ma] Mazur, B., Modular curves and the Eisenstein ideal. *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186.
- [Men] Mennicke, J., On Ihara’s modular group. *Invent. Math.* **4** (1967), 202–228.
- [Mer] Merel, L., Bornes pour la torsion des courbes elliptiques sur les corps de nombres. *Invent. Math.* **124** (1–3) (1996), 437–449.
- [MM] Murty, M. R., Murty, V. K., *Non-vanishing of L -functions and applications*. Progr. Math. 157, Birkhäuser, Basel 1997.
- [PS] Pollack, R., Stevens, G., Explicit computations with overconvergent modular symbols. In preparation.
- [RV] Ricotta, G., Vidick, T., Hauteur asymptotique des points de Heegner. *Canad. J. Math.*, to appear.
- [Se] Serre, J.-P., Complex multiplication. In *Algebraic Number Theory* (Brighton, 1965), Thompson, Washington D.C. 1967, 292–296.
- [SW] Skinner, C. M., Wiles, A. J., Residually reducible representations and modular forms. *Inst. Hautes Études Sci. Publ. Math.* No. **89** (1999), 5–126.
- [Ta] Taylor, R., l -adic representations associated to modular forms over imaginary quadratic fields. II. *Invent. Math.* **116** (1–3) (1994), 619–643.
- [Tr] Trifkovic, M., Stark-Heegner points on elliptic curves over imaginary quadratic fields. Submitted.
- [TZ] Tian, Y., Zhang, S., Book project in progress.
- [Zh1] Zhang, S., Heights of Heegner points on Shimura curves. *Ann. of Math. (2)* **153** (2001), 27–147.
- [Zh2] Zhang, S., Gross-Zagier formula for GL_2 . *Asian J. Math.* **5** (2) (2001), 183–290.
- [Zh3] Zhang, S., Gross-Zagier formula for $GL(2)$. II. In *Heegner points and Rankin L -series* (ed. by H. Darmon and S. Zhang), Math. Sci. Res. Inst. Publ. 49, Cambridge University Press, Cambridge 2004, 191–214.

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West,
Montreal, QC, Canada

E-mail: darmon@math.mcgill.ca