THE BIRCH–SWINNERTON-DYER CONJECTURE AND HEEGNER POINTS: A SURVEY

WEI ZHANG

Contents

1. The Birch–Swinnerton-Dyer conjecture

For a (connected) smooth projective curve C over the rational numbers $\mathbb{Q}$, it is known that the rational points $C(\mathbb{Q})$ depends on the genus $g = g(C)$ of C:

(1) If $g = 0$, then the local-global principle holds for C, i.e.: $C(\mathbb{Q}) \neq \varnothing$ if and only if $C(\mathbb{Q}_p) \neq \varnothing$ for all primes $p \leqslant \infty$ (we understand $\mathbb{Q}_p = \mathbb{R}$ when $p = \infty$). In other words, C is globally solvable if and only if it is locally solvable everywhere. Another way of stating this is: $C \simeq \mathbb{P}^1_{\mathbb{Q}}$ if and only if $C_{\mathbb{Q}_p} \simeq \mathbb{P}^1_{\mathbb{Q}_p}$ for all primes $p \leqslant \infty$. We see that $C(\mathbb{Q})$ is either an empty set or an infinite set.
(2) If $g = 1$, $C(\mathbb{Q})$ may be empty, finite or infinite. This article will focus on this case.
(3) If $g \geqslant 2$, Faltings theorem asserts that $C(\mathbb{Q})$ is always finite.

1.1. **Mordell–Weil group and Tate–Shafarevich group.** For a genus-one curve $C/\mathbb{Q}$, its Jacobian variety $\mathrm{Jac}(C)$ is an "elliptic curve", i.e., a genus-one curve with a distinguished rational point O. We write E for an elliptic curve defined over $\mathbb{Q}$. A theorem of Mordell in 1922 asserts that the abelian group $E(\mathbb{Q})$ is finitely generated. This result was later generalized by Weil to abelian varieties over number fields. Now we call the abelian group $E(\mathbb{Q})$ the Mordell–Weil group of $E/\mathbb{Q}$ and write:

$$E(\mathbb{Q}) \simeq \mathbb{Z}^{r_{MW}} \oplus \text{ finite group},$$

where the integer $r_{MW} = r_{MW}(E/\mathbb{Q})$ is called the Mordell–Weil rank of $E/\mathbb{Q}$.

The failure of the local-global principle for a genus-one curve over $\mathbb{Q}$ is related to the notion of Tate–Shafarevich group of its Jacobian. The Tate–Shafarevich group of an elliptic curve $E/\mathbb{Q}$, denoted by $Ш(E/\mathbb{Q})$, is closely related to the set of isomorphism classes of smooth projective curves $C/\mathbb{Q}$ such that

$$\mathrm{Jac}(C) \simeq E, \quad C(\mathbb{Q}_p) \neq \varnothing, \text{ for all primes } p \leqslant \infty.$$

The more precise definition is to use Galois cohomology

$$Ш(E/\mathbb{Q}) := \mathrm{Ker}(H^1(\mathbb{Q}, E) \to \prod_v H^1(\mathbb{Q}_v, E)),$$

where the map is the product of the localization at all places v of $\mathbb{Q}$ (including the archimedean one). As usual, we denote the Galois cohomology $H^i(k, E) := H^i(\mathrm{Gal}(\overline{k}/k), E)$ for $k = \mathbb{Q}, \mathbb{Q}_p$ and $i \in \mathbb{Z}_{\geqslant 0}$. The first cohomology group $H^1(k, E)$ is called the Weil–Châtelet group. It is the abelian group of principal homogeneous spaces for E over k. The group $H^1(\mathbb{Q}, E)$ is torsion and abelian, hence so is the group $Ш(E/\mathbb{Q})$. Shafarevich and Tate independently made the following fundamental conjecture ([41],[46])

Conjecture 1.1. *Let $E/\mathbb{Q}$ be an elliptic curve. Then the Tate–Shafarevich group $Ш(E/\mathbb{Q})$ is finite.*

Remark 1. One famous example of elliptic curve with nontrivial Ш was discovered by Selmer:

$$x^3 + y^3 + 60z^3 = 0 \subset \mathbb{P}^2_{\mathbb{Q}}.$$

This is the Jacobian of the Selmer curve:

$$3x^3 + 4y^3 + 5z^3 = 0 \subset \mathbb{P}^2_{\mathbb{Q}},$$

which is locally solvable everywhere but does not have a $\mathbb{Q}$-point.

Remark 2. The order of Ш can be arbitrarily large. Cassels proved that in the following family of elliptic curves (with complex multiplication by $\mathbb{Z}[\sqrt{-3}]$):

$$E_n : x^3 + y^3 + nz^3 = 0,$$

the 3-torsion of $Ш(E_n/\mathbb{Q})$ is unbounded. To the author's knowledge, it is still unknown whether for each prime p, there is an elliptic curve $E/\mathbb{Q}$ whose Tate–Shafarevich group contains an element of order p.

Remark 3. There is an alternating pairing, called the "Cassels–Tate pairing"

$$Ш(E/\mathbb{Q}) \times Ш(E/\mathbb{Q}) \to \mathbb{Q}/\mathbb{Z}$$

whose kernel is precisely the divisible subgroup of $Ш(E/\mathbb{Q})$. Assuming the finiteness of Ш, this gives a non-degenerate alternating pairing on Ш and it follows that the elementary divisors of Ш show up by pairs. In particular, the order of Ш is a square.

1.2. **Selmer group.** As a motivation for defining the Selmer group, we recall the Hodge conjecture for a smooth projective algebraic variety $X/\mathbb{C}$ and a fixed $i \in \mathbb{Z}_{\geqslant 0}$. The Betti cohomology of even degree $H^{2i}_B(X(\mathbb{C}), \mathbb{Z})$ is a finitely generated abelian group. Hodge theory allows us to define the subgroup, denoted by $\mathrm{Hg}(X_{\mathbb{C}}, \mathbb{Z})$, of integral Hodge classes in $H^{2i}_B(X(\mathbb{C}), \mathbb{Z})$. This provides a "cohomological description" of algebraic cycle classes in $H^{2i}_B(X(\mathbb{C}), \mathbb{Z})$. Denote by $\mathrm{Alg}(X_{\mathbb{C}}, \mathbb{Z})$ the subgroup of algebraic cycle classes in $H^{2i}_B(X(\mathbb{C}), \mathbb{Z})$. Then the Hodge conjecture is equivalent to the statement that the quotient

$$\text{“}Ш^i(X/\mathbb{C})\text{”} := \frac{\mathrm{Hg}(X_{\mathbb{C}}, \mathbb{Z})}{\mathrm{Alg}(X_{\mathbb{C}}, \mathbb{Z})}$$

is a finite group. We may write this as an exact sequence

$$0 \longrightarrow \mathrm{Alg}(X_{\mathbb{C}}, \mathbb{Z}) \longrightarrow \mathrm{Hg}(X_{\mathbb{C}}, \mathbb{Z}) \longrightarrow Ш^i(X/\mathbb{C}) \longrightarrow 0\ . \tag{1.1}$$

There is also an analogous short exact sequence which gives a cohomological description of rational points on an elliptic curve E over a number field F:

$$0 \longrightarrow E(F)\otimes \mathbb{Q}_p/\mathbb{Z}_p \longrightarrow \mathrm{Sel}_{p^\infty}(E/F) \longrightarrow Ш(E/F)[p^\infty] \longrightarrow 0\ . \tag{1.2}$$

Here p is a prime number, $Ш(E/F)[p^\infty]$ is the p-primary part of Tate–Shafarevich group, and $\mathrm{Sel}_{p^\infty}(E/F)$ is the p^∞-Selmer group defined as follows. Let $E[p^\infty]$ be the group of p-primary torsion points of $E(\overline{\mathbb{Q}})$. The absolute Galois group Gal_F acts on $E[p^\infty]$, which is isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^2$ as an abstract group. Consider the local Kummer map

$$\delta_v : E(F_v)\otimes \mathbb{Q}_p/\mathbb{Z}_p \to H^1(F_v, E[p^\infty]).$$

Then $\mathrm{Sel}_{p^\infty}(E/F)$ is defined as

$$\mathrm{Sel}_{p^\infty}(E/F) := \mathrm{Ker}(H^1(F, E[p^\infty]) \to \prod_v H^1(F_v, E[p^\infty])/\mathrm{Im}(\delta_v)),$$

where the map is the product of the localization at all places v of F. The $\mathbb{Z}_p$-corank of $\mathrm{Sel}_{p^\infty}(E/F)$ is denote by $r_p(E/F)$. As an abstract abelian group, $\mathrm{Sel}_{p^\infty}(E/F)$ is of the form

$$(\mathbb{Q}_p/\mathbb{Z}_p)^{r_p(E/F)} \oplus \text{ a finite group.}$$

We have an inequality

$$0 \leqslant r_{MW}(E/F) \leqslant r_p(E/F) < \infty, \tag{1.3}$$

where the equality $r_{MW}(E/F) = r_p(E/F)$ holds if and only if the p-primary part $Ш(E/F)[p^\infty]$ is finite.

It is also useful to consider the p-Selmer group $\mathrm{Sel}_p(E/F)$ and the p-torsion $Ш(E/F)[p]$ of $Ш(E/F)$. The p-Selmer group can be defined as the fiber product

$$\begin{array}{ccc} \mathrm{Sel}_p(E/F) & \longrightarrow & H^1(F, E[p]) \\ \downarrow & & \downarrow \\ \prod_v E(F_v)/pE(F_v) & \xrightarrow{\prod_v \delta_v} & \prod_v H^1(F_v, E[p]) \end{array}$$

We have the exact sequence of vector spaces over $\mathbb{F}_p$ (the finite field of p elements):

$$0 \to E(F)\otimes \mathbb{Z}/p\mathbb{Z} \to \mathrm{Sel}_p(E/F) \to Ш(E/F)[p] \to 0.$$

This sequence is called the *p-descent (or the first descent) of E/F*. It has its genesis in Fermat's method of descent, used by him in the 17th century to study certain Diophantine equations. The p-Selmer group can be effectively computed (though not necessarily easily).

Comparing with the p^∞-Selmer group, we have an exact sequence of $\mathbb{F}_p$-vector spaces

$$0 \to E(F)[p] \to \mathrm{Sel}_p(E/F) \to \mathrm{Sel}_{p^\infty}(E/F)[p] \to 0, \tag{1.4}$$

where $[p]$ denotes the p-torsion. This sequence turns out to be very useful. For example, if $\mathrm{Sel}_p(E/F)$ is trivial, so is $\mathrm{Sel}_{p^\infty}(E/F)$. If $E(F)[p]$ is trivial and $\dim_{\mathbb{F}_p} \mathrm{Sel}_p(E/F) = 1$, then a simple argument using Cassels–Tate pairing shows

$$\mathrm{Sel}_{p^\infty}(E/F) \simeq \mathbb{Q}_p/\mathbb{Z}_p,$$

in particular, $r_p(E/F) = 1$. Then the finiteness of Ш predicts that $r_{MW}(E/F) = 1$. Such results can indeed be proved unconditionally for nice primes p as we will see.

1.3. **The Birch–Swinnerton-Dyer conjecture.** The origins of this conjecture can be traced back to numerical computations done by Birch and Swinnerton-Dyer ([5]). They were motivated by Siegel's mass formula for quadratic forms. Recall that the mass formula provides a weighted count of integral quadratic forms within a fixed genus class (two forms are said to be in the same genus if they are locally integrally equivalent for all places). Roughly speaking, the formula expresses the weighted count (something "global") as a product of local terms:

$$\prod_p \frac{\#G(\mathbb{F}_p)}{p^{\dim G}}, \tag{1.5}$$

where G is the (special) orthogonal group, defined over $\mathbb{Z}$, attached to any a quadratic lattice within the genus class. From this product, one naturally obtain a product of the values of Riemann zeta function $\zeta(s)$ at certain integers.

Now for an elliptic curve $E/\mathbb{Q}$, it is natural to investigate the product

$$\prod_{p\leqslant X} \frac{\#E(\mathbb{F}_p)}{p}, \tag{1.6}$$

where $\#E(\mathbb{F}_p)$ is the number of points over the finite field $\mathbb{F}_p$.[1]

At this moment let us introduce the Hasse–Weil (complex) L-function. It is defined as an Euler product of local L-factors

$$L(E/\mathbb{Q}, s) = \prod_p L(E/\mathbb{Q}_p, s). \tag{1.7}$$

The local L-factors are defined as

$$L(E/\mathbb{Q}_p, s) = \begin{cases} (1 - a_p p^{-s} + p^{1-2s})^{-1}, & p \nmid N, \\ (1 - a_p p^{-s})^{-1}, & p | N, \end{cases}$$

where N is the conductor of $E/\mathbb{Q}$, $a_p = 1 + p - \#E(\mathbb{F}_p)$ when $p \nmid N$, and $a_p = +1$, -1, 0, respectively, when $p|N$ and $E/\mathbb{Q}_p$ has split, non-split multiplicative, additive reduction, respectively.

Note that for $p \nmid N$, we have

$$\frac{\#E(\mathbb{F}_p)}{p} = 1 - a_p p^{-1} + p^{-1} = \frac{1}{L(E/\mathbb{Q}_p, 1)}.$$

The product is formally

$$\prod_{p\leqslant\infty} \frac{\#E(\mathbb{F}_p)}{p} \text{“}=\text{”} \frac{1}{L(E/\mathbb{Q}, 1)}.$$

Based on their numerical computation on the family of elliptic curves arising from the "congruent number problems"

$$E_n : y^2 = x^3 - n^2 x,$$

Birch and Swinnerton-Dyer were then led to the following conjecture ([5]):

$$\prod_{p\leqslant X} \frac{\#E(\mathbb{F}_p)}{p} \sim c'(E/\mathbb{Q}) \cdot (\log X)^r, \tag{1.8}$$

[1] One needs to be careful at the those (finitely many) bad primes.

and

$$L(E/\mathbb{Q}, s) \sim c(E/\mathbb{Q}) \cdot (s-1)^r, \tag{1.9}$$

where $r = r_{MW}(E/\mathbb{Q})$ is the Mordell–Weil rank of $E/\mathbb{Q}$ and $c'(E/\mathbb{Q})$ and $c(E/\mathbb{Q})$ are some nonzero constants. Indeed, the asymptotic behavior (1.8) is extremely strong! By a result of Goldfeld ([14]), (1.8) implies that $L(E/\mathbb{Q}, s)$ satisfies the Riemann hypothesis and (1.9) holds with the constant

$$c(E/\mathbb{Q}) = \frac{\sqrt{2}e^{r\gamma}}{c'(E/\mathbb{Q})},$$

where γ is the Euler constant.

