Singular and Supersingular Moduli

a senior honors thesis by
Anthony Varilly

Advised by
Prof. Benedict H. Gross

in partial fulfillment of the honors requirements
for the degree of
Bachelor of Arts in Mathematics

Harvard University
Cambridge, Massachusetts
March 31, 2003

Acknowledgements

It is a pleasure to thank a small group of people without whom this paper would not have been possible.

First, I would like to thank my advisor, Benedict Gross, for his continuous support, his dedication and patience, but most importantly for the inspiration he provided during the past two years. Words cannot describe how wonderful the experience has been.

I would also like to thank Professors Frank Calegari, Christophe Cornut and Mark Villarino for countless helpful conversations, their generosity with their time and accessibility. In this regard, I would also like to express my gratitude towards Peter Green, a friend I look up to with utmost admiration.

A special thanks also goes to my friends Nitin Saksena and Wei Ho, with whom I have grown personally and academically throughout my college years. Their role in the completion of this paper cannot be overestimated. My roommates Brian Boyle and Matt Victory also deserve credit here; we have enjoyed many laughs together.

I would like to thank my family. I owe my passion for Mathematics to my father, whom I greatly admire and love; my brother Patrick, as well as Mima and Paola have been great sources of support and stimulation throughout my life. Together with my mother, I owe them the person that I am today.

Quizá el agradecimiento más especial es para mi madre, Susy, quien falleció durante las etapas iniciales de esta tesis, después de una larga lucha contra el cáncer. Su alegría y espíritu de lucha serán siempre infinitas fuentes de inspiración para mí.

To my mother,
in loving memory.

Contents

Chapter 1

Introduction

Elliptic Curves are special geometric objects that come equipped with a rich structure: one may "add" points of an elliptic curve much in the same way that one adds two numbers.

The structure preserving *maps* between elliptic curves are just as important as the curves themselves. These maps are called isogenies. The set of isogenies from an elliptic curve to *itself* is called the endomorphism ring of the curve. A 'generic' elliptic curve has an endomorphism ring that looks like a copy of the integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Some curves, however, possess 'extra' maps. They are said to have *complex multiplication.* Among all curves with complex multiplication, some have so many 'extra' maps that they are given a name of their own: they are known as *supersingular* curves: their endomorphism ring has the structure of a maximal order in a rational quaternion algebra.

Every elliptic curve has a number attached to it, called a j-invariant, which classifies the curve up to isomorphism. The j-invariant of a curve that has complex multiplication (resp. is supersingular) is called a *singular modulus* (resp. *supersingular modulus*). Singular moduli of elliptic curves defined over the complex numbers are associated to imaginary quadratic fields (these are obtained by adjoining square roots of negative numbers to the rational numbers).

This paper studies the absolute norms of differences of singular moduli corresponding to elliptic curves with complex multiplication by a ring of integers in an imaginary quadratic field. We have affectionately called these norms *Gross–Zagier* numbers, after the two mathematicians whose groundbreaking work has led us to an understanding of many properties of them (cf. [G–Z]).

Gross–Zagier numbers can be enormous, yet they appear to have extremely small prime factors. For example, one such number is

$$\begin{aligned} &-19077542993352945680961028994697271308288000000000000 \\ =&-(2^8 \cdot 3^4 \cdot 5^4 \cdot 11^2 \cdot 23^2 \cdot 29^2 \cdot 383)^3 \end{aligned}$$

Intuitively, the number 383 seems too paltry to be the largest prime factor dividing a 53-digit monster. This phenomenon is ubiquitous among singular moduli. The goal of this thesis is to answer a natural question: why?

It is known that, on average, the number of divisors of N, denoted $d(N)$, is of order $\log N$ (cf. [H–W, Theorem 319]). To give the reader some sense of how unusually divisible Gross–Zagier numbers are, we have compiled a table with a few more examples (cf. Table 1.1). In all cases, $d(N)$ is rather large compared to $\log N$. Moreover, a number with many divisors cannot have very big

An imaginary quadratic field K	The Gross–Zagier number $N(j_K - 0)$	$\log(N)$	$d(N)$
$\mathbb{Q}(\sqrt{-1})$	$+(2^2 \cdot 3)^3$	7.45	28
$\mathbb{Q}(\sqrt{-7})$	$-(3 \cdot 5)^3$	8.12	16
$\mathbb{Q}(\sqrt{-5})$	$-(2^4 \cdot 5 \cdot 11)^3$	20.34	208
$\mathbb{Q}(\sqrt{-6})$	$-(2^4 \cdot 3^2 \cdot 17)^3$	23.41	364
$\mathbb{Q}(\sqrt{-23})$	$-(5^3 \cdot 11 \cdot 17)^3$	30.18	160
$\mathbb{Q}(\sqrt{-13})$	$-(2^4 \cdot 3^2 \cdot 5^2 \cdot 23)^3$	33.97	2548
$\mathbb{Q}(\sqrt{-163})$	$-(2^6 \cdot 3 \cdot 5 \cdot 23 \cdot 29)^3$	40.11	4864
$\mathbb{Q}(\sqrt{-21})$	$-(2^8 \cdot 3^5 \cdot 47 \cdot 59)^3$	56.90	6400
$\mathbb{Q}(\sqrt{-133})$	$-(2^8 \cdot 3^4 \cdot 5^4 \cdot 11^2 \cdot 23^2 \cdot 29^2 \cdot 383)^3$	120.38	5796700

Table 1.1: Some Gross–Zagier numbers

prime factors, so the table also supports the empirical claim that Gross–Zagier numbers have small prime divisors.

The reader may wonder why this problem is worth pursuing. Many number theorists are likely to find this problem beautiful in and of itself and hence a worthwhile question. Perhaps a more satisfying answer, however, is the following. A good mathematical problem is one that raises ten questions before it can be answered, and (hopefully) ten more questions *after* it is answered. Moreover, a good problem should either require the development of new mathematical tools for its solution, or it should bring together and connect seemingly unrelated areas of knowledge. Our problem is of the latter kind.

The issue of primes dividing norms of singular moduli raises many questions. For example, before we ask why the prime factors of Gross–Zagier numbers are small, we should wonder why Gross–Zagier numbers are integers at all! There is no *a priori* reason to believe this is the case.

The full answer to our problem lies at the crossroads of many areas of mathematics. For example, we will use analytical methods involving modular forms of weight zero to show singular moduli are algebraic integers and in this way explain the integrality of Gross–Zagier numbers. We will also carefully study the possible endomorphism rings of elliptic curves, with particular attention to supersingular curves. This study will lead us into the realm of rational quaternion algebras (of which Hamilton's quaternions are an example) and a complete classification of them, as well as the theory of formal groups, which we will use to establish a maximality property of the endomorphism ring of a supersingular curve. Although the elliptic curves over the complex numbers that give rise to Gross–Zagier numbers cannot be supersingular, they may *reduce* modulo a prime to a curve (defined over a field of positive characteristic) which *is* supersingular. The technique of reduction will play a fundamental role in the bound given in this paper for the primes dividing Gross–Zagier numbers. Our journey will also take us through a specialized study of the theory of complex multiplication, which, as an added bonus, will give a method for (educatedly) guessing the Hilbert class field for a few imaginary quadratic fields of small class number.

Bounding the size of prime factors that divide Gross–Zagier numbers will only raise many more questions about these numbers. For example, given a singular modulus, how do we know *which* primes divide it's associated Gross–Zagier number? With what *exponent* does a prime divide this number? We will not address these questions, but the interested reader may consult [G–Z].

We now give a brief outline of the paper. In Chapter 2 we review the general theory of elliptic curves, with particular attention to the structure of their endomorphism rings. Chapter 3 contains a discussion of elliptic curves with complex multiplication by the ring of integers of an imaginary quadratic field as well as a proof of the integrality of singular moduli and an introduction to Gross–Zagier numbers. Chapter 4 is intended to lay the foundations for a study of supersingular curves; in it we consider rational quaternion algebras in the abstract (a theory which is quite beautiful) and give a complete classification of them. Chapter 5 then proceeds with the study of supersingular curves. The main result we prove is a *maximality* property of the endomorphism ring of such a curve. Chapter 6 is a brief introduction to the theory of reduction elliptic curves defined over number fields, with particular attention to curves that have supersingular reduction. In this chapter we come back to Gross–Zagier numbers and give a bound for the primes that divide these numbers. Specifically, we will show that if K and K' are imaginary quadratic fields with relatively prime discriminants, then a prime p that divides $N(j_K - j_{K'})$ is at most equal to $DD'/4$, where D and D' are the discriminants of K and K', respectively.

To the best of the author's knowledge, the material we present is somewhat scattered in the literature and has heretofore not been accessible in a single source to the non-specialist. We have reconstructed several proofs of “well-known” theorems in what we hope is a refreshing and welcome presentation.

Chapter 2

Basic Theory of Elliptic Curves

In this chapter we recall some of the basic facts about elliptic curves. We only provide proofs for a few theorems of latter importance to us. The reader is referred to the excellent books by Silverman [Sil 1] (our main source for this section) and Husemöller [Hus] for more detailed and complete treatments; a concise yet informative exposition can be found in [Shi, Ch. 4] or in [Ked, Ch. 5].

Let K be a field. Our primary objects of study are special algebraic curves in $\mathbb{P}^2(\overline{K})$. These are projective algebraic sets defined over K by a principal ideal $I \in K[X, Y, Z]$, or equivalently, they are the sets of K-solutions to a single homogeneous polynomial equation

$$f(X, Y, Z) = 0.$$

Oftentimes we will focus on the restriction of an algebraic curve to the complement of the affine hyperplane $Z = 0$. We set $x = X/Z, y = Y/Z$ and $Z = 1$ to obtain an affine algebraic set given by the zero set of the polynomial $g(x, y) = f(X/Z, Y/Z, 1)$. This restriction is not too severe since we will concentrate on elliptic curves, which only have one point 'at infinity', namely $[0, 1, 0]$. To reverse this process and recover a projective curve from an affine algebraic set defined by a single equation, we insert as many Z's as 'necessary.' For example, when we refer to the *projective* curve

$$y^2 = x^3 + x$$

we really mean the variety in $\mathbb{P}^2$ given by the homogeneous equation

$$Y^2 Z = X^3 + XZ^2.$$

2.1 Elliptic Curves

Assume for simplicity that K is a field whose characteristic is not 2 or 3[1]. Then an elliptic curve E defined over K (denoted E/K) is a nonsingular curve in the projective plane $\mathbb{P}^2$ of the form

$$y^2 = 4x^3 - g_2 x - g_3, \tag{2.1}$$

[1] Much of the theory of elliptic curves that we present in this chapter for fields of positive characteristic $\neq 2, 3$ also applies to fields of arbitrary characteristic. For the more general theory, however, one needs a more general equation than (2.1) to describe an elliptic curve. Silverman and Hüsemoller provide full treatments of elliptic curves over fields of characteristic 2 and 3, cf. [Hus] and [Sil 1, Appendix A], especially.

with $g_2, g_3 \in K$, whose only point on the line at infinity is $O = [0,1,0]$. Every elliptic curve admits the structure of an abelian group; the coordinates of sum of two points are given by regular functions. The point O plays the identity role in the group.

We say two elliptic curves E and E' are *isomorphic* if there exists $u \in K^*$ such that $g_2' = u^4 g_2$ and $g_3' = u^6 g_3$. We define the discriminant $\Delta(E)$ of an elliptic curve E as $\Delta(E) = g_2^3 - 27g_3^2$. The nonsingularity of E is equivalent to $\Delta(E) \neq 0$. The *j-invariant* of E is defined as the quantity

$$j(E) = \frac{1728 g_2^3}{\Delta(E)}.$$

The j-invariant of E classifies the elliptic curve up to isomorphism over $\overline{K}$. For a given $j_0 \in K$ there exists an elliptic curve defined over $K(j_0)$ with j-invariant equal to j_0. For example,

$$\begin{aligned} \text{for } j_0 = 0, &\quad \text{take} \quad y^2 = 4x^3 - 1 \\ \text{for } j_0 = 1728, &\quad \text{take} \quad y^2 = 4x^3 - 12x \\ \text{for } j_0 \neq 0, 1728, &\quad \text{take} \quad y^2 = 4x^3 - \frac{27j_0}{j_0 - 1728}x - \frac{27j_0}{j_0 - 1728}. \end{aligned}$$

2.2 Isogenies

An *isogeny* is a regular map between two elliptic curves $\phi : E_1/K \to E_2/K$ such that $\phi(O) = O$. We say that ϕ is *defined over* K if it commutes with the action of the Galois group $\mathrm{Gal}(\overline{K}/K)$, i.e., if

$$\phi(\sigma(P)) = \sigma(\phi(P)) \quad \text{for all } \sigma \in \mathrm{Gal}(\overline{K}/K).$$

For clarity, we will usually denote the action of a Galois group with a superscript, and write the above equality, for example, as $\phi(P^\sigma) = \phi^\sigma(P)$.