The L-function $L(E/\mathbb{Q}, s)$ introduced earlier, by the theorems of Wiles, Taylor–Wiles and Breuil–Conrad–Diamond–Taylor, is equal to the L-series $L(f_E, s)$ of a weight two newform f_E of level N. Therefore it extends to an entire function on $\mathbb{C}$ and admits a functional equation with center at $s = 1$. The vanishing order $\mathrm{ord}_{s=1} L(E/\mathbb{Q}, s)$ at the center $s = 1$ is denoted by $r_{an}(E/\mathbb{Q})$ and called the *analytic rank* of $E/\mathbb{Q}$.

The Birch–Swinnerton-Dyer conjecture on rank asserts that the analytic rank and the Mordell–Weil rank coincide.

Conjecture 1.2 (Birch–Swinnerton-Dyer conjecture on rank)**.** *Let E be an elliptic curve over $\mathbb{Q}$. Then we have*

$$r_{an}(E/\mathbb{Q}) = r_{MW}(E/\mathbb{Q}).$$

Informally speaking, the conjecture implies that knowing enough information of the L-values means knowing the Mordell–Weil rank.

We would also like to recall the refined conjecture of Birch and Swinnerton-Dyer on the leading coefficient $c(E/\mathbb{Q})$ of the Taylor expansion of the L-function at its center of symmetry. We need to define two more ingredients:

(1) the period $P(E/\mathbb{Q})$,
(2) the regulator $R(E/\mathbb{Q})$.

To define the period $P(E/\mathbb{Q})$, we may take any a nonzero invariant differential $\omega \in H^0(E, \Omega_{E/\mathbb{Q}})$. For each place $p \leqslant \infty$, one may naturally associate a measure $|\omega|_p$ on the compact group $E(\mathbb{Q}_p)$[2]. Then we define

$$P(E/\mathbb{Q}) := \left(\int_{E(\mathbb{R})} |\omega|_\infty\right) \cdot \prod_{p<\infty} \left(L(E/\mathbb{Q}_p, 1) \int_{E(\mathbb{Q}_p)} |\omega|_p\right),$$

where the local L-factor is a normalizing factor such that the local term is equal to one for all but finitely many p. The definition is independent of the choice of a nonzero $\omega \in H^0(E, \Omega_{E/\mathbb{Q}})$ since any other choice differs by a constant $\alpha \in \mathbb{Q}^\times$ and we have the product formula: $\prod_{p\leqslant\infty} |\alpha|_p = 1$. One can also define the period by choosing ω canonically as the Néron differential ω_0, which is a generator of the free $\mathbb{Z}$-module of rank one $H^0(\mathcal{E}, \Omega_{\mathcal{E}/\mathbb{Z}})$ for the Néron model $\mathcal{E}/\mathbb{Z}$ of $E/\mathbb{Q}$. It is unique up to ± 1 and hence we have a well-define period

$$\Omega_E := \int_{E(\mathbb{R})} |\omega_0|_\infty,$$

[2]Note that this definition of the period $P(E/\mathbb{Q})$ is very much like that of the Tamagawa number of a linear algebraic group. This should be natural since Siegel's mass formula was equivalent to the fact that the Tamagawa number of the special orthogonal group is 2.

and we have an explicit formula

$$P(E/\mathbb{Q}) = \Omega_E \cdot \prod_{p|N} c_p, \tag{1.10}$$

where c_p is the so-called local Tamagawa number, i.e., the cardinality of the component group of the Néron model of E over $\mathbb{Z}_p$.

To define the regulator $R(E/\mathbb{Q})$, we choose a basis $P_1, P_2, ..., P_r$ of a free subgroup of finite index I in $E(\mathbb{Q})$ and define

$$R(E/\mathbb{Q}) = \frac{\det(\langle P_i, P_j\rangle)}{I^2},$$

where $\langle P_i, P_j\rangle$ is the Néron–Tate height pairing. It is easy to see that this does not depend on the choice of the P_i. In particular, one may choose P_i to generate the free part of $E(\mathbb{Q})$ (i.e., together with the torsion subgroup $E(\mathbb{Q})_{tor}$ they generate $E(\mathbb{Q})$). Then we have an explicit formula

$$R(E/\mathbb{Q}) = \frac{\det(\langle P_i, P_j\rangle)}{\#E(\mathbb{Q})_{tor}^2}. \tag{1.11}$$

The regulator measures the density of the Mordell–Weil lattice with respect to the metric defined by the Néron–Tate height pairing on $E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$.

Conjecture 1.3 (Birch–Swinnerton-Dyer refined conjecture)**.** *Let E be an elliptic curve over $\mathbb{Q}$ of analytic rank r. Then we have*

$$\lim_{s\to 1} \frac{L(E,s)}{(s-1)^r} = P(E/\mathbb{Q}) \cdot R(E/\mathbb{Q}) \cdot \#Ш(E/\mathbb{Q}). \tag{1.12}$$

This gives a conjectural formula of $c(E/\mathbb{Q})$ in (1.9). In terms of (1.10) and (1.11), the refined formula can be written as

$$\frac{L^{(r)}(E,1)}{r! \cdot \Omega_E} = \#Ш(E/\mathbb{Q}) \cdot \frac{\det(\langle P_i, P_j\rangle)}{\#E(\mathbb{Q})_{tor}^2} \cdot \prod_{p|N} c_p. \tag{1.13}$$

We can state the refined B-SD conjecture in a manner more analogous to Weil's conjecture on Tamagawa numbers. Define the Tamagawa number of $E/\mathbb{Q}$ as

$$\tau(E/\mathbb{Q}) := \frac{\prod_{p\leqslant\infty} L(E/\mathbb{Q}_p, 1) \int_{E(\mathbb{Q}_p)} |\omega|_p}{L^*(E,1)} \cdot R(E/\mathbb{Q}), \quad L^*(E,1) = L^{(r)}(E,1)/r!.$$

Then the refined B-SD conjecture is equivalent to

$$\tau(E/\mathbb{Q}) = \frac{1}{\#Ш(E/\mathbb{Q})}.$$

This resembles the formula proved by T. Ono for the Tamagawa number of an algebraic torus $T \subset \mathrm{GL}_n$ defined over a number field (or a function field over a finite field) F

$$\tau(T/F) = \frac{\#\mathrm{Pic}(T/F)}{\#Ш(T/F)}.$$

Here the "Picard group" $\mathrm{Pic}(T/F)$ is defined as $H^1(F, X(T))$ for the character group $X(T) = \mathrm{Hom}(T, \mathbb{G}_m)$ with the natural Gal_F-action, and $Ш(T/F) := \mathrm{Ker}(H^1(F,T) \to \prod_v H^1(F_v, T))$ is a finite group.

Remark 4. The Birch and Swinnerton-Dyer conjecture for both the rank and the refined formula has a natural generalization to abelian varieties over an arbitrary number field ([48]).

Remark 5. In [6] Bloch gave a volume-theoretical interpretation of the conjectural formula (1.12) as the Tamagawa number conjecture for an algebraic group which is an extension of the elliptic curve E by an algebraic torus.

Remark 6. If we replace the base field $\mathbb{Q}$ by a function field F (such as $\mathbb{F}_q(t)$ over a finite field $\mathbb{F}_q$), the finiteness of $Ш(E/F)$ is equivalent the Birch and Swinnerton-Dyer conjecture 1.2 on rank, and implies the refined conjecture 1.3. It is also equivalent to the Tate conjecture for elliptic surfaces over a finite field.

For some historical account of the B-SD conjecture, the readers are invited to consult the articles [5], [47] and [18].

1.4. **The status to date.** For the rank part, by far the most general result for elliptic curves over $\mathbb{Q}$ is obtained using the Gross–Zagier formula and the Heegner point Euler system of Kolyvagin:

Theorem 1.4 (Gross–Zagier, Kolyvagin)**.** *If $r_{an}(E/\mathbb{Q}) \leqslant 1$, then $r_{an}(E/\mathbb{Q}) = r_{MW}(E/\mathbb{Q})$ and $Ш(E/\mathbb{Q})$ is finite.*

S. Zhang generalized this result to many modular elliptic curves (and modular abelian varieties of GL_2-type) over totally real number fields ([58]).

The proof of this result requires a suitable auxiliary choice of imaginary quadratic field. Such auxiliary choice exists by a non-vanishing result in analytic number theory ([7], [38]).

Remark 7. Kato has an independent proof using his Euler system ([29]) that, if $r_{an}(E/\mathbb{Q}) = 0$, then $r_{MW}(E/\mathbb{Q}) = 0$ and $Ш(E/\mathbb{Q})$ is finite. Bertolini and Darmon in [4] also prove that if $r_{an}(E/\mathbb{Q}) = 0$, then $r_{MW}(E/\mathbb{Q}) = 0$. For elliptic curves $E/\mathbb{Q}$ with complex multiplication, Coates–Wiles ([8]) earlier already proved that if $r_{an}(E/\mathbb{Q}) = 0$, then $r_{MW}(E/\mathbb{Q}) = 0$.

Regarding the refined formula (1.12), Rubin proved that, for elliptic curves $E/\mathbb{Q}$ with complex multiplication by an imaginary quadratic field K, if $r_{an}(E/\mathbb{Q}) = 0$, then (1.12) holds up to some primes dividing the order of the group of units $\mathcal{O}_K^\times$. If $r_{an}(E/\mathbb{Q}) = 1$ for E with complex multiplication, there are results on the refined B-SD formula due to Perrin–Riou ([39], for ordinary primes) and Kobayashi ([30], for supersingular primes).

For elliptic curves $E/\mathbb{Q}$ without complex multiplication, the theorem of Kato and Skinner–Urban on the Iwasawa–Greenberg main conjecture for GL_2 ([44]) implies that the p-part of the formula (1.12) holds for nice p in the case $r(E/\mathbb{Q}) = 0$:

Theorem 1.5 (Kato, Skinner–Urban)**.** *Let $E/\mathbb{Q}$ be an elliptic curve with conductor N. Let $p \geqslant 3$ be a prime such that:*

1. *E has good ordinary reduction at p.*
2. *$\overline{\rho}_{E,p}$ is surjective.*
3. *There exists a prime $\ell \| N$ such that $\overline{\rho}_{E,p}$ is ramified at ℓ.*

If $L(E,1) \neq 0$, then the p-part of the B-SD formula (1.12) holds, i.e.:

$$\left|\frac{L(E,1)}{\Omega_E}\right|_p = \left|Ш(E/\mathbb{Q}) \cdot \prod_{\ell | N} c_\ell\right|_p .$$

Remark 8. The condition (3) can be removed by recent work of X. Wan [54].

A recent result of the author ([67]) is that the p-part of the formula (1.12) holds for nice p in the case $r(E/\mathbb{Q}) = 1$.

Theorem 1.6. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Let $p \geqslant 5$ be a prime such that:*

(1) *E has good ordinary reduction at p.*
(2) *$\overline{\rho}_{E,p}$ is surjective.*
(3) *There exist at least two primes $\ell||N$ where $\overline{\rho}_{E,p}$ is ramified.*
(4) *If $\ell \equiv \pm 1 \mod p$ and $\ell||N$, then $\overline{\rho}_{E,p}$ is ramified at ℓ.*

If $r_{an}(E/\mathbb{Q}) = 1$, then the p-part of the B-SD formula (1.12) holds, i.e.:

$$\left|\frac{L'(E,1)}{\Omega_E \cdot R(E/\mathbb{Q})}\right|_p = \left|\#Ш(E/\mathbb{Q}) \cdot \prod_{\ell|N} c_\ell\right|_p.$$

Remark 9. Implicitly in the last two theorems, the ratios $\frac{L(E,1)}{\Omega_E}$ and $\frac{L'(E,1)}{\Omega_E \cdot R(E/\mathbb{Q})}$ are rational numbers. Moreover, there are similar results for modular abelian varieties of GL_2-type over $\mathbb{Q}$. But it is not yet completely understood how to generalize them to totally real number fields.

Remark 10. For an elliptic curve $E/\mathbb{Q}$, let $j_E \in \mathbb{Q}$ be the j-invariant. Let N be its conductor and Δ_E its minimal discriminant. Then for a prime $\ell||N$, we have

$$v_\ell(\Delta_E) = -v_\ell(j_E).$$

The residual representation $\overline{\rho}_{E,p}$ is ramified at $\ell||N$ if and only if $p \nmid v_\ell(j_E)$.

Remark 11. It seems difficult to obtain the same result for small primes p, especially $p = 2$. However, Tian ([48], [49]) and Tian–Yuan–S. Zhang ([50]) have proved the 2-part of the B-SD formula in the case of analytic rank zero or one, for "many" (expected to be of a high percentage) quadratic twists of the congruent number elliptic curve:

$$E_n : y^2 = x^3 - n^2 x.$$

1.5. **Recent results on Selmer groups.**

Theorem 1.7. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Let $p \geqslant 3$ be a prime such that:*

(1) *E has good ordinary reduction at p.*
(2) *$\overline{\rho}_{E,p}$ is irreducible.*
(3) *There exists a prime $\ell||N$ such that $\overline{\rho}_{E,p}$ is ramified at ℓ.*

Then the following are equivalent

(1) $r_p(E/\mathbb{Q}) = 0$.
(2) $r_{MW}(E/\mathbb{Q}) = 0$ *and the p-primary $Ш(E/\mathbb{Q})[p^\infty]$ is finite.*
(3) $r_{an}(E/\mathbb{Q}) = 0$.

Remark 12. (2) $\Rightarrow$ (1) holds trivially for all p; (3) $\Rightarrow$ (2) follows from Theorem 1.4 (and true for all p) or Kato's theorem; (1) $\Rightarrow$ (3) is due to Skinner–Urban ([44]).

Theorem 1.8. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Let $p \geqslant 5$ be a prime such that:*

(1) *E has good ordinary reduction at p.*
(2) *$\overline{\rho}_{E,p}$ is surjective.*

(3) *There exist at least two primes* $\ell||N$ *where* $\overline{\rho}_{E,p}$ *is ramified.*
(4) *If* $\ell \equiv \pm 1 \mod p$ *and* $\ell||N$*, then* $\overline{\rho}_{E,p}$ *is ramified at* ℓ*.*

Then the following are equivalent

(1) $r_p(E/\mathbb{Q}) = 1$.
(2) $r_{MW}(E/\mathbb{Q}) = 1$ *and the* p*-primary* $Ш(E/\mathbb{Q})[p^\infty]$ *is finite.*
(3) $r_{an}(E/\mathbb{Q}) = 1$.