Every isogeny is in fact a homomorphism that respects the group law of the curves. A nontrivial isogeny is surjective (a fact derived from the general theory of maps between smooth curves). The set $\mathrm{Hom}(E_1, E_2)$ of isogenies between two elliptic curves has a torsion-free $\mathbb{Z}$-module structure, with rank at most equal to 4 (we will prove this later).

Remark 2.1. Whenever we refer to the sets of isogenies $\mathrm{Hom}(E_1, E_2)$ or $\mathrm{End}\, E$ without further qualification, it will be understood we refer to isogenies defined over $\overline{K}$.

Example 2.1. One of the most important examples of isogenies are the 'multiplication by m' maps $[m] : E \to E$ given by $P \mapsto P + \cdots + P$ (m times) where $+$ is used to denote the group law on the points of E. The kernel of these maps are precisely the subgroups of m-torsion points on the curve and are denoted $E[m]$, respectively. We will see that $E[m] \cong (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$ as abstract groups in characteristic 0. For most elliptic curves, the 'multiplication by m' maps are the only nonconstant isogenies in $\mathrm{End}(E)$. Elliptic curves that possess extra endomorphisms are said to admit *complex multiplication.*

Example 2.2. In a field of characteristic $p > 0$, the q^{th}-power Frobenius automorphism $x \mapsto x^q$ ($q = p^r$) induces an isogeny $E \to E^q$ where E^q is the elliptic curve

$$y^2 = 4^q x^3 - g_2^q x - g_3^q.$$

An isogeny ϕ induces an injection of elliptic function fields (i.e., function fields of elliptic curves) $\phi^* : K(E_2) \to K(E_1)$ given by pulling functions back, i.e., if $f \in K(E_2)$ then $\phi^* f = f \circ \phi$. We define the *degree* of ϕ by

$$\deg \phi = [K(E_1) : \phi^* K(E_2)].$$

The separable and inseparable degrees of ϕ (denoted $\deg_s \phi$ and $\deg_i \phi$, respectively) are defined similarly. We also say, for example, that ϕ is a separable map if $K(E_1)$ is a separable extension of $\phi^* K(E_2)$.

Theorem 2.2. *Every isogeny $\phi : E_1 \to E_2$ over a field of positive characteristic factors as*

$$E_1 \xrightarrow{\psi} E_1^q \xrightarrow{\lambda} E_2.$$

where $q = \deg_i \phi$, ψ is the q^{th}-power Frobenius automorphism and λ is a separable map. □

This theorem follows from the proposition that every extension can be written as a purely inseparable extension of a separable extension, obtained by adjoining a certain q^{th} root to the separable extension, where q is the inseparable degree of the original extension.

The basic Galois theory of elliptic function fields can be found in Silverman's book, cf. [Sil 1, § III.4]. We summarize Silverman's treatment in the following useful theorems:

Theorem 2.3 (Galois theory of Elliptic Function Fields). *Let ϕ be a nonconstant isogeny between two elliptic curves E_1/K and E_2/K. Then for any $Q \in E_2$*

$$\#\phi^{-1}(Q) = \deg_s \phi.$$

In particular, if ϕ is separable

$$\# \ker \phi = \deg \phi,$$

moreover, $\overline{K}(E_1)$ is a Galois extension of $\phi^* \overline{K}(E_2)$. □

Theorem 2.4. *Let E be an elliptic curve and let G be a finite group of point of E. There exists a unique elliptic curve E' and a separable isogeny $\phi : E \to E'$ with $\ker \phi = G$.* □

2.3 Invariant Differentials

The space of meromorphic differential forms on an elliptic curve E, denoted Ω_E is the vector space over the field $\overline{K}(E)$ generated by symbols df (where $f \in \overline{K}(E)$) which are subject to the usual formal rules

- $d(f+g) = df + dg$ for all $f, g \in \overline{K}(E)$.
- $d(fg) = f\,dg + g\,df$ for all $f, g \in \overline{K}(E)$.
- $dc = 0$ for all $c \in \overline{K}$.

For a non-trivial isogeny $\phi : E_1 \to E_2$ the map $\phi^* : \overline{K}(E_2) \to \overline{K}(E_1)$ induces a map on differentials, also denoted ϕ^*

$$\phi^* : \Omega_{E_2} \to \Omega_{E_1}$$
$$\phi^* \left(\sum f_i dx_i \right) = \sum (\phi^* f_i) d(\phi^* x_i).$$

When working in characteristic zero, where all field extensions (and hence all isogenies) are separable, the above map is actually injective. We record this criterion for future reference in the following theorem (cf. [Sil 1, Theorem II.4.2(c)]).

Theorem 2.5. *Let $\phi : E_1 \to E_2$ be a non-trivial isogeny. Then ϕ is separable if and only if the map*

$$\phi^* : \Omega_{E_2} \to \Omega_{E_1}$$

is injective. □

We will be interested in those differentials on an elliptic curve defined by functions $f \in K(E)$ without poles, which are invariant under translation by a point on the curve. An example of such differentials is given by $\omega = dx/2y = dy/(12x^2 - g_2)$. These objects are extremely useful because they linearize the complicated addition law on the curve, as the following theorem illustrates (cf [Sil 1, Theorem III.5.2]).

Theorem 2.6. *Let E_1 and E_2 be two elliptic curves, let $\omega \in \Omega_{E_1}$ be an invariant differential and let $\phi, \psi : E_2 \to E_1$ be two isogenies. Then*

$$(\phi + \psi)^*\omega = \phi^*\omega + \psi^*\omega.$$ □

An easy corollary of this theorem is that the pullback of an invariant differential through the 'multiplication by m' isogeny is equal to scalar multiplication by m, i.e., $[m]^*\omega = m\omega$. This means $[m] \neq [0]$, and from Theorem 2.5 it follows the multiplication by m maps are separable.

2.4 Elliptic Curves over $\mathbb{C}$

We make a pause here in our study of elliptic curves over arbitrary fields and specialize to the case $K = \mathbb{C}$. The rich structure of the complex numbers allows one to use lattices to study elliptic curves. The usefulness of such a tool will become apparent in our discussion of the theory of complex multiplication over $\mathbb{C}$. The reader is referred to [Cox, Sil 1, Hus] for a full treatment of the subject.

As a complex analytic manifold, every elliptic curve defined over a subfield of $\mathbb{C}$ is isomorphic to a one-dimensional complex torus $\mathbb{C}/\Lambda$, where Λ is a lattice of $\mathbb{C}$ (i.e., a discrete submodule of rank 2 over $\mathbb{Z}$). To describe this isomorphism explicitly, we introduce the Weierstrass $\wp$-function

$$\wp(z; \Lambda) = \wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda - 0} \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right]$$

The $\wp$-function converges uniformly on compact subsets of $\mathbb{C} - \Lambda$. It is an even function whose only poles are the points of Λ. Furthermore, it is an elliptic function with respect to the lattice Λ, i.e., it is a meromorphic function such that $\wp(z + \omega) = \wp(z)$ for all $\omega \in \Lambda$. The field of elliptic functions of a lattice Λ is generated by the $\wp$-function and its derivative

$$\wp'(z) = -2 \sum_{\omega \in \Lambda - 0} \frac{1}{(z-\omega)^3}.$$

The functions $\wp(z)$ and $\wp'(z)$ have Laurent series expansions around $z = 0$ given by

$$\wp(z) = \frac{1}{z^2} + \sum_{2\leq k} G_{2k}(\Lambda)(2k-1)z^{2k-2},$$
$$\wp'(z) = \frac{-2}{z^3} + \sum_{2\leq k} G_{2k}(\Lambda)(2k-1)(2k-2)z^{2k-3},$$

where $G_{2k}(\Lambda) = \sum_{\omega\in\Lambda-0}\omega^{-2k}$ is the Eisenstein series of weight $2k$ (cf. [Hus, §9.4]). Since $\wp(z)$ and $\wp'(z)$ are elliptic functions, the difference $\wp'(z)^2 - 4\wp(z)^3 + 60G_4\wp(z) + 140G_4$ is also elliptic. However, using the above expansions we note this difference has the form $z \cdot$ (holomorphic function). Now note that a holomorphic elliptic function f is constant because it factors by continuous functions $\mathbb{C} \to \mathbb{C}/\Lambda \to \mathbb{C}$, and since $\mathbb{C}/\Lambda$ is compact, f is bounded, so Liouville's theorem tells us f is constant. Since the above difference vanishes at zero, we have shown the following theorem.

Theorem 2.7. *The Weierstrass $\wp$-function satisfies the differential equation*

$$\wp'(z;\Lambda)^2 = 4\wp(z;\Lambda)^3 - g_2(L)\wp(z;\Lambda) - g_3(\Lambda),$$

where $g_2(\Lambda) = 60G_4(\Lambda)$ and $g_3(\Lambda) = 140G_6(\Lambda)$. □

It follows that the point $(\wp(z), \wp'(z))$ is on the elliptic curve $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$. In fact

Theorem 2.8. *The map*

$$\begin{aligned} \mathbb{C}/\Lambda &\longrightarrow E_\Lambda/\mathbb{C} : y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda) \\ z &\longmapsto (\wp(z,\Lambda), \wp'(z,\Lambda)), \end{aligned}$$

is an isomorphism of complex analytic manifolds. [cf. [Sil 2, Cor. 4.3 § I.4]] □

Since we can associate an elliptic curve to a lattice by the above isomorphism, it is natural to ask whether the converse is true: given an elliptic curve $E/\mathbb{C}$, does there exist a lattice $\Lambda \subset \mathbb{C}$ such that the above map is an isomorphism of complex analytic manifolds? The answer is yes; this is known as the *Uniformization Theorem* (cf. [Sil 2, § I.4]). Furthermore, this lattice is unique up to homothety; recall two lattices Λ_1 and Λ_2 are *homothetic* if there is an $\alpha \in \mathbb{C}^*$ such that $\alpha\Lambda_1 = \Lambda_2$.

In this way an isogeny $\phi : E_1 \to E_2$ of elliptic curves over $\mathbb{C}$ gives rise to a holomorphic map $\phi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ such that $\phi(0) = 0$, where Λ_1 and Λ_2 are the lattices that correspond to E_1 and E_2, respectively. The converse is also true: the natural inclusion

$$\{\text{isogenies } \phi : E_1 \to E_2\} \longrightarrow \{\text{holomorphic maps } \phi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2 \text{ such that } \phi(0) = 0\}$$

is a bijection. The set of holomorphic maps above is in turn in bijection with the set

$$\{\alpha \in \mathbb{C}^* \mid \alpha\Lambda_1 \subset \Lambda_2\}.$$

For a given α in the latter set, then map $\phi_\alpha(z) = \alpha z \bmod \Lambda_2$ is a holomorphic homomorphism that preserves the origin. The map $\alpha \mapsto \phi_\alpha$ gives the desired bijection of sets. We may summarize our discussion in the following two theorems.

Theorem 2.9. *Two elliptic curves E_1 and E_2 are isomorphic over $\mathbb{C}$ if and only if their associated lattices Λ_1 and Λ_2, respectively, are homothetic.* □

Theorem 2.10. *Let $E_\Lambda/\mathbb{C}$ be the elliptic curve associated to the lattice Λ by Theorem 2.8. Then*

$$\operatorname{End}(E_\Lambda) \cong \{\alpha \in \mathbb{C} \mid \alpha\Lambda \subset \Lambda\}. \qquad \square$$

As a consequence of Theorem 2.10 we can show that an elliptic curve over $\mathbb{C}$ cannot be supersingular (cf. Chapter 5). The reader should compare Theorem 2.11 with Theorem 2.17, which gives a complete characterization of the endomorphism ring of an elliptic curve over an arbitrary field.

Theorem 2.11. *Let $E/\mathbb{C}$ be an elliptic curve and let $\Lambda = [1, \tau]$ be the lattice associated to E. Then* $\operatorname{End} E \cong \mathbb{Z}$ *or* $\operatorname{End} E$ *is isomorphic to an order in a the quadratic imaginary field $\mathbb{Q}(\tau)$.*

Proof. We know from Theorem 2.10 that $\operatorname{End} E \cong \{\alpha \in \mathbb{C} \mid \alpha\Lambda \subset \Lambda\}$. Since $[1, \tau]$ is a basis for Λ, for any $\alpha \in \operatorname{End} E$ there are integers a, b, c, d such that

$$\alpha = a + b\tau \quad \text{and} \quad \alpha\tau = c + d\tau.$$

We eliminate τ and obtain

$$\alpha^2 - (a + d)\alpha + (ad - bc) = 0,$$

from which it follows that $\operatorname{End} E$ is an integral extension of $\mathbb{Z}$. If $\operatorname{End} E$ is not isomorphic $\mathbb{Z}$ then take $\alpha \in \operatorname{End} E - \mathbb{Z}$ (i.e. $b \neq 0$). Then eliminating α above gives

$$b\tau^2 - (a - d)\tau - c = 0.$$

This means $\mathbb{Q}(\tau)$ is an imaginary quadratic field, and $\operatorname{End} E$ is an integral extension of $\mathbb{Z}$ contained in this field, i.e., it is an order of $\mathbb{Q}(\tau)$. $\square$

2.5 Dual Isogenies and Tate Modules

We now resume our study of elliptic curves over an arbitrary field K.

Let E_1, E_2 be two elliptic curves over a field K. To every isogeny $\phi : E_1 \to E_2$ of degree m we may associate a unique dual isogeny $\hat{\phi} : E_2 \to E_1$ such that $\hat{\phi} \circ \phi = [m]$. If $\phi = [0]$, we set $\hat{\phi} = 0$. The existence and uniqueness of such a homomorphism is checked with Picard groups (cf. [Sil 1, § III.6]).

Theorem 2.12 (Properties of the dual isogeny). *Let $\phi : E_1 \to E_2$ be an isogeny of degree m, and let $\psi : E_1 \to E_2$, $\lambda : E_2 \to E_3$ be two other isogenies. Then*

(i) $\phi \circ \hat{\phi} = [m]$ *on* E_2.

(ii) $\widehat{\lambda \circ \phi} = \hat{\phi} \circ \hat{\lambda}$.

(iii) $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$; *in particular* $\widehat{[n]} = [n]$ *for all* $n \in \mathbb{Z}$ *and* $\deg[n] = n^2$.

(vi) $\deg \hat{\phi} = \deg \phi$.

(v) $\hat{\hat{\phi}} = \phi$.

If $\operatorname{char} K = 0$ or is prime to the integer m, then the map $[m] \in \operatorname{End} E$ is separable, so $\# \ker[m] = \deg \phi$ (cf. Theorem 2.3). In other words $\#E[m] = m^2$. The equality is true if we replace m by a positive divisor of it; since $E[m]$ is a finite abelian group, by the structure theorem for such groups it follows that

$$E[m] = \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}.$$

If, on the other hand, $\operatorname{char} K = p$, then letting ϕ be the p-th power Frobenius isogeny, Theorem 2.3 tells us that

$$\#E[p^n] = \deg_s [p^n] = (\deg_s \hat{\phi} \circ \phi)^n;$$

since the Frobenius map is purely inseparable, it follows that $\#E[p^n] = (\deg_s \hat{\phi})^n$. If $\hat{\phi}$ is inseparable as well then $\deg_s \hat{\phi} = 1$. Otherwise $\deg_s \hat{\phi} = p$. We collect all our results in the following theorem.

Theorem 2.13. *Let E/K be an elliptic curve and let m be a nonzero integer.*

(i) If $\operatorname{char} K = 0$ or is prime to m then

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$$

(ii) Otherwise K has characteristic p and

$$\begin{aligned} &E[p^e] \cong \{O\} \textit{ for all positive integers } e, \textit{ or} \\ &E[p^e] \cong \mathbb{Z}/p^e\mathbb{Z} \textit{ for all positive integers } e. \end{aligned}$$

□

If $\operatorname{char} K = 0$ or is prime to the integer m, then the Galois group $\operatorname{Gal}(\overline{K}/K)$ acts on $E[m]$ since for each $\sigma \in \operatorname{Gal}(\overline{K}/K)$ and P such that $[m]P = O$ we have $[m]P^\sigma = ([m]P)^\sigma = O$. This action gives a representation $\operatorname{Gal}(\overline{K}/K) \to \operatorname{Aut} E[m] \cong GL_2[\mathbb{Z}/m\mathbb{Z}]$. The Tate module provides a means for fitting these representations together as m ranges through prime numbers l to obtain a useful representation over a matrix ring over a field of characteristic zero.

Definition 2.1. *Let E be an elliptic curve and $l \in \mathbb{Z}$ a prime number. The l-adic Tate module of E is defined as*

$$T_l(E) = \varprojlim_n E[l^n]$$

where the inverse limit is taken with respect to the maps

$$[l] : E[l^{n+1}] \to E[l^n].$$

The Tate module has a $\mathbb{Z}_l$-module structure because each $E[l^n]$ is a $\mathbb{Z}/l^n\mathbb{Z}$-module. In fact, from the above remarks on the structure of $E[m]$, we see that

$$\begin{aligned} &T_l(E) \cong \mathbb{Z}_l \times \mathbb{Z}_l \text{ if } l \neq \operatorname{char} K, \\ &T_l(E) \cong \{0\} \text{ or } \mathbb{Z}_p \text{ if } \operatorname{char} K = p > 0. \end{aligned}$$

Since the action of $\operatorname{Gal}(\overline{K}/K)$ commutes with the multiplication by l maps used to take the inverse limit, it follows that $\operatorname{Gal}(\overline{K}/K)$ also acts on $T_l(E)$. In this way we obtain the desired representation

$$\rho : \operatorname{Gal}(\overline{K}/K) \to \operatorname{Aut}(T_l(E)).$$

The Tate module is a useful tool to study isogenies between two elliptic curves. An isogeny $\phi : E_1 \to E_2$ gives rise to maps

$$\phi : E_1[l^n] \to E_2[l^n],$$

and thus induces a $\mathbb{Z}_l$-linear map

$$\phi_l : T_l(E_1) \to T_l(E_2).$$

In other words, we obtain a homomorphism

$$\operatorname{Hom}(E_1, E_2) \to \operatorname{Hom}(T_l(E_1), T_l(E_2)).$$

Theorem 2.14. *Let E be an elliptic curve over a field K, and let l be a prime number distinct from* char K*. Then the natural map*

$$\begin{gathered}\operatorname{Hom}(E_1, E_2) \otimes \mathbb{Z}_l \to \operatorname{Hom}(T_l(E_1), T_l(E_2)) \\ \phi \mapsto \phi_l\end{gathered}$$

is injective. See [Sil 1, III.7.4] □

This strong proposition will be of great use for us. In particular, when $E_1 = E_2$ we obtain the following important result which we had mentioned.

Corollary 2.15. *The endomorphism ring of an elliptic curve* End E *is a free $\mathbb{Z}$-module of rank at most 4.*

Proof. If there exists an isogeny $\phi \in \operatorname{End} E$ together with a nonzero integer m such that $[m] \circ \phi = [0]$ then we would have $\deg[m] \cdot \deg \phi = 0$. However, we know that $\deg[m] = m^2$ (cf. Theorem2.12(iii)). Hence $\deg \phi = 0$ and consequently $\phi = [0]$. It follows that End E has characteristic zero. Furthermore, this ring is an integral domain (this follows from the fact that $\mathbb{Z}$ is an integral domain by passing from the composition $\phi_1 \circ \phi_2$ to the product $\deg[\phi_1] \cdot \deg[\phi_2]$).

For the rank of End E, first note that

$$\operatorname{rank}_{\mathbb{Z}} \operatorname{End} E = \operatorname{rank}_{\mathbb{Z}_l} \operatorname{End} E \otimes \mathbb{Z}_l,$$

and we know by Theorem 2.14 (letting $E = E_1 = E_2$) that

$$\operatorname{rank}_{\mathbb{Z}_l} \operatorname{End} E \otimes \mathbb{Z}_l \leq \operatorname{rank}_{\mathbb{Z}_l} \operatorname{End} T_l(E).$$

Depending on the characteristic of K, $T_l(E)$ is one of $\{0\}, \mathbb{Z}_l$ or $\mathbb{Z}_l \times \mathbb{Z}_l$. Thus End $T_l(E)$ is a (possibly improper) subset of $M_2(\mathbb{Z}_l)$, the group of 2×2 matrices with entries in $\mathbb{Z}_l$. Accordingly,

$$\operatorname{rank}_{\mathbb{Z}_l} \operatorname{End} T_l(E) \leq 4.$$ □

Recall that an isogeny $\phi : E_1 \to E_2$ is defined over a field K if it commutes with the action of $\mathcal{G} = \operatorname{Gal}(\overline{K}/K)$. The group of isogenies from E_1 to E_2 defined over K is denoted $\operatorname{Hom}_{\mathcal{G}}(E_1, E_2)$. Similarly, we may consider the group of Z_l-linear maps from $T_l(E_1)$ to $T_l(E_2)$ that commute with the action of $\mathcal{G}$; we denote this group by $\operatorname{Hom}_{\mathcal{G}}(T_L(E_1), T_l(E_2))$. By Theorem 2.14, the natural map

$$\operatorname{Hom}_{\mathcal{G}}(E_1, E_2) \otimes \mathbb{Z}_l \to \operatorname{Hom}_{\mathcal{G}}(T_L(E_1), T_l(E_2)) \tag{2.2}$$

is injective. The following theorem, due to Tate (cf. [Tate]), settles the issue of surjectivity of this map for a finite field K. The theorem holds in general for abelian varieties, and its proof uses methods and concepts beyond the scope of this paper.

Theorem 2.16. *Let E_1 and E_2 be two elliptic curves over a finite field K. Then the map (2.2) is an isomorphism.* □

2.6 Characterizing $\operatorname{End} E$

We are now in a position to see what kind of ring $\operatorname{End} E$ can be. Let $\mathcal{K}$ be a finitely generated $\mathbb{Q}$-algebra. Recall a subring $\mathcal{O}$ of $\mathcal{K}$ is called an order if it is a finitely generated $\mathbb{Z}$-module such that $\mathcal{O} \otimes \mathbb{Q} = \mathcal{K}$.

Theorem 2.17. *Let E be an elliptic curve and set $\mathcal{O} = \operatorname{End} E$. Then either*

(i) $\mathcal{O} = \mathbb{Z}$, *or*

(ii) $\mathcal{O}$ *is an order in a quadratic imaginary extension of* $\mathbb{Q}$, *or*

(iii) $\mathcal{O}$ *is an order in a quaternion algebra over* $\mathbb{Q}$.

(A quaternion algebra over $\mathbb{Q}$ is 4-dimensional algebra generated by $\{1, \alpha, \beta, \alpha\beta\}$ over $\mathbb{Q}$ such that α^2, β^2 are negative rational numbers and $\alpha\beta = -\beta\alpha$. We will discuss these algebras in detail in Chapter 4.)