Remark 13. (2) $\Rightarrow$ (1) holds trivially for all p; (3) $\Rightarrow$ (2) follows from Theorem 1.4 (and true for all p); (1) $\Rightarrow$ (3) is a consequence of the Kolyvagin conjecture proved by the author in [67] (also cf. §4, Remark 22). Y. Tian ([48], [49]) first proved a result of the type (1) $\Rightarrow$ (3) for $p = 2$ and many quadratic twists of the congruent number elliptic curve. Skinner ([42]) has also proved a result of the type (2) $\Rightarrow$ (3).

The interest of the result of type (1) $\Leftrightarrow$ (2) in both theorems lies in the fact that the statements are purely algebraic and do not involve the L-values (though the current proof, as we will see, must go through the study of special value of L-function). Much of the appeal stems from the fact that the assumption $r_p(E/\mathbb{Q}) = 0$ or 1 is usually easy to check. For instance, by (1.4), if $E[p](\mathbb{Q}) = 0$ (automatically true when $\overline{\rho}_{E,p}$ is irreducible) and $\dim_{\mathbb{F}_p} \mathrm{Sel}_p(E/\mathbb{Q}) = d \in \{0, 1\}$, then we have $r_p(E/\mathbb{Q}) = d$.

Theorem 1.8 leads to the following converse to the theorem of Gross–Zagier and Kolyvagin 1.4.

Theorem 1.9. *Let* $E/\mathbb{Q}$ *be an elliptic curve of conductor* N*. Assume that there are at least two distinct prime factors* $\ell||N$*. Then we have*

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

This is proved in [67], and earlier by Skinner [42] for square-free N using different methods.

1.6. **Outline of this survey.** This survey article will be devoted to some main ingredients in the proof of Theorem 1.6 and 1.8, as well as some partial generalization to higher rank motives. The complete proof of Theorem 1.6 and 1.8 is given in [67].

The section §2 is devoted to the Waldspurger formula for GL_2, and its generalization to higher rank groups, i..e, the global Gan–Gross–Prasad conjecture and its refinement.

The section §3 is devoted to the Gross–Zagier formula on Shimura curves. We aim to state this formula in the most general form, due to Yuan–Zhang–Zhang [56]. The formulation is completely parallel to that of Waldspurger formula in §2. On the earlier developments on Heegner points and the Gross–Zagier formula with certain ramification hypothesis, the reader may consult a previous CDM article by S. Zhang [60].

The section §4 is devoted to the Heegner point Kolyvagin system and the structure of Selmer groups. Particularly, the attention is paid to the Kolyvagin conjecture on the divisibility of higher Heegner points, which the author proved under a certain local condition in [67].

We have been unable to include the Iwasawa theoretical aspect (which yields the proof of Theorem 1.5 and 1.7 of Kato, Skinner–Urban) into this survey. The reader may consult a previous CDM article by Skinner [43].

2. Waldspurger formula for GL_2 and higher rank groups

2.1. **A formula of Gross.** We start with a formula proved by Gross [15]. Let f be a newform of weight two and level N, with trivial nebentypus. Let $K = \mathbb{Q}[\sqrt{-D}]$ be an imaginary quadratic extension with discriminant $-D < -4, (D, N) = 1$. This determines a unique factorization

$$N = N^+ N^- \tag{2.1}$$

where the prime factors of N^+ (N^-, resp.) are all split (inert, resp.) in K. Assume that N^- is square-free. Then the root number

$$\epsilon(f/K) = 1$$

if and only if N^- has *odd* number of prime factors, which we assume now. Let B be the unique quaternion algebra over $\mathbb{Q}$ which is ramified at exactly primes dividing N^- and ∞. Let π_B be the Jacquet–Langlands correspondence of the automorphic representation π associated to f. We may consider an Eichler order $\mathcal{O}_{B,N^+}$ in $\mathcal{O}_B$ with level N^+ and define the Shimura set as

$$X_{N^+,N^-} := B^\times(\mathbb{Q})\backslash B^\times(\mathbb{A}_f)/(\mathcal{O}_{B,N^+} \otimes \widehat{\mathbb{Z}})^\times. \tag{2.2}$$

This may be interpreted as the set of one-sided ideal classes of the Eichler order. Let ϕ be a new-vector of π_B. It defines a function on X_{N^+,N^-}. Embed K into B so that $K \cap \mathcal{O}_{B,N^+} = \mathcal{O}_K$. This induces an embedding of the ideal class group of K into the Shimura set:

$$\mathrm{Pic}(\mathcal{O}_K) \simeq K^\times \backslash \widehat{K}^\times / \widehat{\mathcal{O}}_K^\times \hookrightarrow X_{N^+,N^-}.$$

For a function $\phi : X_{N_+,N^-} \to \mathbb{C}$ we define

$$\mathscr{P}_K(\phi) = \int_{t \in \mathrm{Pic}(\mathcal{O}_K)} \phi(t) dt,$$

and

$$\langle \phi, \phi \rangle = \int_{X_{N^+,N^-}} \phi(x)\overline{\phi(x)} dx,$$

where dt (dx, resp.) is the counting measure on the finite set $\mathrm{Pic}(\mathcal{O}_K)$ (X_{N^+,N^-}, reap.).

Consider the Petersson inner product defined by

$$(f, f) := 8\pi^2 \int_{\Gamma_0(N)\backslash\mathcal{H}} f(z)\overline{f(z)} dx dy = \int_{X_0(N)} \omega_f \wedge \overline{i\omega}_f, \tag{2.3}$$

where $\omega_f := 2\pi i f(z) dz$. Let $L(f/K, s)$ be the L-function (without the archimedean factor) with the classical normalization, i.e, the center of functional equation is at $s = 1$.

Theorem 2.1. *We have*

$$\frac{L(f/K, 1)}{(f, f)} = \frac{1}{\sqrt{|D|}} \frac{|\mathscr{P}_K(\phi)|^2}{\langle \phi, \phi \rangle}. \tag{2.4}$$

Now we specialize to an f associated to an elliptic curve $E/\mathbb{Q}$. Then we have $L(f/K, s) = L(E/K, s)$. We consider a modular parameterization, still denoted by f,

$$f : X_0(N) \to E, \tag{2.5}$$

which we assume maps the cusp (∞) to zero. The pull-back of the Néron differential ω on E is $f^*\omega = c \cdot \omega_f$ for a constant c (called "Manin constant"). It is known that the constant c is an integer. If f is an optimal parameterization, it is called the Manin constant and conjecturally equal to 1. By a theorem of Mazur, we have $p|c \Longrightarrow p|2N$.

The automorphic form ϕ on the Shimura set X_{N^+,N^-} is normalized so that it is integral and its image contains 1. It is then unique up to ± 1. Then the formula (2.4) can be rewritten as a form that is close the Birch–Swinnertod-Dyer formula for E/K (cf. (1.13)):

$$\frac{L(E/K,1)}{|D|^{-1/2}\int_{E(\mathbb{C})}\omega\wedge\overline{\omega}} = \frac{1}{c^2}\frac{\deg(f)}{\deg(\phi)}|\mathscr{P}_K(\phi)|^2. \tag{2.6}$$

The term $\deg(\phi) := \langle\phi,\phi\rangle$ is an analogue of the modular degree $\deg(f)$, even though there is no physical modular parameterization of E by the set X_{N^+,N^-}.

2.2. **Waldspurger formula for** GL_2**.**

2.2.1. *The period integral.* We fix the following data:

- F a number field with adeles denoted by $\mathbb{A} = \mathbb{A}_F$.
- $G = B^\times$ as an algebraic group over F. Here B is a quaternion algebra over F. Its center is denoted by Z_G.
- E/F a quadratic extension of number fields [3], with a fixed embedding $E \hookrightarrow B$. Denote by $H = E^\times$, viewed as an algebraic group over F. The embedding $E \hookrightarrow B$ makes H a subgroup of G.
- $\eta : F^\times\backslash\mathbb{A}^\times \to \{\pm 1\}$ the quadratic character associated to E/F by class field theory.
- π an irreducible cuspidal automorphic representation of G.
- $\chi : H(F)\backslash H(\mathbb{A}) \to \mathbb{C}^\times$, a (unitary) character compatible with the central character ω_π of π:
$$\omega_\pi \cdot \chi|_{\mathbb{A}^\times} = 1.$$

Then we consider a linear functional on π defined by

$$\mathscr{P}_\chi(\phi) := \int_{Z_G(\mathbb{A})\mathrm{H}(F)\backslash\mathrm{H}(\mathbb{A})} \phi(h)\chi(h)dh, \quad \phi \in \pi. \tag{2.7}$$

Since H is a torus, this is sometimes called a toric period. It is obviously $H(\mathbb{A})$-invariant:

$$\mathscr{P}_\chi \in \mathrm{Hom}_{H(\mathbb{A})}(\pi \otimes \chi, \mathbb{C}).$$

2.2.2. *Branching law.* Analysis of this last space itself leads to interesting questions on branching laws, familiar to us from representation theory. Decompose $\pi = \otimes_v \pi_v$ as a tensor product, and similary $\chi = \otimes\chi_v$. Then we have

- Multiplicity one: $\dim \mathrm{Hom}_{H(F_v)}(\pi_v \otimes \chi_v, \mathbb{C}) \leqslant 1$.
- Dichotomy: Let π_v^0 be an infinite dimensional irreducible representation of $\mathrm{GL}_2(F_v)$. Let B_v be the unique division quaternion algebra over F_v and π_v^1 the representation of $B_v^\times$ associated to π_v^0 by Jacquet and Langlands

[3] There is a notational nightmare at this point: we have been using E for a quadratic extension in many places while it is also customary to use E for an elliptic curve. We warn the reader about the inconsistence in different sections of this article.

(where we set $\pi_v^1 = 0$ if π_v^0 is not a discrete series representation). Then we have a dichotomy:

$$\dim \mathrm{Hom}_{H(F_v)}(\pi_v^0 \otimes \chi_v, \mathbb{C}) + \dim \mathrm{Hom}_{H(F_v)}(\pi_v^1 \otimes \chi_v, \mathbb{C}) = 1.$$

- Root number: It is possible to relate the vanishing or non-vanishing of the above Hom spaces to root numbers. A theorem of of Tunnell and Saito asserts that $\dim \mathrm{Hom}_{H(F_v)}(\pi_v^i \otimes \chi_v, \mathbb{C}) = 1, i \in \{0,1\}$ if and only if
$$\epsilon(1/2, \pi_v, \chi_v) = (-1)^i \chi_v(-1)\eta_v(-1),$$
where $\epsilon(1/2, \pi_v, \chi_v) \in \{\pm 1\}$ is the local root number associated to the representation $\pi_{v,E_v} \otimes \chi_v$, the base change to E_v twisted by χ_v.

2.2.3. *Local canonical invariant form.* So far there seems to be no natural construction of any element in $\mathrm{Hom}_{H(F_v)}(\pi_v \otimes \chi_v, \mathbb{C})$. However, Waldspurger constructed a natural element in the (at most one-dimensional) space

$$\mathrm{Hom}_{H(F_v)}(\pi_v \otimes \chi_v, \mathbb{C}) \otimes \mathrm{Hom}_{H(F_v)}(\pi_v^\vee \otimes \chi_v^{-1}, \mathbb{C}),$$

where $\pi_v^\vee$ is the contragredient of π_v. Let $\langle \cdot, \cdot \rangle$ be the canonical pairing $\pi_v \times \pi_v^\vee \to \mathbb{C}$. Define an average of the matrix coefficient

$$\alpha_v(\phi_v, \varphi_v) = \int_{Z_G(F_v)\backslash H(F_v)} \langle \pi_v(h)\phi_v, \varphi_v \rangle \chi_v(h)\, dh. \tag{2.8}$$

The integral is absolutely convergent for any unitary representation π_v and defines a canonical invariant form:

$$\alpha_v \in \mathrm{Hom}_{H(F_v)}(\pi_v \otimes \chi_v, \mathbb{C}) \otimes \mathrm{Hom}_{H(F_v)}(\pi_v^\vee \otimes \chi_v^{-1}, \mathbb{C}).$$

When π_v is unramified[4], and the vectors ϕ_v, φ_v are fixed by K_v such that $\langle \phi_v, \varphi_v \rangle = 1$, we have

$$\alpha_v(\phi_v, \varphi_v) = \mathscr{L}(\pi_v, \chi_v) := \frac{\zeta_{F_v}(2) L(1/2, \pi_{v,E_v} \otimes \chi_v)}{L(1, \pi_v, \mathrm{Ad}) L(1, \eta_v)}.$$

Therefore, for global purposes, we normalize the canonical invariant form α_v as follows:

$$\alpha_v^\natural(\phi_v, \varphi_v) = \frac{1}{\mathscr{L}(\pi_v, \chi_v)} \alpha_v(\phi_v, \varphi_v). \tag{2.9}$$

2.2.4. *Waldspurger formula.* We define a global pairing $\pi \times \pi^\vee \to \mathbb{C}$ using the Petersson inner product

$$\langle \phi, \varphi \rangle = \int_{Z_G(\mathbb{A})G(F)\backslash G(\mathbb{A})} \phi(g)\varphi(g) dg, \quad \phi \in \pi, \varphi \in \pi^\vee,$$

where we choose the Tamagawa measure on $G(\mathbb{A})$. We normalize the measure dh on $H(\mathbb{A})$ and the measures dh_v on $H(F_v)$ such that $dh = \prod_v dh_v$. Then the Waldspurger formula ([52]) can be stated as follows.

Theorem 2.2 (Waldspurger)**.** *For $\phi = \otimes \phi_v \in \pi, \varphi = \otimes \varphi_v \in \pi^\vee$, we have*

$$\frac{\mathscr{P}(\phi)\mathscr{P}(\varphi)}{\langle \phi, \varphi \rangle} = \frac{1}{4} \frac{\zeta_F(2) L(1/2, \pi_E \otimes \chi)}{L(1, \pi, \mathrm{Ad}) L(1, \eta)} \prod_v \frac{\alpha_v^\natural(\phi_v, \varphi_v)}{\langle \phi_v, \varphi_v \rangle_v}. \tag{2.10}$$

[4]For a non-archimedean place v we say that π_v is unramified if the quadratic extension E/F is unramified at v, the group $G(F_v)$ has a hyperspecial subgroup $K_v = G(\mathcal{O}_v)$ and π_v has a nonzero K_v-fixed vector.