Proof. We closely follow [Sil 1, Theorem III.9.3]. Let $\mathcal{K} = \mathcal{O} \otimes \mathbb{Q}$. By Corollary 2.15 we know that $\mathcal{O}$ is a finitely generated $\mathbb{Z}$-module, so it suffices to show that $\mathcal{K} = \mathbb{Q}$, or $\mathcal{K}/\mathbb{Q}$ is either a quadratic imaginary extension or a quaternion algebra. Let $\phi \in \mathcal{O}$ be an isogeny and $\hat{\phi}$ be its dual isogeny. Recall that $\phi \circ \hat{\phi} = \hat{\phi} \circ \phi = [m]$, where $m = \deg \phi$. Hence we have a map $\mathcal{O} \to \mathbb{Z}$ given by $\phi \mapsto m$. We may extend this map to $\mathcal{K}$. In this way we define the reduced *norm* and *trace* of ϕ as

$$n(\phi) = \phi \circ \hat{\phi} \quad \text{and} \quad t(\phi) = \phi + \hat{\phi}.$$

Let us make two remarks about the trace map. Note that

$$t(\phi) = 1 + n(\phi) - n(\phi - 1)$$

and so $t(\phi) \in \mathbb{Q}$ in the sense that $t(\phi) = [m] \otimes 1/n$, for some $m, n \in \mathbb{Z}$. Next, the isogeny ϕ is a root of the polynomial $X^2 - t(\phi)X + n(\phi)$, so if $t(\phi) = 0$ we have

$$\phi^2 = -n(\phi),$$

and $-n(\phi)$ is a negative rational number.

If $\mathcal{K} = \mathbb{Q}$ we are done. Otherwise there is an element $\phi \in \mathcal{K} - \mathbb{Q}$. Replacing ϕ with $\phi - (1/2)t(\phi)$ if necessary, we may assume that $t(\phi) = 0$ because the trace is $\mathbb{Q}$-linear and $t(\phi) = 2\phi$ for $\phi \in \mathbb{Q}$. Hence $\phi \in \mathbb{Q}$ and $\phi^2 < 0$, from which $\mathbb{Q}(\phi)$ is an quadratic imaginary extension.

If $\mathcal{K} = \mathbb{Q}(\phi)$ we are done. Otherwise let $\psi \in \mathcal{K} - \mathbb{Q}(\psi)$. Replacing ψ by

$$\psi - \frac{1}{2}t(\psi) - \frac{1}{2}(t(\phi\psi)/\phi^2)\phi,$$

we may assume that $t(\psi) = t(\phi\psi) = 0$. Thus ψ^2 is a negative rational number. Furthermore, $\mathbb{Q}[\phi, \psi]$ is a quaternion algebra because $\phi\psi = -\psi\phi$. Indeed, the equations

$$t(\phi) = t(\psi) = t(\phi\psi) = 0$$

imply that

$$\phi = -\hat{\phi}, \quad \psi = -\hat{\psi}, \quad \phi\psi = -\widehat{\phi\psi},$$

from which the desired equality follows because $\widehat{\phi\psi} = \hat{\psi}\hat{\phi}$.

It remains to show this is it, i.e., $\mathcal{K} = \mathbb{Q}[\phi, \psi]$. Since $\mathcal{O}$ has rank at most 4 as a $\mathbb{Z}$-module, it follows that $\mathcal{K}$ is at most 4-dimensional as a $\mathbb{Q}$-vector space. So it suffices to show that the set $\{1, \phi, \psi, \phi\psi\}$ is linearly independent over $\mathbb{Q}$. Suppose

$$a + b\phi + c\psi + d\phi\psi = 0 \quad a, b, c, d \in \mathbb{Q}.$$

Taking traces of both sides we see that $2a = 0$ and so $a = 0$. Composing the remaining elements with ϕ on the left and ψ on the right we obtain

$$(b\phi^2)\psi + (c\psi^2)\phi + (d\phi^2\psi^2) = 0.$$

But this is a $\mathbb{Q}$-linear dependence relation for the set $\{1, \phi, \psi\}$. We know, however, that this set is $\mathbb{Q}$-linearly independent, so we arrive at a contradiction unless $b = c = d = 0$. □

2.7 Divisors and the Weil Paring

So far we have managed to avoid the concept of a divisor. This is only because our summary has not made explicit use of them. It would be a mistake to soft-pedal their importance. They can be used, for example, to construct the unique dual isogeny $\hat{\phi}$ through Picard groups (cf. [Sil 1, § III.4]). They are also essential in the construction of a pairing $e : E[m] \times E[m] \to \mu_m$ (here μ_m is the group of m-th roots of unity). This pairing, together with l-adic Tate modules, provides a means to study objects like $\mathrm{Hom}(E_1, E_2)$, and consequently the endomorphism ring of an elliptic curve. As usual, we do not provide proofs for theorems we cite in this summary.

Let E/K be an elliptic curve and let $\overline{K}[E]_{M_P}$ be the localization of the coordinate ring of E (with coefficients in $\overline{K}$) at the maximal ideal $M_P = \{f \in \overline{K}[E] : f(P) = 0\}$. We define a map

$$\begin{aligned} \mathrm{ord}_P : \overline{K}[E]_{M_P} &\to \mathbb{N} \cup \{\infty\} \\ f &\mapsto \max\{n \in Z : f \in M_P^n\}. \end{aligned}$$

This map is easily extended to $\overline{K}(E)_{M_P}$ by setting $\mathrm{ord}_P(f/g) = \mathrm{ord}_P\, f - \mathrm{ord}_P\, g$.

The *divisor* of $f \in \overline{K}(E)$, denoted $\mathrm{div}\, f$, is the finite formal sum

$$\mathrm{div}\, f = \sum_{P \in E} \mathrm{ord}_P\, f[P].$$

More generally, a *divisor* of a curve E is a formal sum

$$\sum_{P \in E} n_P[P],$$

with $n_P \in \mathbb{Z}$, only finitely of which are nonzero. In other words, the collection of divisors of a curve E, denoted $\mathrm{Div}\, E$, is the free abelian group generated by the points of the curve (hence $\mathrm{Div}\, E$ is a $\mathbb{Z}$-module). The *degree* of a divisor is $\sum n_P$. A divisor is called *principal* when it is of the form $\mathrm{div}\, f$ for some $f \in \overline{K}(E)$.

Two divisors D_1 and D_2 are said to be linearly equivalent if $D_1 - D_2$ is principal. With this equivalence relation we define the Picard group $\mathrm{Pic}(E)$ as the quotient of $\mathrm{Div}(E)$ by the subgroup of principal divisors. Let $\mathrm{Div}^0(E)$ denote the submodule of $\mathrm{Div}\, E$ of divisors of degree zero, and let $\mathrm{Pic}^0(E)$ be the quotient of $\mathrm{Div}^0(E)$ by the subgroup of principal divisors (this quotient makes sense because principal divisors are of zero degree–cf. [Sil 1, § II.3]). A theorem of Abel and Jacobi shows that the map

$$\begin{aligned} E &\to \mathrm{Pic}^0(E) \\ P &\mapsto [P]-[O], \end{aligned}$$

is a group isomorphism (cf. [Hus, Theorem 9.3.5]). The following theorem is an immediate and important consequence of this isomorphism.

Theorem 2.18. *Given a finite collection of integers $\{n_i\}_{i\in I}$ and a corresponding collection of points $\{P_i\}_{i\in I}$ on an elliptic curve E/K such that*

$$\sum_{i\in I} n_i = 0 \quad \textit{and} \quad \sum_{i\in I}[n_i]P_i = O,$$

where the latter sum refers to the group law on the curve, then the divisor $D=\sum n_i[P_i]$ is principal. □

Let $P \in E[m]$. Since $[m]P-[m]O=O$, Theorem 2.18 tells us $m[P]-m[O]=\operatorname{div} f$ for some $f\in \overline{K}(E)$. Since the multiplication-by-m map is surjective for nonzero m (it is a nonconstant isogeny) there is a point $Q\in E$ such that $[m]Q=P$. By Theorem 2.18 there exits another function $g\in\overline{K}(E)$ such that

$$\operatorname{div} g = \sum_{R\in E[m]} [Q+R]-[R].$$

The functions $f\circ[m]$ and g^m have the same divisor. Rescaling f by a constant factor if necessary, it follows that $f\circ[m]=g^m$. Hence, if $S\in E[m]$ and X is any point of E

$$g^m(X+S)=f\circ[m](X)+f\circ[m](S)=f\circ[m](X)=g^m(X),$$

from which it follows that $g(X+S)=e_m(S,P)g(X)$, for a certain m-th root of unity $e_m(S,P)$.

Theorem 2.19 (Properties of the Weil pairing). *Let E/K be an elliptic curve and let m be a nonzero integer prime to the characteristic of K. Then the pairing $e_m : E[m]\times E[m]\to\mu_m$ defined above has the following properties, cf. [Shi, §4.3] or [Sil 1, § III.8].*

(i) It is bilinear, i.e.

$$\begin{aligned} e_m(S_1+S_2,P) &= e_m(S_1,P)e_m(S_2,P), \\ e_m(S,P_1+P_2) &= e_m(S,P_1)e_m(S,P_2). \end{aligned}$$

(ii) It is alternating: $e_m(S,P)e_m(P,S)=1$.

(iii) It is nondegenerate: if $e_m(S,P)=1$ for all $S\in E[m]$, then $P=O$.

(vi) It is Galois invariant: for every $\sigma\in\mathrm{Gal}(\overline{K}/K)$

$$e_m(S,P)^\sigma = e_m(S^\sigma,P^\sigma).$$

(v) It is compatible: given $S \in E[mn]$ and $P \in E[n]$ then

$$e_{mn}(S, P) = e_m([n]S, P). \qquad \square$$

Let l be a prime number different from $\operatorname{char} K$. We would like to fit the pairings $e_{l^n} : E[l^n] \times E[l^n] \to \mu_{l^n}$ together to get a pairing on the Tate module $e : T_l(E) \times T_l(E) \to T_l(\mu)$. In order to do this, the parings e_{l^n} must be compatible with the inverse limit construction of the Tate modules

$$T_l(E) = \varprojlim_n E[l^n] \quad \text{and} \quad T_l(\mu) = \varprojlim_n \mu_{l^n},$$

which are taken with respect to the multiplication by $[l]$ and l maps, respectively. It suffices to show

$$e_{l^{n+1}}(S, P)^l = e_{l^n}([l]S, [l]P). \tag{2.3}$$

This is done using the linearity and compatibility of the paring. We conclude this chapter with the following important result.

Theorem 2.20. *Given an elliptic curve E, there exists a pairing $e : T_l(E) \times T_l(E) \to T_l(\mu)$ that is bilinear, alternating nondegenerate and Galois invariant. If $\phi : E_1 \to E_2$ is an isogeny then ϕ and $\hat{\phi}$ are adjoints for the pairing.* $\square$

Chapter 3

Complex Multiplication over $\mathbb{C}$

In this chapter we will discuss some basic aspects of the theory of Complex Multiplication (CM) over the field of complex numbers and will introduce Gross–Zagier numbers, the driving force behind this paper.

Recall an elliptic curve $E/\mathbb{C}$ is said to admit complex multiplication when $\operatorname{End}(E)$ contains endomorphisms other than the multiplication by m maps; in such a case we say E is a CM-curve. We showed in Chapter 2 that if Λ is a lattice associated to a CM-curve E, with basis $[1, \tau]$, then $\mathbb{Q}(\tau)$ is a quadratic imaginary field and $\operatorname{End}(E)$ is an order in this field (cf. Theorem 2.11). This means $\operatorname{End}(E) \otimes \mathbb{Q}$ is isomorphic to $\mathbb{Q}(\tau)$. If $\operatorname{End}(E) \cong \mathcal{O} \subset \mathbb{C}$ and $K = \mathcal{O} \otimes \mathbb{Q}$ we will say "E has complex multiplication by $\mathcal{O}$."

Following ideas set out in [Sil 2, Ch. II] and [Se 2] we will focus on elliptic curves with complex multiplication by the ring of integers $\mathcal{O}_K$ (the *maximal* order) of a given imaginary quadratic field K—complex multiplication by non-maximal order is much harder, and we will not need to delve into it to study Gross–Zagier numbers.

The j-invariant of a CM-curve is called a *singular modulus*; it is an algebraic integer and it generates a finite abelian extension of K. As an example of the theory developed in this chapter, we will compute the irreducible polynomial for j in the cases $K = \mathbb{Q}(\sqrt{-21}), \mathbb{Q}(\sqrt{-133})$. The constant term of these polynomials are the first examples of Gross–Zagier numbers we will meet.

3.1 Complex Multiplication over $\mathbb{C}$

We begin our discussion of CM-curves with an example.

Example 3.1. Let $E/\mathbb{C}$ be a curve whose automorphism group $\operatorname{Aut} E$ is strictly larger than $\{\pm 1\}$. Then E must admit complex multiplication, otherwise $\operatorname{End} E \cong \mathbb{Z}$ and therefore $\operatorname{Aut} E \cong \{\pm 1\}$. Let Λ be a lattice for E. By Theorem 2.10, $\operatorname{Aut}(E) \cong \{\alpha \in \mathbb{C} \mid \alpha\Lambda = \Lambda\}$. Suppose $\{\pm 1, \pm i\} \subset \operatorname{Aut}(E)$, i.e., $i\Lambda = \Lambda$. Then

$$g_3(\Lambda) = g_3(i\Lambda) = i^6 g_3(\Lambda) = -g_3(\Lambda)$$

and so $g_3(\Lambda) = 0$, which means $j(E) = 1728$. Since the j-invariant classifies curves up to isomorphism, any curve of the form $y^2 = 4x^3 - g_2 x$ admits complex multiplication.

Given an elliptic curve $E/\mathbb{C}$ with complex multiplication by the ring of integers $\mathcal{O}_K$ of the quadratic imaginary field K, we would like to embed $\mathcal{O}_K$ (as a subset of the complex numbers) in $\operatorname{End} E$. One way to do this is to choose, without loss of generality, a lattice Λ such that $E \cong E_\Lambda$;

by Theorem 2.10, $\operatorname{End}(E_\Lambda) \cong \{\alpha \in \mathbb{C} \mid \alpha\Lambda \subset \Lambda\} = \mathcal{O}_K$. Each $\alpha \in \mathcal{O}_K$ gives rise to a map $[\alpha] \in \operatorname{End}(E_\Lambda)$ determined by the commutativity of the following diagram

$$\begin{tikzcd}
\mathbb{C}/\Lambda \arrow[r, "\phi_\alpha", "z\mapsto \alpha z"'] \arrow[d, "f"'] & \mathbb{C}/\Lambda \arrow[d, "f"'] \\
E_\Lambda \arrow[r, "{[\alpha]}"] & E_\Lambda
\end{tikzcd}$$

where f is the map of Theorem 2.8. The advantage of this embedding is that it behaves nicely with respect to invariant differentials. Indeed, let $\omega \in \Omega_E$ be an invariant differential. Note that any two nonzero invariant differentials in Ω_{E_Λ} differ by a multiplicative constant because their quotient is translation invariant. Then, using the commutativity of the above diagram, and the fact that the pullback of an invariant differential ω through f is a constant multiple of the invariant differential dz of $\mathbb{C}/\Lambda$, we conclude that

$$[\alpha]^*\omega = \alpha\omega \quad \text{for all } \alpha \in \mathcal{O}. \tag{3.1}$$

The embedding $\mathcal{O}_K \hookrightarrow \operatorname{End} E$ thus obtained is known as the *normalized embedding.* Two isogenous elliptic curves have normalized embeddings equal up to conjugation by an isogeny between the curves (cf. [Sil 2, § II.1]).

Suppose we start now with an arbitrary quadratic imaginary field K. There is a practical way of finding elliptic curves with complex multiplication by the ring of integers $\mathcal{O}_K$. The key idea is to note that a nonzero fractional ideal $\mathfrak{a}$ of K is a lattice in $\mathbb{C}$. By Theorem 2.10 we know $\operatorname{End}(E_\mathfrak{a}) \cong \{\alpha \in \mathbb{C} \mid \alpha\mathfrak{a} \subset \mathfrak{a}\}$. But this last set is just $\{\alpha \in K \mid \alpha\mathfrak{a} \subset \mathfrak{a}\}$ because $\mathfrak{a} \subset K$, and this set in turn is $\mathcal{O}_K$ since $\mathfrak{a}$ is a fractional ideal.

Recall that two homothetic lattices $\mathfrak{a}$ and $c\mathfrak{a}$ give rise to isomorphic elliptic curves (Theorem 2.9), i.e., two elements in the same ideal class of the ideal class group $C(\mathcal{O}_K)$ give rise to isomorphic elliptic curves. This shows that the map from $C(\mathcal{O}_K)$ to the set

$$\begin{aligned}
\mathfrak{E}(\mathcal{O}_K) &= \frac{\{\text{elliptic curves } E \mid \operatorname{End} E \cong \mathcal{O}_K\}}{\mathbb{C}\text{-isomorphism}} \\
&= \frac{\{\text{lattices } \Lambda \mid \operatorname{End} E_\lambda \cong \mathcal{O}_K\}}{\text{homothety}}
\end{aligned}$$

given by $\bar{\mathfrak{a}} \mapsto E_\mathfrak{a}$ is well defined. Moreover, this map gives rise to an action of $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ which is simply transitive. We will now study this action.

Lemma 3.1. *Let $\mathfrak{a}$ be a nonzero fractional ideal of $\mathcal{O}_K$ and Λ a lattice such that the elliptic curve E_Λ belongs to $\mathfrak{E}(\mathcal{O}_K)$. Then the product*

$$\mathfrak{a}\Lambda = \{\alpha_1\lambda_1 + \cdots + \alpha_r\lambda_r \mid \alpha_i \in \mathfrak{a}, \lambda_i \in \Lambda\}$$

is a lattice in $\mathbb{C}$.

Proof. By definition of fractional ideal, there is an integer d_1 such that $d_1\mathfrak{a} \subset \mathcal{O}_K$. Since $\operatorname{End}(E_\Lambda) = \mathcal{O}_K$, it follows that

$$d_1\mathfrak{a} \subset \mathcal{O}_K \implies \mathfrak{a}\Lambda \subset (1/d_1)\Lambda$$

On the other hand we can choose an integer d_2 such that $d_2\mathcal{O}_K \subset \mathfrak{a}$, so that $d_2\Lambda \subset \mathfrak{a}\Lambda$. We conclude that $\mathfrak{a}\Lambda$ is a discrete $\mathbb{Z}$-module of the same rank as Λ, i.e., it is a lattice in $\mathbb{C}$. □

From Theorem 2.10, it follows that

$$\mathrm{End}(\mathfrak{a}\Lambda) \cong \{\alpha \in \mathbb{C} \mid \alpha\mathfrak{a}\Lambda \subset \mathfrak{a}\Lambda\} = \{\alpha \in \mathbb{C} \mid \alpha\Lambda \subset \Lambda\} = \mathrm{End}(E_\Lambda) \cong \mathcal{O}_K.$$

Hence $E_{\mathfrak{a}\Lambda} \in \mathfrak{E}(\mathcal{O}_K)$. We define the action of $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ by

$$\overline{\mathfrak{a}} * E_\Lambda = E_{\mathfrak{a}^{-1}\Lambda}.$$

(It is straightforward to check the above *is* an action.) This action is well-defined, i.e., $E_{\mathfrak{a}^{-1}\Lambda} \cong E_{\mathfrak{b}^{-1}\Lambda}$ if and only if $\overline{\mathfrak{a}} = \overline{\mathfrak{b}}$. Indeed, $E_{\mathfrak{a}^{-1}\Lambda} \cong E_{\mathfrak{b}^{-1}\Lambda}$ if and only if $\mathfrak{a}^{-1}\Lambda$ and $\mathfrak{b}^{-1}\Lambda$ are homothetic lattices. This means there exists $\alpha \in \mathbb{C}^*$ such that $\mathfrak{a}^{-1}\Lambda = \alpha\mathfrak{b}^{-1}\Lambda$; equivalently, $\Lambda = \alpha\mathfrak{a}^{-1}\mathfrak{b}\Lambda$ (or $\Lambda = \alpha^{-1}\mathfrak{a}\mathfrak{b}^{-1}\Lambda$ because $\mathfrak{b}^{-1}\Lambda = \Lambda\mathfrak{b}^{-1}$). If both $\alpha\mathfrak{a}^{-1}\mathfrak{b}$ and $\alpha^{-1}\mathfrak{a}\mathfrak{b}^{-1}$ preserve Λ, they must be contained in $\mathcal{O}_K$, and of course this happens if and only if $\overline{\mathfrak{a}} = \overline{\mathfrak{b}}$ in $C(\mathcal{O}_K)$.

Theorem 3.2. *The action of $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ given by $\overline{\mathfrak{a}} * E_\Lambda = E_{\mathfrak{a}^{-1}\Lambda}$ is simply transitive.*

Proof. Let E_{Λ_1} and E_{Λ_2} be elements of $\mathfrak{E}(\mathcal{O}_K)$. To show the action is transitive, it suffices to exhibit a nonzero fractional ideal $\mathfrak{a}$ of K such that $\overline{\mathfrak{a}} * E_{\Lambda_1} = E_{\Lambda_2}$. Let λ_1 be a nonzero element of Λ_1 and set $\mathfrak{a}_1 = (1/\lambda_1)\Lambda_1$.

Since $\mathfrak{a}_1$ is a lattice, Theorem 2.10 tell us that

$$\mathrm{End}(E_{\mathfrak{a}_1}) \cong \{\alpha \in \mathbb{C} \mid (\alpha/\lambda_1)\Lambda_1 \subset (1/\lambda_1)\Lambda_1\} = \{\alpha \in \mathbb{C} \mid \alpha/\Lambda_1 \subset \Lambda_1\} \cong \mathcal{O}_K;$$

hence $\mathfrak{a}_1 \subset K$. By assumption $\mathfrak{a}_1$ is a finitely generated $\mathcal{O}_K$-module, so it is a fractional ideal of K. Similarly, taking a nonzero element λ_2 of Λ_2 and setting $\mathfrak{a}_2 = (1/\lambda_2)\Lambda_2$ we obtain another fractional ideal of K.

It is easy to see that $(\lambda_2/\lambda_1)\mathfrak{a}_2\mathfrak{a}_1^{-1}\Lambda_1 = \Lambda_2$. Now set $\mathfrak{a} = \mathfrak{a}_2^{-1}\mathfrak{a}_1$. Then a straightforward manipulation shows $\overline{\mathfrak{a}} * E_{\Lambda_1} = E_{\Lambda_2}$, as desired.

Simplicity of the action follows from the fact that if $\mathfrak{a} * E_\Lambda = \mathfrak{b} * E_\Lambda$ then $\overline{\mathfrak{a}} = \overline{\mathfrak{b}}$, something we have already shown. □

Corollary 3.3. *Let K be a quadratic imaginary field. Then there are finitely many isomorphism classes of elliptic curves with complex multiplication by $\mathcal{O}_K$.*

Proof. Since the action of $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ is simply transitive, there are as many isomorphism classes of curves with complex multiplication by $\mathcal{O}_K$ as there are ideal classes in $C(\mathcal{O}_K)$. The group $C(\mathcal{O}_K)$ is finite (cf. [Mar, Ch. 5]) and the result follows immediately. □

Given the group action above, it is natural to look at the map $E \mapsto \overline{\mathfrak{a}} * E$. The kernel of this map leads to the study of certain finite subgroups of points of E. Indeed, let Λ be a lattice such that $E \cong E_\Lambda$. Then the kernel of $E \mapsto \overline{\mathfrak{a}} * E$ is the kernel of the homomorphism $\mathbb{C}/\Lambda \to \mathbb{C}/\mathfrak{a}^{-1}\Lambda$ given by $z \mapsto z$. This set, however, is just $\mathfrak{a}^{-1}\Lambda/\Lambda$. Hence

$$\begin{aligned}\ker(E \mapsto \overline{\mathfrak{a}} * E) &= \{z \in \mathbb{C}/\Lambda \mid \alpha z = 0 \ \forall\, \alpha \in \mathfrak{a}\} \\ &\cong \{P \in E \mid [\alpha]P = 0 \ \forall\, \alpha \in \mathfrak{a}\} =: E[\mathfrak{a}].\end{aligned}$$

We say $E[\mathfrak{a}]$ is the group of $\mathfrak{a}$-torsion points of E. The reader should note that the above isomorphism depends on the embedding $[\cdot] : \mathcal{O}_K \to \mathrm{End}\, E$; we always use the normalized embedding.