The formula of Gross (2.4) in the beginning of this section can be viewed as an explicit Waldspurger formula for the new vectors in the representation space $\pi, \pi^\vee$, when $F = \mathbb{Q}, E = \mathbb{Q}[\sqrt{-D}]$, the central character ω_π and the character χ are trivial. In various setting, such explicit formulae were also considered by S. Zhang et al in [59], [61], [50]. It is well-known that they play an important role in the arithmetic study of elliptic curves and modular forms, intimately related to some "explicit reciprocity law" for instance.

2.3. **The global Gan-Gross-Prasad conjecture and its refinement.** Study of the Waldspurger formula for GL_2 has exploded into a broad web of conjectures on branching problems, automorphic periods, and L-functions, due to Gross–Prasad [19] [20] in the 1990s, Gan–Gross–Prasad more recently [9], and Ichino–Ikeda [26], N. Harris [23]. To describe them, let G be a reductive group and H a subgroup defined over a number field F (with adeles $\mathbb{A}$). Let π be an automorphic cuspidal representation of G. Then we consider the automorphic period integral, as a linear functional on π:

$$\mathscr{P}_{\mathrm{H}}(\phi) := \int_{\mathrm{H}(F)\backslash \mathrm{H}(\mathbb{A})} \phi(h)dh, \quad \phi \in \pi, \tag{2.11}$$

in the orthogonal and Hermitian cases in [9]. To describe them let F be a number field and let $E = F$ in the quadratic case and E a quadratic extension of F in the Hermitian case. Let W_{n+1} be a quadratic space or Hermitian space with E-dimension $n+1$. Let $W_n \subset W_{n+1}$ be a non-degenerate subspace of codimension one. Let G_i be $\mathrm{SO}(W_i)$ or $\mathrm{U}(W_i)$ for $i = n, n+1$. Then the Gan-Gross-Prasad period is the period integral (2.11) attached to the pair (H, G) where $\mathrm{G} = \mathrm{G}_n \times \mathrm{G}_{n+1}$ and $\mathrm{H} \subset \mathrm{G}$ is the diagonal embedding of G_n.

Let $\pi = \pi_n \otimes \pi_{n+1}$ be a cuspidal automorphic representation of $G(\mathbb{A})$. Let $\Pi_{i,E}$ be the standard functoriality transfer from G_i to suitable $\mathrm{GL}_N(\mathbb{A}_E)$: in the Hermitian case, this is the base change of π_i to $\mathrm{GL}_i(\mathbb{A}_E)$; in the orthogonal case, this is the endoscopic transfer from $G_i(\mathbb{A})$ to $\mathrm{GL}_i(\mathbb{A})$ ($\mathrm{GL}_{i-1}(\mathbb{A})$, resp.) if i is even (odd, resp.). We will assume the expected properties of the theory of the endoscopic functoriality transfer from classical group to the general linear group.

Two Langlands L-functions enter the stage:

- The L-function $L(s, \pi, \mathrm{R})$ for a certain representation R of the L-group ${}^L G$.
- The adjoint L-function $L(s, \pi, Ad)$ (cf. [9, §7]).

The first L-function $L(s, \pi, \mathrm{R})$ can be defined more explicitly as $L(s, \Pi_{n,E} \times \Pi_{n+1,E})$, the Rankin-Selberg convolution L-function due to Jacquet–Piatetski-Shapiro–Shalika ([27]) (known to be the same as the one defined by the Langlands–Shahidi method).

Denote $\Delta_{n+1} = L(M^\vee(1))$ where $M^\vee$ is the motive dual to the motive M associated to G_{n+1} defined by Gross ([17]). It is a product of (special values of) Artin L-functions. We will be interested in the following combination of L-functions

$$\mathscr{L}(s, \pi) = \Delta_{n+1} \frac{L(s, \pi, \mathrm{R})}{L(s + \frac{1}{2}, \pi, Ad)}. \tag{2.12}$$

They are completely analogous to the L-functions appeared in Waldspurger formula (2.10). We also write $\mathscr{L}(s, \pi_v)$ for the corresponding local factor at v.

The global Gan–Gross–Prasad conjecture asserts that the non-vanishing of the linear functional $\mathscr{P}_{\mathrm{H}}$ on π (possibly by varying the orthogonal/Hermitian spaces

(W, V) and switching to another member in the Vogan L-packet (cf. [9, §9-11]) of π) is equivalent to the non-vanishing of the central value $L(\frac{1}{2}, \pi, \mathrm{R})$ of the Rankin-Selberg L-function. This conjectural equivalence is proved in the Hermitian case for π satisfying some local conditions in [64]. One direction of the equivalence for both the orthogonal and Hermitian cases had also been proved by Ginzburg–Jiang–Rallis (cf. [12], [13]) when the group G is quasi-split and the representation π is generic.

For arithmetic applications, it is necessary to have the refinement of the Gan-Gross-Prasad conjecture, namely a Waldspurger formula for the period $\mathscr{P}_{\mathrm{H}}$ analogous to (2.10). For simplicity let $[H]$ denote the quotient $H(F)\backslash H(\mathbb{A})$; similarly define $[G]$. We endow $H(\mathbb{A})$ ($G(\mathbb{A})$, resp.) with their Tamagawa measures and $[H]$ ($[G]$, resp.) with the quotient measure by the counting measure on $H(F)$ ($G(F)$, resp.). Let $\langle \cdot, \cdot \rangle$ be the Peterson inner product

$$\langle \phi, \varphi \rangle = \int_{[G]} \phi(g)\varphi(g)\, dg, \quad \phi \in \pi, \varphi \in \pi^\vee. \tag{2.13}$$

To define a local canonical invariant form, we again consider the integration of matrix coefficients: for $\phi_v, \varphi_v \in \pi_v^\vee$,

$$\alpha_v(\phi_v, \varphi_v) = \int_{H_v} \langle \pi_v(h)\phi_v, \varphi_v \rangle_v \, dh, \quad H_v = H(F_v). \tag{2.14}$$

Here we normalize the measure dh on $H(\mathbb{A})$ and the measures dh_v on $H(F_v)$ such that $dh = \prod_v dh_v$. Ichino and Ikeda showed that this integral converges absolutely. When π_v is unramified and the vectors ϕ_v, φ_v are fixed by a hyperspecial compact open $G(\mathcal{O}_v)$ such that $\langle \phi_v, \varphi_v \rangle_v = 1$, we have

$$\alpha_v(\phi_v, \varphi_v) = \mathscr{L}(\frac{1}{2}, \pi_v) \cdot \mathrm{vol}(H(\mathcal{O}_v)).$$

Analogous to (2.9) we normalize the local canonical invariant form α_v:

$$\alpha_v^\natural(\phi_v, \varphi_v) = \frac{1}{\mathscr{L}(\frac{1}{2}, \pi_v)} \int_{H_v} \langle \pi_v(h)\phi_v, \varphi_v \rangle_v \, dh. \tag{2.15}$$

The refined (global) Gan–Gross–Prasad conjecture as formulated by Ichino–Ikeda and N. Harris (cf. [26], [23]) states:

Conjecture 2.3. *Assume that π is* tempered*, i.e., π_v is tempered for all v. For $\phi = \otimes \phi_v \in \pi, \varphi = \otimes \varphi_v \in \pi^\vee$, we have*

$$\frac{\mathscr{P}(\phi)\mathscr{P}(\varphi)}{\langle \phi, \varphi \rangle} = \frac{1}{|S_\pi|} \mathscr{L}(\frac{1}{2}, \pi) \prod_v \frac{\alpha_v^\natural(\phi_v, \varphi_v)}{\langle \phi_v, \varphi_v \rangle_v}, \tag{2.16}$$

where S_π is a finite elementary 2-group: the component group associated to the L-parameter of $\pi = \pi_n \otimes \pi_{n+1}$.

Remark 14. The refined conjecture for $\mathrm{SO}(3) \times \mathrm{SO}(4)$, concerning "the triple product L-function", was established after the work by Garrett [11], Piatetski-Shapiro–Rallis, Harris–Kudla [22], Gross–Kudla, Watson [55], and Ichino [25]. Gan and Ichino ([10]) established some new cases for $\mathrm{SO}(4) \times \mathrm{SO}(5)$. Liu ([35]) proves some endoscopic cases for $\mathrm{SO}(2) \times \mathrm{SO}(5)$ and $\mathrm{SO}(3) \times \mathrm{SO}(6)$. All of these results utilize the theta correspondence.

In [65], the above refined Gan–Gross–Prasad conjecture is proved in the Hermitian case, under the following local conditions:

(i) There exists a split place v such that the local component π_v is supercuspidal.
(ii) If π_v is not unramified, then either v is split in E/F, or H_v is compact or π_v is supercuspidal.
(iii) If π_v is unramified, then its residue characteristic of F_v is larger than a constant $c(n)$. The constant is defined such that the Jacquet–Rallis fundamental lemmaholds when the residue characteristic $p \geqslant c(n)$ for a constant $c(n)$ depending only on n (cf. [57] and its appendix).

The method of the proof of both the unrefined (in [64]) and the refined (in [65]) conjecture in the Hermitian case is to study the relative trace formula initiated by Jacquet–Rallis ([28]). One crucial ingredient—the fundamental lemma—is proved by Z. Yun ([57]). For other ingredients, one may consult the expository article [66].

In [9], Gan–Gross–Prasad also proposed a local conjecture to address the three questions on the branching law: the multiplicity one, the dichotomy and the relation to local root number. The local conjecture specifies the pure inner form of the reductive group G and the representation in the L-packet which makes the local canonical invariant form non-vanishing. There has been substantial progress towards a complete resolution to the local conjectures (at least for p-adic local fields), due to many people: Aizenbud–Gourevitch–Rallis–Schiffmann [1] and Sun–Zhu [45] for the multiplicity-one theorem, Waldspurger [53] (orthogonal groups) and Beuzart-Plessis [2] (unitary groups) for the dichotomy and the relation to root numbers.

Gan–Gross–Prasad also made analogous conjectures for (W, V) where W is not not necessarily of codimension one. Towards this, in the Hermitian cases, Yifeng Liu in [34] has generalized Jacquet–Rallis's construction of relative trace formulae, and proved some cases of relevant fundamental lemmas. Recently, Yifeng Liu in [35] has also extended the formulation of the refined conjecture to more general Bessel periods.

3. Gross–Zagier formula for GL_2

In this section, we recall the formulation of the general Gross–Zagier formula for GL_2, following the joint work by X. Yuan, S. Zhang and the author in [56].

3.1. Gross–Zagier fromula for X_{N^+,N^-}.

As an example, we first give an explicit version of the formula in [56] in a special case, which is used in the proof of Theorem 1.6.

In 1950s, Heegner first realized that modular parameterization could be used to construct rational points on elliptic curves [24]. By applying this idea to the congruent number curve $y^2 = x^3 - n^2x$, Heegner proved that all primes $p \equiv 5 \pmod 8$ are congruent numbers. His method also essentially proved that Gauss's list of imaginary quadratic fields with class number one is complete.

Heegner's idea was to use the theory of complex multiplication to construct special points on the modular curves $X_0(N)$. This led to the computation of Birch and Stephens on what are now called Heegner points, cf. a historical account by Birch [3]. To define these points, one needs an auxiliary imaginary quadratic field $K = \mathbb{Q}[\sqrt{-D}]$ with discriminant $-D < 0$. For simplicity we assume $D > 4$. We impose the *Heegner hypothesis*: *every prime factor $\ell | N$ is split in K*. It follows that

there exists an ideal $\mathcal{N}$ of $\mathcal{O}_K$, the ring of integers of K, such that

$$\mathcal{O}_K/\mathcal{N} \simeq \mathbb{Z}/N\mathbb{Z}.$$

Then the elliptic curves $\mathbb{C}/\mathcal{O}_K$ and $\mathbb{C}/\mathcal{N}^{-1}$ are naturally related by a cyclic isogeny of degree N, therefore define a point, denoted by $x(1)$, on $X_0(N)$. By the theory of complex multiplication, the point $x(1)$ is defined over the Hilbert class field H of K. By class field theory, the Galois group $\mathrm{Gal}(H/K)$ is isomorphic to the class group $\mathrm{Pic}(\mathcal{O}_K)$.

To have a complete analogue of §2.1, we assume that the factorization $N = N^+N^-$ in (2.1) satisfies the generalized Heegner hypothesis: *N^- is square-free and has even number of prime factors.* The Heegner hypothesis corresponds to the case $N^- = 1$. We need to use the Shimura curve X_{N^+,N^-} associated to quaternion algebra ramified precisely at prime factors of N^-, whose definition is given in the next paragraph. Then one may similary define a point $x(1) \in X_{N^+,N^-}(H)$.

Let $E/\mathbb{Q}$ be an elliptic curve with conductor N and f the associated weight two newform. Consider a modular parameterization

$$\phi : X_{N^+,N^-} \longrightarrow E.$$

If $N^- = 1$ we require this morphism to send (∞) to zero; in general one needs to normalize it in a certain way we will describe later on. Define

$$y(1) = \phi(x(1)) \in E(H), \quad y_K = \mathrm{tr}_{H/K}\, y(1) \in E(K). \tag{3.1}$$

The generalized Heegner hypothesis ensures that the global root number

$$\epsilon(E/K) = -1.$$

Then the Gross–Zagier formula for X_{N^+,N^-} is as follows, stated analogous to the formula (2.4):

Theorem 3.1. *We have*

$$\frac{L'(f/K,1)}{(f,f)} = \frac{1}{\sqrt{|D|}} \frac{\langle y_K, y_K\rangle}{\deg(\phi)}, \tag{3.2}$$

where (f,f) is the Petersson inner product (2.3), and $\langle y_K, y_K\rangle$ is the Néron–Tate height pairing over K.

This is proved by Gross–Zagier in [21] when $N^- = 1$, in general by S. Zhang in [58] and Yuan–Zhang–Zhang in [56].

To have an analogue to the formula (2.6), we choose a parameterization by the modular curve $X_0(N)$, already appearing in §2.1 (2.5):

$$f : X_0(N) \to E.$$

Let c be the constant such that $f^*\omega = c \cdot \omega_f$. Then an equivalent form of the formula (3.2) is

$$\frac{L'(E/K,1)}{|D|^{-1/2} \int_{E(\mathbb{C})} \omega \wedge \overline{\omega}} = \frac{1}{c^2} \frac{\deg(f)}{\deg(\phi)} \langle y_K, y_K\rangle. \tag{3.3}$$

In the rest of this section, we state the most general Gross–Zagier formula for GL_2 over a totally real field, following the book [56].

3.2. **Quaternions and Shimura curves.** Let F be a number field with adele ring $\mathbb{A} = \mathbb{A}_F$ and let $\mathbb{A}_f$ be the ring of finite adeles. Let Σ be a finite set of places of F. We then have *the quaternion algebra* $\mathbb{B}$ *over* $\mathbb{A}$ *with ramification set* $\Sigma(\mathbb{B}) := \Sigma$, i.e., the unique (up to isomorphism) $\mathbb{A}$-algebra, free of rank 4 as an $\mathbb{A}$-module, whose localization $\mathbb{B}_v := \mathbb{B} \otimes_{\mathbb{A}} F_v$ is isomorphic to $M_2(F_v)$ if $v \notin \Sigma$ and to the unique division quaternion algebra over F_v if $v \in \Sigma$. If $\#\Sigma$ is even then $\mathbb{B} = B \otimes_F \mathbb{A}$ for a quaternion algebra B over F. In this case, we call $\mathbb{B}$ a *coherent* quaternion algebra. If $\#\Sigma$ is odd, then $\mathbb{B}$ is not from any quaternion algebra over F. In this case, we call $\mathbb{B}$ an *incoherent* quaternion algebra (cf. Kudla's notion of *incoherent collections of quadratic spaces*, [33]).

Now assume that F is a totally real number field in the rest of this section and that $\mathbb{B}$ is a *totally definite* incoherent quaternion algebra over $\mathbb{A}$. Here "totally definite" means that Σ contains all archimedean places, i.e., $\mathbb{B}_\tau$ is the Hamiltonian quaternion for every archimedean place τ of F. We then have a (compactified) Shimura curve X_U over F indexed by open compact subgroups U of $\mathbb{B}_f^\times := (\mathbb{B} \otimes_{\mathbb{A}} \mathbb{A}_f)^\times$. For any embedding $\tau : F \hookrightarrow \mathbb{C}$, let $B(\tau)$ be the unique quaternion algebra over F with ramification set $\Sigma\backslash\{\tau\}$, and identify $\mathbb{B}_f$ with $B(\tau)_{\mathbb{A}_f}$ as an $\mathbb{A}_f$-algebra. Then the complex points of $X_{U,\tau} := X_U \times_{F,\tau} \mathbb{C}$ form a Riemann surface with a uniformization:

$$X_{U,\tau}(\mathbb{C}) \simeq B(\tau)^\times \backslash \mathcal{H}^\pm \times \mathbb{B}_f^\times / U \cup \{\text{cusps}\}, \quad \mathcal{H}^\pm := \mathbb{C}\backslash\mathbb{R}, \tag{3.4}$$

where $B(\tau)^\times$ acts on $\mathcal{H}^\pm$ through an isomorphism $B(\tau)_\tau \simeq M_2(\mathbb{R})$. The set $\{\text{cusps}\}$ is non-empty if and only if $F = \mathbb{Q}$ and $\Sigma = \{\infty\}$, in which case the Shimura curve X_U is a modular curve.

For later purposes, we will give a class of examples of Shimura curves which resemble the classical modular curve $X_0(N)$ with $\Gamma_0(N)$-level structure. Let $F = \mathbb{Q}$ and fix positive integers N^+ and N^- such that $(N^+, N^-) = 1$ and N^- is square-free with *even* number of prime factors (cf. the Shimura set X_{N^+,N^-} (2.2) when $\nu(N^-)$ is odd). We consider the indefinite quaternion algebra B over $\mathbb{Q}$ that is ramified precisely at all factors of N^-. Then the Shimura curve X_{N^+,N^-} is X_U where the compact open $U \subset B^\times(\mathbb{A}_f)$ is prescribed by

$$U_{N^+,N^-} = \prod_{\ell<\infty} U_\ell, \quad U_\ell = \begin{cases} \Gamma_0(N), & \ell | N^+, \\ \mathcal{O}_{B_\ell}^\times, & \ell \nmid N^+. \end{cases}$$

Equivalently, we may consider an Eichler order $\mathcal{O}_{B,N^+}$ in $\mathcal{O}_B$ with level N^+ and define

$$U_{N^+,N^-} = (\mathcal{O}_{B,N^+} \otimes \widehat{\mathbb{Z}})^\times. \tag{3.5}$$

In particular, if $N^- = 1$, the curve X_{N^+,N^-} is the classical modular curve $X_0(N^+)$, whose complex points are $\Gamma_0(N^+)\backslash\mathcal{H}$ together with cusps.

For any two open compact subgroups $U_1 \subset U_2$ of $\mathbb{B}_f^\times$, one has a natural surjective morphism $\pi_{U_1,U_2} : X_{U_1} \to X_{U_2}$. Let X be the projective limit of the system $\{X_U\}_{U \subset \mathbb{B}_f^\times}$. It is a regular scheme over F, locally noetherian but not of finite type.

On the curve X_U, there is a distinguished class $\xi_U \in \mathrm{Pic}(X_U)_{\mathbb{Q}}$ with degree equal to one on every connected component of X_U. In the case of the modular curve $X_0(N)$, one may work with the divisor class of the cusp (∞). In general, one uses a normalized Hodge class, cf. [56, §3.1.3] for details.

3.3. **Abelian varieties parametrized by Shimura curves.** We will be interested in the isogeny factors of the Jacobian of Shimura curves. For a *simple* abelian variety A over F, we say that A *is parametrized by* X if there is a non-constant morphism $X_U \to A$ over F for some U. If A is parametrized by X, then A *is of strict GL(2)-type* in the sense that

$$M = \mathrm{End}^0(A) := \mathrm{End}_F(A) \otimes_{\mathbb{Z}} \mathbb{Q}$$

is a field and $\mathrm{Lie}(A)$ is a free module of rank one over $M \otimes_{\mathbb{Q}} F$ by the induced action.

We now define a $\mathbb{Q}$-vector space:

$$\pi_A = \mathrm{Hom}^0_{\xi}(X, A) := \varinjlim_{U} \mathrm{Hom}^0_{\xi_U}(X_U, A),$$

where $\mathrm{Hom}^0_{\xi_U}(X_U, A)$ consists of morphisms in $\mathrm{Hom}_F(X_U, A) \otimes_{\mathbb{Z}} \mathbb{Q}$ with the following property: if ξ_U is represented by a divisor $\sum_i a_i x_i$ on $X_{U,\overline{F}}$, then $f \in \mathrm{Hom}_F(X_U, A) \otimes_{\mathbb{Z}} \mathbb{Q}$ is in π_A if and only if $\sum_i a_i f(x_i) = 0$ in $A(\overline{F})_{\mathbb{Q}}$. For example, if A is an elliptic curve over $\mathbb{Q}$ of conductor N, then there exists a non-constant morphism $\phi : X_0(N) \to A$ that maps the cusp (∞) to $0 \in A$. Then such a modular parameterization ϕ defines an element in π_A.

The curve X admits a $\mathbb{B}_f^\times$-action by Hecke correspondence. Making $\mathbb{B}_\infty^\times$ act trivially, we have a $\mathbb{B}^\times$-action. Then the space π_A admits a natural $\mathbb{B}^\times$-module structure. Note that π_A also admits an M-action. In [56] we say that π_A is an *automorphic representation of* $\mathbb{B}^\times$ *over* $\mathbb{Q}$, if the Jacquet–Langlands transfer of $\pi_{A,\mathbb{C}}$ to $\mathrm{GL}_2(\mathbb{A})$ is automorphic. It is proved in [56] that

$$\mathrm{End}_{\mathbb{B}^\times}(\pi_A) = M$$

and that π_A has a decomposition as a restricted tensor product

$$\pi_A = \bigotimes_M \pi_v,$$

where π_v is an absolutely irreducible representation of $\mathbb{B}_v^\times$ over M. Denote by ω_A the central character of π_A.

Remark 15. We note that the way we form the representation space π_A is analogous to the passage from a classical holomorphic modular form to the associated automorphic representation: if we have a Hecke eigenform f which is a classical holomorphic modular form of a certain weight, then one may take all Hecke translations to form an automorphic representation of $\mathrm{GL}_2(\mathbb{A})$. For comparison, for a reductive group G over F, an automorphic form is a function $f : G(F)\backslash G(\mathbb{A}) \to \mathbb{C}$ with values in $\mathbb{C}$; an element in π_A can be viewed as an "automorphic form" on a Shimura curve with values in an abelian variety A.

We may then define the L-function attached to the M-representation π_A

$$L(s, \pi_A) = \prod_v L_v(s, \pi_v) \in M \otimes_{\mathbb{Q}} \mathbb{C}.$$

It has an analytic continuation to an entire function of $s \in \mathbb{C}$. This is done by invoking the Jacquet–Langlands transfer of π_A to $\mathrm{GL}_2(\mathbb{A})$. We will take the complete L-series using suitable Γ-functions at archimedean places. One may think of $L(s, \pi_A)$ as a tuple $L(s, \pi_A, \iota)$ indexed by $\iota \in \mathrm{Hom}(M, \mathbb{C})$.

Remark 16. There is also a motivic definition of the L-function of A. For example, when A is an elliptic curve over $\mathbb{Q}$, the L-function $L(s, A)$ is the L-function defined by (1.6) completed by a certain Γ-function.

3.4. **Duality.** If A is parametrized by a Shimura curve X, then the dual abelian variety $A^\vee$ is also parametrized by X. Then the endomorphism field $M^\vee = \mathrm{End}^0(A^\vee)$ is canonically isomorphic to M by sending a homomorphism $m : A \to A$ to its dual $m^\vee : A^\vee \to A^\vee$.

One may define a perfect $\mathbb{B}^\times$-invariant pairing

$$(\cdot,\cdot) : \quad \pi_A \times \pi_{A^\vee} \longrightarrow M$$

as follows. Let $f_{1,U} \in \mathrm{Hom}^0_{\xi_U}(X_U, A), f_{2,U} \in \mathrm{Hom}^0_{\xi_U}(X_U, A^\vee)$. By Albanese functoriality, $f_{1,U}$ induces an element in $\mathrm{Hom}(J_U, A)$, still denoted by $f_{1,U}$, where J_U is the Jacobian of X_U. Similarly we have $f_{2,U} \in \mathrm{Hom}(J_U, A^\vee)$. Let $f^\vee_{2,U} : A \to J_U$ be the dual of $f_{2,U}$, where we identify $J_U^\vee \simeq J_U$. We then define

$$(f_1, f_2) = \frac{(f_{1,U} \circ f^\vee_{2,U})}{\mathrm{vol}(X_U)} \in M,$$

where $f_{1,U} \circ f^\vee_{2,U} \in \mathrm{End}^0(A) = M$, and the volume factor is defined

$$\mathrm{vol}(X_U) := \int_{X_U(\mathbb{C})} \frac{dxdy}{2\pi y^2}.$$

It takes value in $\mathbb{Q}$. This is independent of the choice of the compact open subgroup $U \subset \mathbb{B}_f^\times$. It follows that $\pi_{A^\vee}$ is dual to π_A as M-representations of $\mathbb{B}^\times$.

Remark 17. We note that the pairing $(\cdot,\cdot)$ here plays the role of Petersson inner product for an automorphic representation. This becomes more obvious when we compare the Gross–Zagier formula with the Waldspurger formula (cf. (3.2) and (2.4)).

Remark 18. When A is an elliptic curve, we have $M = \mathbb{Q}$ and π_A is self-dual. For any morphism $f \in \pi_A$ represented by a direct system $\{f_U\}_U$, we have

$$(f, f) = \mathrm{vol}(X_U)^{-1} \deg f_U.$$

Here $\deg f_U$ denotes the degree of the finite morphism $f_U : X_U \to A$, usually referred as the "modular degree", an invariant that contains important arithmetic information of the elliptic curve A and has received wide attention.

3.5. **Height pairing.** The Néron–Tate height pairing is a $\mathbb{Q}$-bilinear non-degenerate pairing

$$\langle\cdot,\cdot\rangle_{\mathrm{NT}} : A(\overline{F})_\mathbb{Q} \times A^\vee(\overline{F})_\mathbb{Q} \longrightarrow \mathbb{R}.$$

The field $M = \mathrm{End}^0(A)$ acts on $A(\overline{F})_\mathbb{Q}$ by definition, and acts on $A^\vee(\overline{F})_\mathbb{Q}$ through the duality. Then one may define an M-bilinear pairing, called *an M-linear Néron–Tate height pairing*

$$\langle\cdot,\cdot\rangle_M : A(\overline{F})_\mathbb{Q} \otimes_M A^\vee(\overline{F})_\mathbb{Q} \longrightarrow M \otimes_\mathbb{Q} \mathbb{R}$$

such that

$$\langle\cdot,\cdot\rangle_{\mathrm{NT}} = \mathrm{tr}_{M\otimes\mathbb{R}/\mathbb{R}}\langle\cdot,\cdot\rangle_M.$$

3.6. **CM points.** To introduce CM points, we let E/F be a totally imaginary quadratic extension, with a fixed embedding $\mathbb{A}_E \hookrightarrow \mathbb{B}$ over $\mathbb{A}$. Then $\mathbb{A}_E^\times$ acts on X by the right multiplication via $\mathbb{A}_E^\times \hookrightarrow \mathbb{B}^\times$. Let $X^{E^\times}$ be the subscheme of X of fixed points of X under $E^\times$. The scheme $X^{E^\times}$ is defined over F. By the theory of complex multiplication, every point in $X^{E^\times}(\overline{F})$ is defined over E^{ab} and the Galois action of $\mathrm{Gal}(E^{ab}/E)$ is given by the Hecke action under the reciprocity law.

Fix a base point $P \in X^{E^\times}(E^{\mathrm{ab}})$ given by a system of points $P_U \in X_U(E^{\mathrm{ab}})$ indexed by $U \subset \mathbb{B}_f^\times$. For $\tau \in \mathrm{Hom}(F, \mathbb{C})$, via the complex uniformization of $X_{U,\tau}(\mathbb{C})$ by (3.4), the point P_U can be chosen to be represented by the double coset of $[z_0, 1]_U$, where $z_0 \in \mathcal{H}$ is the unique fixed point of $E^\times$ in $\mathcal{H}$ via the action induced by an embedding $E \hookrightarrow B(\tau)$.