If $E \in \mathfrak{E}(\mathcal{O}_K)$, one may use techniques of commutative algebra to show $E[\mathfrak{a}]$ is a free $\mathcal{O}_K/\mathfrak{a}$-module of rank 1 (see, for example, [Sil 2, § II.1]). In this case

$$\deg(E \mapsto \bar{\mathfrak{a}} * E) = \#E[\alpha] = \#|\mathcal{O}_K/\mathfrak{a}| = N^K_{\mathbb{Q}}\mathfrak{a}. \tag{3.2}$$

In particular, the endomorphism $[\alpha] : E \to E$ has degree $|N^K_{\mathbb{Q}}\alpha|$ because for a nonconstant separable isogeny ϕ it is true that $\deg\phi = \#\ker\phi$, and so

$$\deg[\alpha] = \#\ker[\alpha] = \#E[\alpha\mathcal{O}_K] = |N^K_{\mathbb{Q}}\alpha|. \tag{3.3}$$

3.2 The Field of Definition of a CM–Curve

We now turn our attention to the field of definition of a CM–curve. We will show that every elliptic curve with complex multiplication is defined over an algebraic extension of $\mathbb{Q}$.

Theorem 3.4. *Let E be an elliptic curve with complex multiplication by $\mathcal{O}_K$. Then $j(E)$ is an algebraic number.*

Proof. We will show $j(E)^\sigma$ takes on finitely many values as σ ranges through $\operatorname{Aut}\mathbb{C}$. To begin, note that $j(E^\sigma) = j(E)^\sigma$ because E^σ is obtained from E by applying σ to the coefficients of the equation for E and $j(E)$ is a rational function of these coefficients.

On the other hand, it is clear that if $\sigma \in \operatorname{Aut}\mathbb{C}$ and $\phi \in \operatorname{End}E$ then $\phi^\sigma \in \operatorname{End}E^\sigma$, so that $\operatorname{End}E^\sigma \cong \operatorname{End}E \cong \mathcal{O}_K$. Hence $j(E)^\sigma$ is the j-invariant of another curve in one of the isomorphism classes of $\mathfrak{E}(\mathcal{O}_K)$. We know from Corollary 3.3 that there are finitely many such classes, and since the isomorphism class of an elliptic curve is determined by its j-invariant, it follows that $j(E)^\sigma$ takes on finitely many values as σ ranges through $\operatorname{Aut}\mathbb{C}$, so $[\mathbb{Q}(j(E)) : \mathbb{Q}] < \infty$ and $j(E)$ is algebraic. □

Remark 3.5. We will later strengthen the above theorem and prove that singular j's are algebraic integers (cf. Theorem 3.20).

Corollary 3.6.

$$\mathfrak{E}(\mathcal{O}_K) \cong \frac{\{\textit{elliptic curves } E/\overline{\mathbb{Q}} \textit{ with } \operatorname{End}(E) \cong \mathcal{O}_K\}}{\textit{isomorphism over } \overline{\mathbb{Q}}}.$$

Proof. Let F be a field and denote

$$\mathfrak{E}_F(\mathcal{O}_K) = \frac{\{\text{elliptic curves } E/F \text{ with } \operatorname{End}(E) \cong \mathcal{O}_K\}}{\text{isomorphism over } F}.$$

If we fix an embedding of $\overline{\mathbb{Q}}$ into $\mathbb{C}$ then there is a natural map

$$f : \mathfrak{E}_{\overline{\mathbb{Q}}}(\mathcal{O}_K) \to \mathfrak{E}_{\mathbb{C}}(\mathcal{O}_K).$$

We will show this map is a bijection. Let $E/\mathbb{C}$ represent an element of $\mathfrak{E}_{\mathbb{C}}(\mathcal{O}_K)$. There exists a curve $E'/\mathbb{Q}(j(E))$ with $j(E') = j(E)$ (cf. §2.1). This means E' is isomorphic to E over $\mathbb{C}$, and since $j(E)$ is algebraic one sees that $E' \in \mathfrak{E}_{\overline{\mathbb{Q}}}(\mathcal{O}_K)$. Hence $f(E') = E$, which shows f is surjective.

To see injectivity suppose $E_1/\overline{\mathbb{Q}}$ and $E_2/\overline{\mathbb{Q}}$ are two elliptic curves such that $f(E_1) = f(E_2)$. It follows that $j(E_1) = j(E_2)$, and since E_1 and E_2 are defined over $\overline{\mathbb{Q}}$, equality of their j-invariants means they are isomorphic over $\overline{\mathbb{Q}}$, i.e., they belong to the same class in $\mathfrak{E}_{\overline{\mathbb{Q}}}(\mathcal{O}_K)$. □

In the sequel, we will identify $\mathfrak{E}(\mathcal{O}_K)$ with $\mathfrak{E}_{\overline{\mathbb{Q}}}(\mathcal{O}_K)$.

Although a CM–elliptic curve E is defined over an algebraic extension of $\mathbb{Q}$, its endomorphisms need not be defined over the same extension. They almost are, however, as the following theorem shows (cf. [Sil 2, Theorem II.2.2(c)]).

Theorem 3.7. *Let E be an elliptic curve defined over a subfield L of $\mathbb{C}$, with complex multiplication by $\mathcal{O}_K$. Then an endomorphism of E is defined over the compositum LK.* □

There is a natural action of $\mathrm{Gal}(\overline{K}/K)$ on $\mathfrak{E}(\mathcal{O}_K)$ that sends E to E^σ for a given $\sigma \in \mathrm{Gal}(\overline{K}/K)$. On the other hand, the action of the class group $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ is simply transitive (Theorem 3.2), so there exists an $\overline{\mathfrak{a}} \in C(\mathcal{O}_K)$ such that $\overline{\mathfrak{a}} * E = E^\sigma$. We can therefore define a map

$$\Theta : \mathrm{Gal}(\overline{K}/K) \to C(\mathcal{O}_K),$$

characterized by $E^\sigma = \Theta(\sigma) * E$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$. This map is the key to understanding the field extension $K(j(E))$.

Since the j-invariant classifies elliptic curves with complex multiplication by $\mathcal{O}_K$ up to $\overline{\mathbb{Q}}$-isomorphism (Corollary 3.6), the map Θ is also characterized by $j(E_\Lambda)^\sigma = j(E_{\Theta(\sigma)^{-1}\Lambda})$. So we see $\Theta(\sigma)$ depends on how the lattice associated to an elliptic curve changes under multiplication by an ideal. The map Θ provides a bridge between the algebraic action of σ and the analytic action of multiplication by $\Theta(\sigma)^{-1}$ (because j is an analytic function of Λ).

The map Θ is a homomorphism. Indeed,

$$\Theta(\sigma\tau) * E = E^{\sigma\tau} = (E^\sigma)^\tau = (\Theta(\sigma) * E)^\tau = \Theta(\tau) * (\Theta(\sigma) * E) = (\Theta(\sigma)\Theta(\tau)) * E.$$

The last equality holds because $*$ is an action and $C(\mathcal{O}_K)$ is an abelian group.

Our next task is to show that the property $E^\sigma = \Theta(\sigma) * E$ determines Θ uniquely. We will need the following lemma, whose proof we omit (cf. [Sil 2, Proposition II.2.5]). The proof is quite hard, and in some sense it is at the heart of the analytic–algebraic bridge described above, yet we hope the reader may still appreciate the power of complex multiplication without the technicalities of commutative algebra involved in this proof.

Lemma 3.8. *Let $E/\overline{\mathbb{Q}}$ represent an element in $\mathfrak{E}(\mathcal{O}_K)$, let $\overline{\mathfrak{a}} \in C(\mathcal{O}_K)$ and let $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Then*

$$(\overline{\mathfrak{a}} * E)^\sigma = \overline{\mathfrak{a}}^\sigma * E^\sigma.$$

□

Theorem 3.9. *The homomorphism $\Theta : \mathrm{Gal}(\overline{K}/K) \to C(\mathcal{O}_K)$, characterized by the property that $E^\sigma = \Theta(\sigma) * E$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$ and all $E \in \mathfrak{E}(\mathcal{O}_K)$ is independent of the choice of $E \in \mathfrak{E}(\mathcal{O}_K)$.*

Proof. Let E_1 and E_2 be two curves in $\mathfrak{E}(\mathcal{O}_K)$ and let $\sigma \in \mathrm{Gal}(\overline{K}/K)$. Then by Theorem 3.2 $E_1^\sigma = \overline{\mathfrak{a}_1} * E_1$ and $E_2^\sigma = \overline{\mathfrak{a}_2} * E_2$ for some $\overline{\mathfrak{a}_1}, \overline{\mathfrak{a}_2} \in C(\mathcal{O}_K)$. We want to show that $\overline{\mathfrak{a}_1} = \overline{\mathfrak{a}_2}$. Since $C(\mathcal{O}_K)$ acts transitively on $\mathfrak{E}(\mathcal{O}_K)$ there is a $\overline{\mathfrak{b}} \in C(\mathcal{O}_K)$ such that $E_2 = \overline{\mathfrak{b}} * E_1$. Then

$$(\overline{\mathfrak{b}} * E_1)^\sigma = E_2^\sigma = \overline{\mathfrak{a}_2} * E_2 = \overline{\mathfrak{a}_2} * (\overline{\mathfrak{b}} * E_1) = (\overline{\mathfrak{a}_2}\overline{\mathfrak{b}}\overline{\mathfrak{a}_1}^{-1}) * E_1^\sigma.$$

By Lemma 3.8, $(\overline{\mathfrak{b}} * E)^\sigma = \overline{\mathfrak{b}}^\sigma * E^\sigma$ and since $\overline{\mathfrak{b}}^\sigma = \overline{\mathfrak{b}}$ (because $\overline{\mathfrak{b}} \in K$ and $\sigma \in \mathrm{Gal}(\overline{K}/K)$), we conclude that

$$\overline{\mathfrak{b}} * E_1^\sigma = (\overline{\mathfrak{a}_2}\overline{\mathfrak{b}}\overline{\mathfrak{a}_1}^{-1}) * E_1^\sigma.$$

Since the action of $C(\mathcal{O}_K)$ on $\mathfrak{E}(\mathcal{O}_K)$ is well-defined, it follows that $\overline{\mathfrak{a}_1} = \overline{\mathfrak{a}_2}$. □

3.3 The Hilbert Class Field of an Imaginary Quadratic Field

Let L and K be number fields, L a finite extension of K. A prime ideal $\mathfrak{p} \in \mathcal{O}_K$ is said to be *unramified* in L if

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{P}_1\mathfrak{P}_2 \cdots \mathfrak{P}_k,$$

where the $\mathfrak{P}_i$ are *distinct* prime ideals of $\mathcal{O}_L$. Given an elliptic curve $E/\mathbb{C}$ with complex multiplication by $\mathcal{O}_K$, the extension $K(j(E))$ is the maximal unramified abelian extension of K, i.e., it is an extension such that

(i) Every prime ideal of $\mathcal{O}_K$ is unramified in $K(j(E))$.

(ii) The group $\mathrm{Gal}(K(j(E))/K)$ is abelian.

(iii) Every extension L of K that satisfies properties analogous to (i) and (ii) is contained in $K(j(E))$.

We say $K(j(E))$ is the *Hilbert class field* of K.

We will show $K(j(E))$ satisfies property (ii) above, and content ourselves with the statement that $K(j(E))$ is also the maximal unramified extension among all abelian extensions of K. A proof of this theorem is somewhat difficult and requires an understanding of the statements of class field theory. In a sense, the general theory of complex multiplication is best thought of as an explicit realization of the class field theory of quadratic imaginary fields. We do not follow this avenue of thought, however. A proof of the theorem, together with the necessary background in class field theory is in [Sil 2, Ch. II.3–4].

Theorem 3.10. *Let K be a quadratic imaginary field, and let $E/\mathbb{C}$ be an elliptic curve with complex multiplication by $\mathcal{O}_K$. Then the extension $K(j(E))$ is finite abelian.*

Proof. We know the extension $K(j(E))$ is finite because $j(E)$ is an algebraic number (Theorem 3.4). Let L/K be the fixed field of the kernel of Θ, i.e., $\mathrm{Gal}(\overline{K}/L) = \ker\Theta$. We claim that $L = K(j(E))$. Indeed, given $\sigma \in \mathrm{Gal}(\overline{K}/K)$, the series of equivalences

$$\begin{aligned}\Theta(\sigma) = 1 &\iff \Theta(\sigma) * E = E\\ &\iff E^\sigma = E\\ &\iff j(E^\sigma) = j(E)\\ &\iff j(E)^\sigma = j(E)\end{aligned}$$

shows that $\mathrm{Gal}(\overline{K}/L) = \mathrm{Gal}(\overline{K}/K(j(E)))$, so $L = K(j(E))$. On the other hand, the map Θ takes $\mathrm{Gal}(L/K)$ injectively into $C(\mathcal{O}_K)$. Indeed, if $\Theta(\sigma) = \overline{1}$ for some $\sigma \in \mathrm{Gal}(L/K)$ then σ belongs to $\ker\Theta$, which means σ fixes L by the definition of L and hence $\sigma = 1$. Since $C(\mathcal{O}_K)$ is abelian, the injection $\mathrm{Gal}(L/K) \hookrightarrow C(\mathcal{O}_K)$ shows $\mathrm{Gal}(L/K) = \mathrm{Gal}(K(j(E))/K)$ is abelian. □

Theorem 3.11. *Let K be a quadratic imaginary field, and let E be an elliptic curve with complex multiplication by $\mathcal{O}_K$. Then the Hilbert Class Field of K is $K(j(E))$.* □

Corollary 3.12. *The degree of the extension $\mathbb{Q}(j(E))$ over $\mathbb{Q}$ is equal to the class number h of K.*

Proof. From class field theory, the Hilbert class field of K has Galois group isomorphic to $C(\mathcal{O}_K)$, hence $[K(j(E)) : K] = h$. The tower law then tells us that

$$[K(j(E)) : \mathbb{Q}] = [K(j(E)) : K] \cdot [K : \mathbb{Q}] = 2h.$$

On the other hand, it follows from $[K : \mathbb{Q}] = 2$ that $[K(j(E)) : \mathbb{Q}(j(E))] \leq 2$, and the proof of Theorem 3.4 shows that

$$[\mathbb{Q}(j(E)) : \mathbb{Q}] \leq \#C(\mathcal{O}_K) = h.$$

Therefore

$$2h = [K(j(E)) : \mathbb{Q}] = [K(j(E)) : \mathbb{Q}(j(E))] \cdot [\mathbb{Q}(j(E)) : \mathbb{Q}] \leq 2 \cdot h,$$

and $[\mathbb{Q}(j(E)) : \mathbb{Q}] = h$, as desired. □

Corollary 3.13. *If $\{E_1, \ldots, E_h\}$ is a set of representatives of $\mathfrak{E}(\mathcal{O}_K)$, then $J = \{j(E_1), \ldots, j(E_h)\}$ is a full set of* $\mathrm{Gal}(\overline{K}/K)$ *conjugates for the j-invariant of any curve in $\mathfrak{E}(\mathcal{O}_K)$.*

Proof. The group $C(\mathcal{O}_K)$ acts on J by $\bar{\mathfrak{a}} \cdot j(E) \mapsto j(\bar{\mathfrak{a}} * E)$ and the group $\mathrm{Gal}(\overline{K}/K)$ acts on J by $\sigma \cdot j(E) \mapsto j(E^\sigma) = j(\Theta(\sigma) * E)$. The map Θ identifies the two actions. From Theorem 3.2 it follows that $\mathrm{Gal}(\overline{K}/K)$ acts transitively on J, which means J is a full set of $\mathrm{Gal}(\overline{K}/K)$ conjugates. □

Corollary 3.14. *Let $J = \{j(E_1), \ldots, j(E_h)\}$ be as in Corollary 3.13. Then J is a full set of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *conjugates for the j-invariant of any curve in $\mathfrak{E}(\mathcal{O}_K)$. The product*

$$N(j_K) := \prod_{i=1}^{h} j(E_i)$$

is called the absolute norm *of the j-invariant of a curve in $\mathfrak{E}(\mathcal{O}_K)$.*

Proof. This result follows directly from Corollaries 3.6 and 3.13. □

3.4 The Hilbert Class Field of $\mathbb{Q}(\sqrt{-21})$

In this section we apply our results above to compute the Hilbert class field of $\mathbb{Q}(\sqrt{-21})$. One may check that the ideal class group of this field is the Klein group. Explicitly,

$$C(\mathcal{O}_K) = \{[\mathcal{O}_K], [P_2], [P_3], [P_5]\},$$

where

$$\begin{aligned} P_2 &= (2, \sqrt{-21} - 1), \\ P_3 &= (3, \sqrt{-21}), \\ P_5 &= (5, \sqrt{-21} - 3), \end{aligned}$$

and the relations $[P_2]^2 = [\mathcal{O}_K]$, $[P_3]^2 = [\mathcal{O}_K]$ and $[P_5] = [P_2] \cdot [P_3]$ hold.

As discussed earlier, one way to obtain an elliptic curve with complex multiplication by $\mathcal{O}_K$ is to consider the curve that a nonzero fractional ideal (as a lattice in $\mathbb{C}$) gives rise to via the isomorphism of Theorem 2.8. For example, consider $P_2 = [2, \sqrt{-21} - 1]$ as a lattice; then the elliptic curve E_{P_2} has equation

$$y^2 = 4x^3 - g_2(P_2)x - g_3(P_2)$$

(cf. §2.1); it has complex multiplication by $\mathcal{O}_K$. In general, if $\Lambda = [\omega_1, \omega_2]$ is a lattice in $\mathbb{C}$ for which $\tau := \omega_2/\omega_1$ has positive imaginary part we have

$$j(\tau) = j(\Lambda) := j(E_\Lambda) = \frac{1728 g_2(\Lambda)^3}{g_2(\Lambda)^3 - 27 g_3(\Lambda)^2}.$$

Note that it makes sense to speak of $j(\tau)$ since $[1, \tau]$ and Λ are homothetic lattices provided $\mathrm{Im}(\tau) \neq 0$, and the j-invariant is determined up to lattice homothety (i.e., elliptic curve isomorphism).

With the aid of the widely available PARI-GP software (which uses the method of q-expansions to compute j-invariants—cf.), we compute the approximation

$$j(P_2) = j\left(\frac{\sqrt{-21} - 1}{2}\right) = -1787216.6012476570198674 - 4.17619485 \times 10^{-51} i$$

Similarly, we compute

$$\begin{aligned} j(\mathcal{O}_K) &= j(\sqrt{-21}) = 3196802718613.9132928032899986 + 10^{-45} i \\ j(P_3) &= j\left(\frac{\sqrt{-21}}{3}\right) = 15488.6808931242445923 + 10^{-53} i \\ j(P_5) &= j\left(\frac{\sqrt{-21} - 3}{5}\right) = 58.0070617294852765 + 2.6896583624964495 \times 10^{-55} i \end{aligned}$$

We know that the Hilbert class field of K is just $K(j(\mathcal{O}_K))$, yet the above approximation does not give us this field explicitly. We remedy this situation by "guessing" the minimal polynomial for $j(\mathcal{O}_K)$ and solving it explicitly. By Corollary 3.13, the set $J = \{j(\mathcal{O}_K), j(P_2), j(P_3), j(P_5)\}$ is a full set of $\mathrm{Gal}(\overline{K}/K)$ conjugates for $j(\mathcal{O}_K)$ in $\mathfrak{E}(\mathcal{O}_K)$. Since each of these conjugates is in the Hilbert class field of K, the irreducible polynomial for $j(\mathcal{O}_K)$ is just

$$P(X) = (X - j(\mathcal{O}_K))(X - j(P_2))(X - j(P_3))(X - j(P_5)).$$

Using the above approximations for the elements of J, we conjecture, with a wide margin of error, that the irreducible polynomial for $j(\mathcal{O}_K)$ is

$$\begin{aligned} P(X) = x^4 &- 3196800946944 x^3 - 5663679223085309952 x^2 \\ &+ 888212465898100893941 76 x - 5133201653210986057826304 \end{aligned} \tag{3.4}$$

Assuming that this *is* the irreducible polynomial, we may hope to solve the quartic equation by radicals and compute $j(\mathcal{O}_K)$ exactly. Using Descartes' method for solving quartic equations by radicals (cf. [Es, §2.3]), we find that

$$\begin{aligned} j(\mathcal{O}_K) = &\sqrt{7}(2^{10} \cdot 3^3 \cdot 7 \cdot 13 \cdot 19 \cdot 71 \cdot 89) + \sqrt{3}(2^8 \cdot 3^4 \cdot 7 \cdot 13 \cdot 29)(\sqrt{7} \cdot 3187 + 2^4 \cdot 17 \cdot 31) \\ &+ 799200236736 \end{aligned} \tag{3.5}$$

Hence, the Hilbert class field of K is $K(\sqrt{7}, \sqrt{3})$. This computation relies heavily on the assumption that (3.4) is the irreducible polynomial for $j(\mathcal{O}_K)$. Fortunately, we can use results from the genus theory of binary quadratic forms to confirm our observation. The *genus field* of a quadratic imaginary field K is the maximal unramified extension of K that is abelian over $\mathbb{Q}$.

Theorem 3.15. *Let K be an imaginary quadratic field of discriminant d_K. Let $p_1, \dots, p_r$ be the odd primes dividing d_K. Set $p_i^* = (-1)^{(p_i-1)/2} p_i$. Then $K(\sqrt{p_1^*}, \dots, \sqrt{p_r^*})$ is the genus field of K (cf. [Cox, Theorem 6.1]).* □

Sometimes the genus field and the Hilbert class field coincide. A discriminant d_K is said to be a *convenient number* if every element of the ideal class group $C(\mathcal{O}_K)$ has order 2. When the discriminant of a quadratic imaginary field is a convenient number the genus field and the Hilbert class field are the same. Since the ideal class group of $\mathbb{Q}(\sqrt{-21})$ is the Klein group, it follows from Theorem 3.15 that the Hilbert class field of $K = \mathbb{Q}(\sqrt{-21})$ is $K(\sqrt{-7}, \sqrt{-3}) = K(\sqrt{7}, \sqrt{3})$, as we predicted.

The happy coincidence which allowed us to check our result seems to be rare. It was conjectured by Gauss that there are only 65 convenient numbers (Euler gave a list of them), but this question is still an open problem.

The numbers in computations like the ones above can be enormous. For example, if we try to compute the Hilbert class field of $K = \mathbb{Q}(\sqrt{-133})$ by the above method we find that the ideal class group $C(\mathcal{O}_K)$ is once again the Klein group:

$$C(\mathcal{O}_K) = \{[\mathcal{O}_K], [P_2], [P_7], [Q]\},$$

where

$$\begin{aligned} P_2 &= (2, \sqrt{-133} + 1), \\ P_7 &= (7, \sqrt{-133}), \\ Q &= (7 + \sqrt{-133}, -2\sqrt{-133}), \end{aligned}$$

and the relations $[P_2]^2 = [\mathcal{O}_K]$, $[P_7]^2 = [\mathcal{O}_K]$ and $[Q] = [P_2] \cdot [P_7]$ hold. This time the conjectured polynomial for $j(\mathcal{O}_K)$ comes out to be

$$\begin{aligned} P(X) &= (X - j(\mathcal{O}_K))(X - j(P_2))(X - j(P_7))(X - j(Q)) \\ &= x^4 - 2947890901909813907417747913600 0x^3 \\ &\quad - 160054212938390343773833947283393690785408000000x^2 \\ &\quad + 513153774061019296207088016300696964327219200000000 0x \\ &\quad - 190775429933529456809610289946972713082880000000000 00. \end{aligned}$$

Using Descartes' method for solving quartics is a cumbersome task to carry out with such large coefficients. Fortunately in this case d_K is also a convenient number, so the Hilbert class field in this case is $K(\sqrt{-7}, \sqrt{-19})$ (cf. Theorem 3.15).