Let A be an abelian variety over F parametrized by X with $M = \mathrm{End}^0(A)$. Let $\chi : \mathrm{Gal}(E^{\mathrm{ab}}/E) \to L^\times$ be a character of *finite order*, valued in a finite extension L of M. For any $f \in \pi_A$, the image $f(P)$ is a well-defined point in $A(E^{\mathrm{ab}})_{\mathbb{Q}}$. Consider the integration

$$\mathscr{P}_\chi(f) = \int_{\mathrm{Gal}(E^{\mathrm{ab}}/E)} f(P^\tau) \otimes_M \chi(\tau) d\tau \ \in A(E^{\mathrm{ab}})_{\mathbb{Q}} \otimes_M L,$$

where P^τ is the Galois action of τ on P, the Haar measure on $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ has total volume 1. It is essentially a finite sum, and it is easy to see that

$$P_\chi(f) \in A(\chi) := (A(E^{\mathrm{ab}})_{\mathbb{Q}} \otimes_M L_\chi)^{\mathrm{Gal}(E^{\mathrm{ab}}/E)}.$$

Here L_χ denotes the M-vector space L with the action of $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ given by the multiplication by the character χ. For $\mathscr{P}_\chi(f) \neq 0$, a necessary condition is that the central character ω_A of π_A should be compatible with χ:

$$\omega_A \cdot \chi|_{\mathbb{A}^\times} = 1.$$

From now on we assume this compatibility. Consider the L-vector space $\mathrm{Hom}_{\mathbb{A}_E^\times}(\pi_A \otimes \chi, L)$. The map $f \mapsto \mathscr{P}_\chi(f)$ defines an element:

$$\mathscr{P}_\chi \in \mathrm{Hom}_{\mathbb{A}_E^\times}(\pi_A \otimes \chi, L) \otimes_L A(\chi),$$

where we recall that the Hom space is at most 1-dimensional.

3.7. **Local canonical invariant form.** We need a local canonical invariant form in the space

$$\mathrm{Hom}_{E_v^\times}(\pi_{A,v} \otimes \chi_v, L) \otimes \mathrm{Hom}_{E_v^\times}(\pi_{A,v}^\vee \otimes \chi_v^{-1}, L).$$

This is almost the same as the one $\alpha^\natural$ appearing in the Waldspurger formula except we need to make it defined over L. Let $(\cdot,\cdot)_v : \pi_{A,v} \times \pi_{A,v}^\vee \to M$ be the canonical pairing. We define α_v formally by

$$\alpha_v(f_1, f_2) = \int_{E_v^\times/F_v^\times} (\pi_v(t) f_1, f_2)_v \ \chi_v(t) dt, \qquad f_1 \in \pi_v, \ f_2 \in \widetilde{\pi}_v. \tag{3.6}$$

and normalize it by

$$\alpha_v^\natural(f_1, f_2) = \frac{L(1, \eta_v) L(1, \pi_v, \mathrm{Ad})}{\zeta_{F_v}(2) L(\frac{1}{2}, \pi_v, \chi_v)} \alpha_v(f_1, f_2) \tag{3.7}$$

More precisely, we choose measures so that $\mathrm{vol}(E_v^\times/F_v^\times) \in \mathbb{Q}$ and take an embedding $\iota : L \hookrightarrow \mathbb{C}$ and define the above integral with value in $\mathbb{C}$. We then show that, for

all places v, the value of $\alpha_v^\natural(f_1, f_2)$ lies in L, and does not depend on the choice of the embedding ι.

3.8. **Gross–Zagier formula.** Let $\pi_{A,E}$ denotes the base change of π_A to E. View χ as a character of $E^\times\backslash\mathbb{A}_E^\times$ via the reciprocity law $E^\times\backslash\mathbb{A}_E^\times \longrightarrow \mathrm{Gal}(E^{\mathrm{ab}}/E)$ which maps uniformizers to geometric Frobenii. Define the L-function

$$L(s, \pi_A, \chi) = L(s, \pi_{A,E} \otimes \chi).$$

For example, if A is an elliptic curve over $F/\mathbb{Q}$ and χ is trivial, then the L-function is the Hasse–Weil L-function associated to the base change A/E.

We identify the contragredient $\widetilde{\pi}_A = \pi_{A^\vee}$ by the duality map

$$(\cdot,\cdot) : \pi_A \times \pi_{A^\vee} \longrightarrow M.$$

Then we have the following Gross–Zagier formula on Shimura curves ([56, Theorem 1.2]), parallel to the Waldspuger formula (2.10).

Theorem 3.2. *For any $f_1 = \otimes_v f_{1,v} \in \pi_A$ and $f_2 = \otimes_v f_{2,v} \in \pi_{A^\vee}$, we have*

$$\frac{\langle \mathscr{P}_\chi(f_1),\ \mathscr{P}_{\chi^{-1}}(f_2)\rangle_L}{(f_1, f_2)} = \frac{1}{4}\frac{\zeta_F(2)L'(1/2, \pi_A, \chi)}{L(1,\eta)^2 L(1, \pi_A, \mathrm{Ad})}\prod_v \frac{\alpha^\natural(f_{1,v}, f_{2,v})}{(f_{1,v}, f_{2,v})_v} \tag{3.8}$$

as an identity in $L \otimes_{\mathbb{Q}} \mathbb{C}$. Here $\langle\cdot,\cdot\rangle_L : A(\chi) \times A^\vee(\chi^{-1}) \to L \otimes_{\mathbb{Q}} \mathbb{R}$ is the L-linear Néron–Tate height pairing induced by the M-linear Néron–Tate height pairing $\langle\cdot,\cdot\rangle_M$ between $A(\overline{F})$ and $A^\vee(\overline{F})$.

Note that, in contrast to the Waldspurger formula, the appearance of $L(1,\eta)^2$ is caused by the different choice of the measure on $\mathbb{A}_E^\times/\mathbb{A}_F^\times$.

A necessary condition for $\mathscr{P}_\chi \neq 0$ is that $\mathrm{Hom}_{\mathbb{A}_E^\times}(\pi_A \otimes \chi, L) \neq 0$. By the theorem of Saito–Tunnell (cf. §2), the space $\mathrm{Hom}_{\mathbb{A}_E^\times}(\pi_A \otimes \chi, L)$ is at most one-dimensional, and it is one-dimensional if and only if the ramification set $\Sigma(\mathbb{B})$ of $\mathbb{B}$ is equal to the set

$$\Sigma(A, \chi) := \{ \text{ places } v \text{ of } F :\ \epsilon(1/2, \pi_{A,v}, \chi_v) \neq \chi_v(-1)\eta_v(-1)\}.$$

In that case, since $\#\Sigma(\mathbb{B})$ is odd, the global root number

$$\epsilon(1/2, \pi_A, \chi) = -1$$

and hence $L(1/2, \pi_A, \chi) = 0$. If $\Sigma(\mathbb{B}) \neq \Sigma(A, \chi)$, the vector space $\mathrm{Hom}_{\mathbb{A}_E^\times}(\pi_A \otimes \chi, L)$ is zero and thus both sides of the formula (3.8) are zero.

Back to the Gross-Zagier formula (3.2) or (3.3), the character $\chi = 1$ and the generalized Heegner condition implies that

$$\Sigma(A, \chi) = \{\ell : \ell | N^-\} \cup \{\infty\}.$$

The relevant Shimura curves are then the curves X_{N^+,N^-}.

3.9. **Higher rank cases: the arithmetic Gan–Gross–Prasad conjecture.** There is also a natural generalization of the Gross–Zagier formula to higher rank, called the arithmetic Gan–Gross–Prasad conjecture (and its refinement), formulated in [9], [62], [63]. A relative trace formula approach has also been proposed by the author in [63]. However, in contrast to the generalization of Waldspurger formula, there are fewer results in this direction. One obstruction is the still unproven "arithmetic fundamental lemma", which will be stated below. This is only known in the lower rank cases ([63]) or in some special cases ([40]). Moreover, one needs

also an arithmetic version of "smooth matching" for the relevant Shimura varieties at all places.

3.10. **The arithmetic fundamental lemma.** We state the conjectural arithmetic fundamental lemma [63]. Fix a prime p and consider the following data:

- $E = \mathbb{Q}_{p^2}$, unramified quadratic extension of $\mathbb{Q}_p$.
- V: Hermitian space of E-dimension $n+1$, with the Hermitian pairing $(\cdot,\cdot) : V \times V \to E$.
- $u \in V$, $(u,u) = 1$.
- For *regular* $g \in U(V)$, define L_g to be the $\mathbb{Z}_{p^2}$-lattice in V generated by $u, gu, ..., g^n u$. Here "regular" means that L_g has full rank.
- $\tau : V \to V$: an E-linear involution (depending on g) characterized by $\tau(g^i u) = g^{-i} u$.

For comparison, we also recall the Jacquet–Rallis fundamental lemma, which is a fundamental ingredient to proving the generalized Waldspurger formulas in §2. We consider two type of lattices: *self-dual lattices* and *conjugate-invariant lattices*:

$$\mathscr{L}^{self} := \{\Lambda \subset V : \mathbb{Z}_{p^2}\text{-lattice}, \Lambda^* = \Lambda\},$$

where $\Lambda^* = \{v \in V : (v,\Lambda) \subset \mathbb{Z}_{p^2}\}$, and

$$\mathscr{L}^{conj} := \{\Lambda \subset V : \mathbb{Z}_{p^2}\text{-lattice}, \Lambda^\tau = \Lambda\}.$$

Then the Jacquet-Rallis fundamental lemma is a family of identities relating weighted counts of the two type lattices, indexed by (regular) $g \in U(V)$:

$$\sum_{\{\Lambda \in \mathscr{L}^{conj} : g\Lambda = \Lambda, L_g \subset \Lambda \subset L_g^*\}} (-1)^{\ell(\Lambda)} = \sum_{\{\Lambda \in \mathscr{L}^{self} : g\Lambda = \Lambda, L_g \subset \Lambda \subset L_g^*\}} 1, \tag{3.9}$$

where $\ell(\Lambda)$ is the length of $\mathbb{Z}_{p^2}$-module Λ/L_g. This form of the statement was proved by Z. Yun ([57]) for function fields of characteristic p¿n. In an appendix to [57], J. Gordon showed that this implies the characteristic zero version when the residue characteristic p is sufficiently large (compared to n).

To state the arithmetic fundamental lemma, we introduce the set of a third type lattices which we call *almost self-dual lattices*:

$$\mathscr{L}^{A-self} := \{\Lambda \subset V : \mathbb{Z}_{p^2}\text{-lattice}, p\Lambda \subset \Lambda^* \subset \Lambda\},$$

i.e., Λ/Λ^* is killed by p. The almost-self dual lattices appear in parameterizing irreducible components of the supersingular locus of unitary Shimura variety (type $U(n,1)$) in characteristic p. More precisely, by the Bruhat–Tits stratification on the Rapoport–Zink space of unitary type $(n,1)$ ([51]), one may associate a stratum to each almost self-dual lattice Λ. Then one may define a certain arithmetic intersection multiplicity $\mathrm{mult}(\Lambda)$ along this stratum (cf. [40]).

Then the conjectural arithmetic fundamental lemma [63] can be stated as a family of identities relating weighted counts of the two type lattices, indexed by (regular) $g \in U(V)$.

Conjecture 3.3. *If $\mathscr{L}^{self}$ is empty, then we have*

$$\sum_{\{\Lambda \in \mathscr{L}^{conj} : g\Lambda = \Lambda, L_g \subset \Lambda \subset L_g^*\}} (-1)^{\ell(\Lambda)} \ell(\Lambda) = \sum_{\{\Lambda \in \mathscr{L}^{A-self} : g\Lambda = \Lambda, L_g \subset \Lambda \subset L_g^*\}} \mathrm{mult}(\Lambda).$$

4. Kolyvagin conjecture and the structure of Selmer groups

4.1. **Shimura curves and Shimura sets X_{N^+,N^-}.** Let $N = N^+N^-$ be a factorization of a positive integer N such that $(N^+, N^-) = 1$ and N^- is square-free. Then we have defined a Shimura set (a Shimura curve over $\mathbb{Q}$, resp.) by (2.2) (by (3.4), (3.5), resp.) when $\nu(N^-)$ is odd (even, resp.), denoted by X_{N^+,N^-}. Its complex points can be uniformly described as (possibly also joint with cusps when $N^- = 1$)

$$X_{N^+,N^-}(\mathbb{C}) \simeq B^\times(\mathbb{Q})\backslash B^\times(\mathbb{A})/\mathbb{R}_+^\times U_\infty \cdot (\mathcal{O}_{B,N^+} \otimes \widehat{\mathbb{Z}})^\times,$$

where

- B is the unique-up-to-isomorphism quaternion algebra B over $\mathbb{Q}$ ramified exactly at prime factors of N^- if $\nu(N^-)$ is even, and ∞ if $\nu(N^-)$ is odd.
- U_∞ is the maximal connected compact subgroup of $B_\infty^\times$.
- $\mathcal{O}_{B,N^+}$ is an Eichler order of level N^+.

4.2. **Heegner points on Shimura curves.** Let $K = \mathbb{Q}[\sqrt{-D}]$ be an imaginary quadratic extension with discriminant $-D < -4, (D, N) = 1$. This determines a unique factorization $N = N^+N^-$ as in (2.1). N^- is assume to be square-free and $\nu(N^-)$ *even* so that we have a Shimura curve X_{N^+,N^-}. For simplicity we write $X = X_{N^+,N^-}$.

For $n \in \mathbb{Z}_{>0}$ we let $\mathcal{O}_{K,n} = \mathbb{Z} + n\mathcal{O}_K$ be the order of $\mathcal{O}_K$ with conductor n. Denote by $K[n]$ the ring class field of K with conductor n, characterized by the reciprocity law

$$\mathrm{rec} : \mathrm{Gal}(K[n]/K) \simeq K^\times\backslash\widehat{K}^\times/\widehat{\mathcal{O}}_{K,n}^\times,$$

In particular, $K[1] = H_K$ is the Hilbert class field of K, and $K[n]$ is an abelian extension of K only ramified at primes above factors of n, with Galois group $\mathrm{Gal}(K[n]/K) \simeq \mathrm{Pic}(\mathcal{O}_{K,n})$.

In §3 (3.1), we have defined the Heegner point $x(1) \in X(H_K), y(1) \in E(H_K), y_K \in E(K)$. There exists a collection of points $x(n)$ on X defined over $K[n]$. They are sometimes called higher Heegner points, to be distinguished with $x(1)$. We describe them in the case of modular curve $X = X_0(N)$, i.e., when $N^- = 1$. Let $\mathcal{N}$ be an ideal such that $\mathcal{O}_K/\mathcal{N} \simeq \mathbb{Z}/N\mathbb{Z}$. Then the ideal $\mathcal{N}_n = \mathcal{N} \cap \mathcal{O}_{K,n}$ is an invertible $\mathcal{O}_{K,n}$-module. Then the elliptic curves $\mathbb{C}/\mathcal{O}_{K,n}$ and $\mathbb{C}/\mathcal{N}_n^{-1}$ are naturally related by a cyclic isogeny of degree N, therefore define a point, denoted by $x(n)$, on $X_0(N)$. By the theory of complex multiplication, the point $x(n)$ is defined over the ring class field $K[n]$.

Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Then it is parameterized by the Shimura curve X_{N^+,N^-}:

$$\phi : X_{N^+,N^-} \longrightarrow E. \tag{4.1}$$

Then one may define points

$$y(n) = \phi(x(n)) \in E(K[n]).$$

The definition of $y(n)$ depends on ϕ. From now on we will assume that ϕ is an optimal parameterization.

4.3. **Kolyvagin cohomology classes.** We first define Kolyvagin's derived Heegner points and cohomology classes. Let p be a prime.

Definition 4.1. (1) *For a prime $\ell \nmid N$, the p-divisibility of ℓ is defined as*

$$M(\ell) = v_p(gcd(a_\ell, \ell+1)),$$

where $a_\ell = \ell + 1 - \#E(\mathbb{F}_\ell)$. If n is square-free and $(n, N) = 1$, we define

$$M(n) = \min_{\ell | n} M(\ell)$$

We also set

$$M(1) = \infty.$$

(2) *A prime $\ell \nmid NDp$ is a Kolyvagin prime if it is inert in K and $M(\ell) > 0$.*

Now let Λ be the set of square-free products of Kolyvagin primes and set

$$\Lambda_r = \{n \in \Lambda : \nu(n) = r\}, \quad \Lambda(M) = \{n \in \Lambda : M(n) \geqslant M\},$$

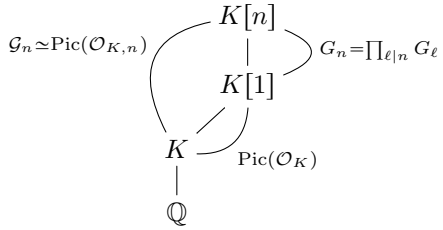
where

$$\nu(n) = \#\{\ell : \ell | n\}. \tag{4.2}$$

For $n \in \Lambda$, we denote $G_n = \mathrm{Gal}(K[n]/K[1])$ and $\mathcal{G}_n = \mathrm{Gal}(K[n]/K)$ for $n \in \Lambda$. Then we have a canonical isomorphism:

$$G_n = \prod_{\ell | n} G_\ell,$$

where the group $G_\ell = \mathrm{Gal}(K[\ell]/K[1])$ is cyclic of order $\ell+1$. We have the following diagram:



Fix a generator σ_ℓ of G_ℓ for each prime $\ell \in \Lambda$. We define the Kolyvagin derivative operator

$$\mathbb{D}_\ell := \sum_{i=1}^{\ell+1} i\sigma_\ell^i \in \mathbb{Z}[G_\ell],$$

and

$$\mathbb{D}_n := \prod_{\ell | n} \mathbb{D}_\ell \in \mathbb{Z}[G_n].$$

Fix a set $\mathscr{G}$ of representative of $\mathcal{G}_n/G_n$. Then we define the derived Heegner point

$$P(n) := \sum_{\sigma \in \mathscr{G}} \sigma(\mathbb{D}_n y(n)) \in E(K[n]). \tag{4.3}$$

From the short exact sequence

$$0 \longrightarrow E[p^M] \longrightarrow E \xrightarrow{\times p^M} E \longrightarrow 0\ ,$$

we have an induced Kummer map

$$E(K) \otimes \mathbb{Z}/p^M \longrightarrow H^1(K, E[p^M]).$$

We have a commutative diagram of Kummer maps:

$$\begin{array}{ccc} E(K)\otimes\mathbb{Z}/p^M & \longrightarrow & H^1(K,E[p^M]) \\ \downarrow & & \downarrow{\scriptstyle \mathrm{Res}} \\ (E(K[n])\otimes\mathbb{Z}/p^M)^{\mathcal{G}_n} & \longrightarrow & H^1(K[n],E[p^M])^{\mathcal{G}_n} \end{array}$$

Now assume that $\overline{\rho}_{E,p}$ is surjective. Then for $n\in\Lambda$, we have ([16, Lemma 4.3])

$$E[p^M](K[n])=0.$$

Hence the restriction map

$$\mathrm{Res}: \qquad H^1(K,E[p^M])\xrightarrow{\simeq} H^1(K[n],E[p^M])^{\mathcal{G}_n}$$

is an isomorphism. When $M\leqslant M(n)$, the derived point $P(n)$ defines a $\mathcal{G}_n$-invariant element in $E(K[n])\otimes\mathbb{Z}/p^M$. Hence the Kummer image of $P(n)$ in $H^1(K[n],E[p^M])$ descends to a cohomology class denoted by

$$c_M(n)\in H^1(K,E[p^M]). \tag{4.4}$$

When $n=1$, we find that

$$y_K=P(1)=\mathrm{tr}_{K[1]/K}y(1)\in E(K), \tag{4.5}$$

and $c_M(1)$ is the Kummer image of y_K.

Definition 4.2. *The* mod p^M *Kolyvagin system is the collection of cohomology classes*

$$\kappa_{p^M}=\{c_M(n)\in H^1(K,E[p^M]) : n\in\Lambda(M)\}. \tag{4.6}$$

We also write

$$\kappa_{p^\infty}=\{c_M(n)\in H^1(K,E[p^M]) : n\in\Lambda, M(n)\geqslant M\}. \tag{4.7}$$

Remark 19. One could also describe the action of the complex conjugation on all of the classes $c(n)$. Let $\epsilon=\epsilon(E/K)\in\{\pm1\}$ be the root number $E/\mathbb{Q}$. Define

$$\epsilon_\nu=\epsilon\cdot(-1)^{\nu+1}\in\{\pm1\}. \tag{4.8}$$

Then the class $c(n)$ lies in the $\epsilon_{\nu(n)}$-eigenspace under complex conjugation ([16, Prop. 5.4]):

$$c(n)\in H^1(K,E[p^M])^{\epsilon_{\nu(n)}}.$$

4.4. **Kolyvagin conjecture.** Define $\mathscr{M}(n)\in\mathbb{Z}_{\geqslant0}\cup\{\infty\}$ to be the divisibility of the derived Heegner point $P(n)$, i.e., the maximal $\mathscr{M}\in\mathbb{Z}_{\geqslant0}\cup\{\infty\}$ such that

$$P(n)\in p^{\mathscr{M}}E(K[n])$$

(or we may write this as $p^{\mathscr{M}(n)}||P(n)$). Let $\mathscr{M}_r$ be the minimal $\mathscr{M}(n)$ for all $n\in\Lambda_r$. Then in [32] Kolyvagin shows that for all $r\geqslant0$:

$$\mathscr{M}_r\geqslant\mathscr{M}_{r+1}\geqslant0. \tag{4.9}$$

Define

$$\mathscr{M}_\infty=\lim_{r\to\infty}\mathscr{M}_r \tag{4.10}$$

as the minimum of $\mathscr{M}_r$ for varying $r\geqslant0$.

Then the conjecture of Kolyvagin [32, Conj. A] (generalized to Shimura curves) asserts that

Conjecture 4.3. *Let $E/\mathbb{Q}$ be an elliptic curve with conductor N. Let $K = \mathbb{Q}[\sqrt{-D}]$ be an imaginary quadratic extension with discriminant $-D < -4, (D, N) = 1$. Assume that the residue representation $\overline{\rho}_{E,p}$ is surjective. Then the Kolyvagin system (4.7)*

$$\kappa_{p^\infty} \neq \{0\},$$

or, equivalently, $\mathscr{M}_\infty < \infty$.

Definition 4.4. *The vanishing order of $\kappa_{p^\infty} \in \mathbb{Z}_{\geqslant 0} \cup \{\infty\}$ is defined as*

$$\operatorname{ord} \kappa_{p^\infty} := \min\{\nu(n) : n \in \Lambda, M \leqslant M(n), c_M(n) \neq 0\},$$

(∞ if all $c_M(n)$ vanish.) Similarly one may define the vanishing order of κ_{p^M}

$$\operatorname{ord} \kappa_{p^M} := \min\{\nu(n) : n \in \Lambda(M), c_M(n) \neq 0\}.$$

Clearly we have

$$\operatorname{ord} \kappa_{p^\infty} \leqslant ... \leqslant \operatorname{ord} \kappa_{p^M} \leqslant \operatorname{ord} \kappa_{p^{M-1}} \leqslant ... \leqslant \operatorname{ord} \kappa_p.$$

The conjecture 4.3 is equivalent to

$$\operatorname{ord} \kappa_{p^\infty} < \infty.$$

Remark 20. Clearly we have

$$\operatorname{ord} \kappa_{p^\infty} = 0 \iff y_K \in E(K) \text{ is non-torsion.} \tag{4.11}$$

Hence the conjecture 4.3 holds trivially if the Heegner point y_K is non-torsion.

It is important to know the value $\mathscr{M}_\infty$. Indeed, one may refine the conjecture to predict $\mathscr{M}_\infty$ in terms of local Tamagawa numbers, based on the following ingredients when $r_{an}(E/K) = 1$:

- Birch–Swinnerton-Dyer conjecture (the refined formula (1.13)).
- Gross–Zagier formula for X_{N^+,N^-} (3.3).
- Kolyvagin's theorem that $\#Ш(E/K)[p^\infty] = p^{2(\mathscr{M}_0 - \mathscr{M}_\infty)}$.
- Ribet–Takahashi's comparison of modular degrees in (3.3):

$$\left|\frac{\deg(f)}{\deg(\phi)}\right|_p = \left|\prod_{\ell | N^-} c_\ell\right|_p,$$

  for a prime p where $\overline{\rho}_{E,p}$ is irreducible.
- Mazur's theorem on the Manin constant c asserts that $p|c \implies p^2|4N$.

Then the refined Kolyvagin conjecture can be stated as follows.

Conjecture 4.5. *Assume that $\overline{\rho}_{E,p}$ is surjective. Then*

$$\mathscr{M}_\infty = v_p(\prod_{\ell | N^+} c_\ell).$$

In [67] we prove the refined Kolyvagin conjecture for $p \geqslant 5$ satisfying certain local conditions. For application to the Birch–Swinnerton-Dyer conjecture for $E/\mathbb{Q}$, these local conditions are mild by a careful choice of the auxiliary K and the Shimura curve.

Theorem 4.6. *Let E be a semistable elliptic curve with conductor N. Let p, K satisfy the following conditions*

(1) *E has good ordinary reduction at $p \geqslant 5$.*

(2) $\overline{\rho}_{E,p}$ *is surjective.*
(3) *If* $\ell \equiv \pm 1 \mod p$ *and* $\ell | N$, *then* $\overline{\rho}_{E,p}$ *is ramified at* ℓ.
(4) $\nu(N^-)$ *is even and* $\overline{\rho}_{E,p}$ *is ramified at all* $\ell | N^+$.

We have

$$\kappa_p = \{c_1(n) \in H^1(K, E[p]) : n \in \Lambda\} \neq \{0\},$$

or equivalently $\mathscr{M}_\infty = 0$ *(or equivalently,* $p \nmid P(n)$ *for some* $n \in \Lambda$*). In particular,* $\kappa_{p^\infty} \neq 0$.

Remark 21. Under the hypothesis (4), we have $v_p(\prod_{\ell | N^+} c_\ell) = 0$.

Remark 22. The refined Kolyvagin conjecture—combined with the five items above—implies the p-part of the B-SD formula for E/K when $r_{an}(E/K) = 1$. By a suitable choice of an auxiliary K using the result of [7] or [38], we may then prove Theorem 1.6.

4.5. **The structure of Selmer group.** Under the irreducibility of $\overline{\rho}_{E,p}$, we have an injection

$$H^1(K, E[p^M]) \hookrightarrow H^1(K, E[p^{M+M'}]), \quad M, M' \geqslant 1.$$

The group $H^1(K, E[p^M])$ can be viewed as the kernel of the multiplication by p^M on $H^1(K, E[p^{M+M'}])$. If an element $c \in H^1(K, E[p^{M+M'}])$ is killed by p^M, we will view c as an element in $H^1(K, E[p^M])$. More generally, we have a short exact sequence:

$$0 \longrightarrow H^1(K, E[p^M]) \longrightarrow H^1(K, E[p^\infty]) \xrightarrow{p^M} H^1(K, E[p^\infty]).$$

In this way we naturally view $c(n) \in H^1(K, E[p^M])$ as an element of $H^1(K, E[p^\infty])$.

Let $\mathrm{Sel}_{p^\infty}^{\pm}(E/K)$ be the two eigenspaces under the action of the nontrivial element in $\mathrm{Gal}(K/\mathbb{Q})$. Let $r_p^{\pm}(E/\mathbb{Q})$ be the corresponding $\mathbb{Z}_p$-corank.

Theorem 4.7 (Kolyvagin, [31], [32])**.** *Let* $E/\mathbb{Q}$ *be an elliptic curve of conductor* N. *Consider* $K = \mathbb{Q}[\sqrt{-D}]$ *with discriminant* $-D < -4$, $(D, N) = 1$, $\nu(N^-)$ *even. Let* $p \geqslant 3$ *be a prime where* $\overline{\rho}_{E,p}$ *is surjective and* $p \nmid ND$. *Assume that Conjecture 4.3 holds, i.e.,*

$$\kappa_{p^\infty} \neq 0,$$

or equivalently, $\mathscr{M}_\infty < \infty$. *Then we have*

(1) $\max\{r_p^+, r_p^-\} = \operatorname{ord} \kappa_{p^\infty} + 1$. *Indeed, denoting* $\nu = \operatorname{ord} \kappa_{p^\infty}$, *then we have that* $r_p^{\epsilon_\nu} = \nu + 1$, *and* $0 \leqslant \nu - r_p^{-\epsilon_\nu} \equiv 0 \mod 2$. *Here* ϵ_ν *is as in (4.8).*
(2) *The group* $\mathrm{Sel}_{p^\infty}^{\epsilon_\nu}(E/K)$ *is contained in the subgroup of* $H^1(K, E[p^\infty])$ *generated by all* $c_M(n), n \in \Lambda, M \leqslant M(n)$.
(3) *As abstract abelian groups, we write* $\mathrm{Sel}_{p^\infty}^{\pm}(E/K) \simeq (\mathbb{Q}_p/\mathbb{Z}_p)^{r_p^{\pm}} \oplus \widetilde{Ш}(E/K)_{p^\infty}^{\pm}$, *where* $\widetilde{Ш}(E/K)_{p^\infty}^{\pm}$ *is a finite group. Then we have*

$$\widetilde{Ш}(E/K)^{\pm}[p^\infty] \simeq \bigoplus_{i \geqslant 1} (\mathbb{Z}/p^{a_i^{\pm}}\mathbb{Z})^2, \quad a_1^{\pm} \geqslant a_2^{\pm} \geqslant \ldots,$$

where

$$\begin{cases} a_i^{\epsilon_\nu} = \mathscr{M}_{\nu+2i-1} - \mathscr{M}_{\nu+2i}, & i \geqslant 1, \\ a_{i+(\nu - r_p^{-\epsilon_\nu})}^{-\epsilon_\nu} = \mathscr{M}_{\nu+2i-2} - \mathscr{M}_{\nu+2i-1}, & i \geqslant 1. \end{cases}$$

In particular, we have

$$\#\widetilde{Ш}(E/K)[p^\infty] \geqslant p^{2(\mathscr{M}_\nu - \mathscr{M}_\infty)},$$

and the equality holds when $\nu = 0$.

Remark 23. When $r_{an}(E/K) = 1$, the Gross–Zagier formula (3.3) implies that the Heegner point $y_K \in E(K)$ is non torsion. It then follows that $\kappa_{p^\infty} \neq 0$, and indeed $\operatorname{ord} \kappa_{p^\infty} = 0$ by (4.11). Then the theorem above implies Theorem 1.4.

Remark 24. The group $\widetilde{Ш}(E/K)[p^\infty]$ is the hypothetically finite group $Ш(E/K)[p^\infty]$. Indeed, if $Ш(E/K)[p^\infty]$ is finite, then $\widetilde{Ш}(E/K)[p^\infty] \simeq Ш(E/K)[p^\infty]$.

Therefore, under the hypothesis of Theorem 4.6, all (1)–(3) in Kolyvagin's theorem 4.7 hold. This then implies Theorem 1.8 by a suitable choice of auxiliary K. Moreover, one may then show that for a broad class of $E/\mathbb{Q}$, $\mathrm{Sel}_{p^\infty}(E/\mathbb{Q})$ (of arbitrary rank) can be constructed from Heegner points defined over ring class fields of a suitable K.

Acknowledgement

The author thanks B. Gross for pointing out the paper of Bloch [6] on a Tamagawa formulation of the Birch and Swinnerton-Dyer conjecture, B. Poonen for pointing out some inaccuracy during the CDM talks. He also thanks Rahul Krishna for many comments on the exposition of the article. The author was supported in part by NSF Grant DMS-1301848 and a Sloan research fellowship.

References

[1] Aizenbud, Gourevitch, Rallis, Schiffmann, *Multiplicity one theorems*, Ann. of Math. (2) 172 (2010), no. 2, 1407–1434.
[2] R. Beuzart-Plessis, *La conjecture locale de Gross-Prasad pour les reprsentations tempres des groupes unitaires*, arXiv:1205.2987
[3] Birch, B. *Heegner points: the beginnings.* Heegner points and Rankin L-series, 1-10, Math. Sci. Res. Inst. Publ., 49, Cambridge Univ. Press, Cambridge, 2004
[4] Bertolini, M.; Darmon, H. *Iwasawa's main conjecture for elliptic curves over anticyclotomic $\mathbb{Z}_p$-extensions.* Ann. of Math. (2) 162 (2005), no. 1, 1-64.
[5] R. Birch, B. J.; Swinnerton-Dyer, H. P. F. *Notes on elliptic curves. II.* J. Reine Angew. Math. 218 1965 79-108.
[6] S. Bloch. *A note on height pairings, Tamagawa numbers, and the Birch and Swinnerton-Dyer conjecture.* Invent. Math. 58 (1980), no. 1, 65–76.
[7] D. Bump, S. Friedberg, and J. Hoffstein, *Nonvanishing theorems for L-functions of modular forms and their derivatives*, Invent. Math. 102 (1990), no. 3, 543–618.
[8] Coates, J., Wiles, A. *On the conjecture of Birch and Swinnerton-Dyer.* Invent. Math. 39 (1977), no. 3, 223-251.
[9] Wee Teck Gan, Benedict H. Gross and Dipendra Prasad. *Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups.* Asterisque 346, 1-110.
[10] Wee Teck Gan, Atsushi Ichino. *On endoscopy and the refined Gross-Prasad conjecture for* $(\mathrm{SO}_5, , \mathrm{SO}_4)$. J. Inst. Math. Jussieu 10 (2011), no. 2, 235-324.
[11] P. Garrett, *Decomposition of Eisenstein series: Rankin triple products.* Ann. of Math. (2) 125 (1987), no. 2, 209–235.
[12] Ginzburg, David; Jiang, Dihua; Rallis, Stephen *On the nonvanishing of the central value of the Rankin-Selberg L-functions.* J. Amer. Math. Soc. 17 (2004), no. 3, 679–722
[13] Ginzburg, David; Jiang, Dihua; Rallis, Stephen. *Models for certain residual representations of unitary groups.* Automorphic forms and L-functions I. Global aspects, 125–146, Contemp. Math., 488, Amer. Math. Soc., Providence, RI, 2009.

[14] Goldfeld, D. *Sur les produits partiels eulriens attachs aux courbes elliptiques.* C. R. Acad. Sci. Paris Sr. I Math. 294 (1982), no. 14, 471-474.
[15] Gross, B. H. *Heights and the special values of L-series.* Number theory (Montreal, Que., 1985), 115187, CMS Conf. Proc., 7, Amer. Math. Soc., Providence, RI, 1987.
[16] Gross, B.H. *Kolyvagin's work on modular elliptic curves.* L-functions and arithmetic (Durham, 1989), 235-256, London Math. Soc. Lecture Note Ser., 153, Cambridge Univ. Press, Cambridge, 1991.
[17] Gross, B.H. *On the motive of a reductive group.* Invent. Math. 130 (1997), no. 2, 287–313.
[18] Gross, B.H. *The arithmetic of elliptic curves—an update.* Arabian Journal of Science and Engineering 1 (2009), 95–103.
[19] B.H. Gross, D. Prasad, *On the decomposition of a representation of* SO_n *when restricted to* SO_{n-1}. Canad. J. Math. 44 (1992), no. 5, 974–1002.
[20] B.H. Gross, D. Prasad, *On irreducible representations of* $\mathrm{SO}_{2n+1} \times \mathrm{SO}_{2m}$. Canad. J. Math. 46 (1994), no. 5, 930–950.
[21] B. Gross; D. Zagier: *Heegner points and derivatives of L-series.* Invent. Math. 84 (1986), no. 2, 225–320.
[22] M. Harris and S. Kudla, *On a conjecture of Jacquet.* Contributions to automorphic forms, geometry, and number theory, 355–371, Johns Hopkins Univ. Press, Baltimore, MD, 2004.
[23] R. Neal Harris, *A refined Gross–Prasad conjecture for unitary groups*, arXiv:1201.0518.
[24] Heegner, K. *Diophantische Analysis und Modulfunktionen.* Math. Z. 56, (1952). 227-253.
[25] Ichino Atsushi. *Trilinear forms and the central values of triple product L-functions.* Duke Math. J. Volume 145, Number 2 (2008), 281–307.
[26] A. Ichino and T. Ikeda. *On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture,* Geom. Funct. Anal. 19 (2010), no. 5, 1378–1425.
[27] Jacquet, H.; Piatetski-Shapiro, I. I.; Shalika, J. A. *Rankin- Selberg convolutions.* Amer. J. Math. 105 (1983), no. 2, 367–464.
[28] H. Jacquet, S. Rallis. *On the Gross-Prasad conjecture for unitary groups.* in *On certain L-functions*, 205264, Clay Math. Proc., 13, Amer. Math. Soc., Providence, RI, 2011.
[29] K. Kato, *p-adic Hodge theory and values of zeta functions of modular forms.* Cohomologies p-adiques et applications arithmtiques. III. Astrisque No. 295 (2004), ix, 117-290.
[30] Kobayashi, S. *The p-adic Gross-Zagier formula for elliptic curves at supersingular primes.* Invent. Math. 191 (2013), no. 3, 527–629.
[31] Kolyvagin, V. A. *On the structure of Shafarevich-Tate groups.* Algebraic geometry (Chicago, IL, 1989), 94–121, Lecture Notes in Math., 1479, Springer, Berlin, 1991.
[32] Kolyvagin, V. A. *On the structure of Selmer groups.* Math. Ann. 291 (1991), no. 2, 253–259.
[33] Kudla, S. *Central derivatives of Eisenstein series and height pairings.* Ann. of Math. (2) 146 (1997), no. 3, 545–646.
[34] Y. Liu. *Relative trace formulae toward Bessel and Fourier-Jacobi periods of unitary groups.* Manuscripta Mathematica, 145 (2014) 1–69.
[35] Y. Liu. *Refined Gan-Gross-Prasad conjecture for Bessel periods.* Journal fur die reine und angewandte Mathematik, to appear.
[36] Mazur, B. *Rational isogenies of prime degree* (with an appendix by D. Goldfeld). Invent. Math. 44 (1978), no. 2, 129162.
[37] Mazur, B.; Tate, J. *Refined conjectures of the "Birch and Swinnerton-Dyer type".* Duke Math. J. 54 (1987), no. 2, 711-750.
[38] M.R. Murty and V.K. Murty, *Mean values of derivatives of modular L-series.* Ann. of Math. (2) 133 (1991), no. 3, 447-475.
[39] Perrin-Riou, B. *Points de Heegner et dérivées de fonctions L p-adiques.* Invent. Math. 89 (1987), no. 3, 455-510.
[40] M. Rapoport, U.Terstiege, W. Zhang. *On the arithmetic fundamental lemma in the minuscule case.* Compositio Math., 2013, vol. 149, issue 10, pp. 1631–1666.
[41] Shafarevich, I. R. *The group of principal homogeneous algebraic manifolds* (in Russian), Doklady Akademii Nauk SSSR (1959) 124: 42-43, ISSN 0002-3264.
[42] Skinner, C. *A converse to a theorem of Gross, Zagier, and Kolyvagin*, preprint 2013.
[43] Skinner, C. *Main conjectures and modular forms.* Current developments in mathematics, 2004, 141–161, Int. Press, Somerville, MA, 2006.
[44] Skinner, C., Urban, E. *The Iwawasa main conjectures for* GL_2, Invent. Math., 195 (2014) no.1, pp 1–277

[45] Sun B., Zhu C. *Multiplicity one theorems: the Archimedean case*, Ann. of Math. (2) 175 (2012), no. 1, 23–44.
[46] Tate, J, *Duality theorems in Galois cohomology over number fields*, Proceedings of the International Congress of Mathematicians (Stockholm, 1962), Djursholm: Inst. Mittag-Leffler, pp. 288-295.
[47] Tate, J, *The arithmetic of elliptic curves.* Invent. Math. 23 (1974), 179-206.
[48] Ye Tian, *Congruent numbers and Heegner points*, Cambridge Journal of Mathematics, 2 (2014) no.1, 117–161.
[49] Y. Tian, *Congruent Numbers with many prime factoirs*, PNAS, Vol 109, no. 52. 21256-21258.
[50] Y. Tian, Xinyi Yuan, and Shou-Wu Zhang, *Genus periods, Genus points and Congruent number problem.* preprint.
[51] I. Vollaard, T. Wedhorn, *The supersingular locus of the Shimura variety for $GU(1, n-1)$, II.*, Invent. math. **184** (2011), 591–627.
[52] J. Waldspurger: *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie.* Compositio Math. 54 (1985), no. 2, 173–242.
[53] J. Waldspurger: *Une formule intgrale relie la conjecture locale de Gross-Prasad, 2me partie: extension aux reprsentations tempres .* arXiv:0904.0314.
[54] X. Wan, *Iwasawa Main Conjecture for Hilbert Modular Forms*, preprint, 2013.
[55] T. Watson: *Rankin triple products and quantum chaos*, to appear in Ann. of Math. (2), Ph.D. dissertation, Princeton University, Princeton, 2002.
[56] Xinyi Yuan, Shou-Wu Zhang, Wei Zhang. *The Gross–Zagier formula on Shimura curves.* Ann. of Math. Studies #184, Princeton University Press, 2012.
[57] Zhiwei Yun: *The fundamental lemma of Jacquet and Rallis.* With an appendix by Julia Gordon. Duke Math. J. 156 (2011), no. 2, 167–227.
[58] S. Zhang. *Heights of Heegner points on Shimura curves.* Ann. of Math. (2) 153 (2001), no. 1, 27–147
[59] S. Zhang. *Gross-Zagier formula for GL(2).* Asian J. Math. 5 (2001), no. 2, 183–290.
[60] S. Zhang. *Elliptic curves, L-functions, and CM-points.*Current developments in mathematics, 2001, 179–219, Int. Press, Somerville, MA, 2002.
[61] S. Zhang. *Gross-Zagier formula for GL(2). II.* Heegner points and Rankin L-series, 191–214, Math. Sci. Res. Inst. Publ., 49, Cambridge Univ. Press, Cambridge, 2004.
[62] S. Zhang. *Linear forms, algebraic cycles, and derivatives of L-series*, preprint.
[63] W. Zhang. *On arithmetic fundamental lemmas*, Invent. Math., 188, No. 1 (2012), 197–252.
[64] W. Zhang. *Fourier transform and the global Gan–Gross–Prasad conjecture for unitary groups*, Ann. of Math., to appear.
[65] W. Zhang: *Automorphic period and the central value of Rankin–Selberg L-function.* J. Amer. Math. Soc. (2014), 541-612.
[66] W. Zhang: *Harmonic analysis for relative trace formula.* in *Automorphic Representations and L-functions*, edited by: D. Prasad, C. S. Rajan, A. Sankaranarayanan, and J. Sengupta, Tata Institute of Fundamental Research, Mumbai, India
[67] W. Zhang: *Selmer group and the indivisibility of Heegner points.* preprint 2013.

Department of Mathematics, Columbia University, MC 4423, 2990 Broadway, New York, NY 10027

Current address: Department of Mathematics, Columbia University, MC 4423, 2990 Broadway, New York, NY 10027

E-mail address: wzhang@math.columbia.edu