3.5 Primes Dividing $N(j_K)$: a few Examples

When computing the minimal polynomial of $j(\mathcal{O}_K)$ for $K = \mathbb{Q}(\sqrt{-133})$ we saw that

$$\begin{aligned} N(j_K) &= -190775429933529456809610289946972713082880000000000 00 \\ &= -(2^8 \cdot 3^4 \cdot 5^4 \cdot 11^2 \cdot 23^2 \cdot 29^2 \cdot 383)^3 \end{aligned}$$

There are two surprising things about this computation. First, the absolute norm is an integer. So far we had only shown that singular moduli are algebraic numbers, so we only expected its absolute

K	Class Number	$N(j_K)$
$\mathbb{Q}(\sqrt{-1})$	1	$(2^2 \cdot 3)^3$
$\mathbb{Q}(\sqrt{-7})$	1	$-(3 \cdot 5)^3$
$\mathbb{Q}(\sqrt{-163})$	1	$-(2^6 \cdot 3 \cdot 5 \cdot 23 \cdot 29)^3$
$\mathbb{Q}(\sqrt{-5})$	2	$-(2^4 \cdot 5 \cdot 11)^3$
$\mathbb{Q}(\sqrt{-6})$	2	$-(2^4 \cdot 3^2 \cdot 17)^3$
$\mathbb{Q}(\sqrt{-13})$	2	$-(2^4 \cdot 3^2 \cdot 5^2 \cdot 23)^3$
$\mathbb{Q}(\sqrt{-23})$	3	$-(5^3 \cdot 11 \cdot 17)^3$
$\mathbb{Q}(\sqrt{-21})$	4	$-(2^8 \cdot 3^5 \cdot 47 \cdot 59)^3$
$\mathbb{Q}(\sqrt{-133})$	4	$-(2^8 \cdot 3^4 \cdot 5^4 \cdot 11^2 \cdot 23^2 \cdot 29^2 \cdot 383)^3$

Table 3.1: Absolute norms of a few singular moduli

norm to be rational. Second, the primes dividing this norm are quite small, and all norms seem to be perfect cubes. Neither observation is a total coincidence. Table 3.1 supports these claims (Gross and Zagier have tabulated $N(j_K)$ for all known fundamental discriminants of class number 1 or 3—cf. [G–Z, Table 1]).

3.6 Integrality of Singular Moduli

Since the j-invariant of a CM–curve is an algebraic number we know its absolute norm $N(j)$ is rational. However, in all our examples above these norms were in fact integers. This is always the case, and the goal of this section is to explain this phenomenon. It will suffice to show that singular moduli are algebraic integers—the norm of an algebraic integer is always an integer (cf. [Mar, p. 22]). More concretely, we will show that given $\tau \in \mathbb{C}$ such that $\text{Im}(\tau) > 0$ and $[\mathbb{Q}(\tau) : \mathbb{Q}] = 2$ then $j(\tau)$ is an algebraic integer. The proof we give is analytic and has the advantage of explicitly showing a monic polynomial with integer coefficients that $j(\tau)$ satisfies. Our exposition is a hybrid of [Cox], [Se 1], [Sil 2] and [Lang]. There are two other known proofs of the integrality of singular moduli, due to Serre and Tate [S–T] (the l-adic good reduction argument) and Serre [Sil 2, § V.6] (the p-adic bad reduction argument); both arguments go beyond the scope of this paper.

3.6.1 Modular Functions of Weight 0

We denote the group $SL_2(\mathbb{Z})$ of 2×2 matrices with coefficients in $\mathbb{Z}$ and determinant 1 by Γ. The subgroup $\Gamma_0(m)$ of is defined as

$$\Gamma_0(m) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z});\ c \equiv 0 \bmod m \right\}.$$

Let f be a meromorphic function on the upper half plane $\mathbb{H}$. We say that f is $\Gamma_0(m)$-invariant if

$$f(\gamma\tau) := f\left(\frac{a\tau + b}{c\tau + d}\right) = f(\tau) \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(m).$$

Such a function f satisfies $f(\gamma(\tau+m)) = f(\gamma\tau)$ for any $\gamma \in \Gamma$. Indeed, if $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$, then $\tau + m = A\tau$, and since $\gamma A \gamma^{-1}$ is in $\Gamma_0(m)$, we conclude that

$$f(\gamma(\tau+m)) = f(\gamma A\tau) = f(\gamma A\gamma^{-1}\gamma\tau) = f(\gamma\tau).$$

It follows that f has an expansion in the variable $q^{1/m}(\tau) = e^{2\pi i\tau/m}$ in the region $0 < |q^{1/m}| < 1$,

$$f = \sum a_n q^{n/m}.$$

We will call this a "q-expansion" for f. The function f is said to be meromorphic (resp. holomorphic) at infinity if $a_n = 0$ for $n \ll 0$ (resp. $a_n = 0$ for $n < 0$).

Definition 3.1. *A modular function f of weight zero for $\Gamma_0(m)$ is a meromorphic function on $\mathbb{H}$ that is $\Gamma_0(m)$-invariant and is meromorphic at infinity.*

Remark 3.16. If $m = 1$ then $\Gamma_0(1) = SL_2(\mathbb{Z}) = \Gamma$ and the definition above coincides with that of a modular function of weight 0 that appears in the literature [Se 1, Ch. VIII]. Whenever we refer to a modular function of weight zero without qualification, we mean the function is modular for Γ.

Example 3.2. Let $\Lambda = [1, \tau]$ be a lattice in $\mathbb{C}$ (say $\mathrm{Im}(\tau) > 0$). The j-invariant $j(\tau)$ of Λ is modular of weight zero and is holomorphic on $\mathbb{H}$; it has a simple pole at infinity. The q-expansion of $j(\tau)$ has integer coefficients, the first few of which are

$$j = \frac{1}{q} + 744 + 196884q + \cdots$$

(cf. [Se 1, Ch. VIII.2–4]).

Example 3.3. A modular function that is holomorphic everywhere (including infinity) is constant. (cf. [Cox, Lemma 10.11]).

Example 3.4. The function $j(m\tau)$ is modular of weight 0 for $\Gamma_0(m)$. Indeed, $j(m\tau)$ is certainly holomorphic on $\mathbb{H}$. To see $j(m\tau)$ is invariant under $\Gamma_0(m)$, we compute

$$j(m\gamma\tau) = j\left(\frac{m(a\tau+b)}{c\tau+d}\right) = j\left(\frac{am\tau+bm}{c/m\cdot m\tau+d}\right) \quad \text{for } \gamma \in \Gamma_0(m),$$

so setting $\gamma' = \begin{pmatrix} a & bm \\ c/m & d \end{pmatrix} \in \Gamma$ it follows that $j(m\gamma\tau) = j(\gamma' m\tau) = j(m\tau)$, where the last equality is a consequence of the Γ-invariance of j.

To see that $j(m\tau)$ is meromorphic at infinity, we want to compute its q-expansion. For this purpose we introduce the set

$$C(m) = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix};\ ad = m,\ 0 \le b < d, (a,b,d) = 1 \right\}.$$

Let $\sigma_0 = \begin{pmatrix} m & 0 \\ 0 & 1 \end{pmatrix} \in C(m)$. Since $\Gamma\sigma_0\Gamma = \bigcup_{\sigma\in C(m)} \Gamma\sigma$ (cf. [Shi, p. 108]), if we fix $\gamma \in \Gamma$ there are $\sigma \in C(m)$ and $\overline{\gamma} \in \Gamma$ such that $\sigma_0^{-1}\overline{\gamma}\sigma = \gamma$. It follows that

$$j(m\gamma\tau) = j(\sigma_0\gamma\tau) = j(\overline{\gamma}\sigma\tau) = j(\sigma\tau),$$

where the last equality holds because j is Γ-invariant. Let

$$j(\tau) = \frac{1}{q} + \sum_{n=0}^{\infty} a_n q^n$$

be the q-expansion of j. If $\sigma\tau = (a\tau + b)/d$ then $q(\sigma\tau) = e^{2\pi i b/d} q^{a/d} = \zeta_m^{ab} (q^{1/m})^{a^2}$, where $\zeta_m = e^{2\pi i/m}$. Hence

$$j(m\gamma\tau) = j(\sigma\tau) = \frac{\zeta_m^{-ab}}{(q^{1/m})^{a^2}} + \sum_{n=0}^{\infty} a_n \zeta_m^{abn} (q^{1/m})^{a^2 n}. \tag{3.6}$$

This shows that there are only finitely many terms of negative order in the q-expansion of $j(m\gamma\tau)$, which is to say that $j(m\gamma\tau)$ is meromorphic at infinity. Thus it is a modular function of weight 0 for $\Gamma_0(m)$, as claimed.

There is a sense in which $j(\tau)$ is the only modular function of weight 0 that is holomorphic on $\mathbb{H}$. The following lemma makes this notion precise.

Lemma 3.17 (Hasse q-expansion principle). *Every modular function $f(\tau)$ of weight 0 that is holomorphic on $\mathbb{H}$ is a polynomial in $j(\tau)$ over the $\mathbb{Z}$-module generated by the coefficients $a_{-N}, \ldots, a_0$ of the q-expansion $\sum_{-N}^{\infty} a_n q^n$ of f.*

Proof. Since f is meromorphic at infinity, we may write

$$f = \frac{a_{-N}}{q^N} + (\text{higher order terms}).$$

This means $f - a_{-N} j^N$ is holomorphic on the upper half plane and that its q-expansion has at worst a polar term of order $N - 1$. By repeating this process we construct a polynomial $P(X) \in \mathbb{Z}[a_{-N}, \ldots, a_0][X]$ such that $f - P(j(\tau))$ is a holomorphic modular function with a q-expansion that contains only terms of positive order, i.e., $f - P(j(\tau))$ vanishes at infinity. This difference is therefore constant (see Example 3.3). □

3.6.2 The Modular Equation

We will now construct a monic polynomial that $j(\tau)$ satisfies. Consider the polynomial

$$\prod_{\sigma \in C(m)} (X - j(\sigma\tau)) = \sum_{i=0}^{N} s_i X^i,$$

where the s_i are elementary symmetric functions on the $j(\sigma\tau)$ and thus are holomorphic functions on $\mathbb{H}$. Recall that $\Gamma\sigma_0\Gamma = \bigcup_{\sigma \in C(m)} \Gamma\sigma$ and so

$$\bigcup_{\sigma \in C(m)} \Gamma\sigma = \bigcup_{\sigma \in C(m)} \Gamma\sigma\gamma \quad \text{for any } \gamma \in \Gamma.$$

Hence, for a fixed σ and γ there are elements $\overline{\gamma} \in \Gamma$ and $\overline{\sigma} \in C(m)$ such that $\sigma\gamma = \overline{\gamma}\,\overline{\sigma}$. Since $j(\sigma\gamma\tau) = j(\overline{\gamma}\,\overline{\sigma}\tau) = j(\overline{\sigma}\tau)$, we have

$$\prod_{\sigma \in C(m)} (X - j(\sigma\tau)) = \prod_{\sigma \in C(m)} (X - j(\sigma\gamma\tau)) \quad \text{for any } \gamma \in \Gamma,$$

It follows that the s_i are invariant under Γ. The expansion (3.6) then shows that an elementary symmetric function on the $j(\sigma\tau)$'s has finitely many polar terms, and hence the functions s_i are holomorphic modular functions of weight 0. By Lemma 3.17 they are polynomials in $j(\tau)$. We conclude from all this that there exists $\Phi_m(X,Y) \in \mathbb{C}[X,Y]$ such that

$$\Phi_m(X, j(\tau)) = \prod_{\sigma \in C(m)} (X - j(\sigma\tau)). \tag{3.7}$$

The polynomial $\Phi_m(X,Y)$ is referred to (by abuse of language) as the *modular equation.*

The modular equation has some wonderful algebraic properties.

Theorem 3.18. *The Modular Equation has integer coefficients, i.e., $\Phi(X,Y) \in \mathbb{Z}[X,Y]$.*

Proof. By (3.7) it is enough to show that an elementary symmetric function $s(\tau)$ on the $j(\sigma\tau)$ is in fact a polynomial in $j(\tau)$ with integer coefficients.

The q-expansion of $j(\sigma\tau)$ given in (3.6) shows that $j(\sigma\tau)$ is in the field $\mathbb{Q}(\zeta_m)((q^{1/m}))$ of meromorphic functions in $q^{1/m}$ over $\mathbb{Q}(\zeta_m)$. We can do slightly better than this: $j(\sigma\tau)$ is contained in $\mathbb{Q}((q^{1/m}))$. To see why first note that any automorphism $\psi \in \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}))$ induces an automorphism of $\mathbb{Q}(\zeta_m)((q^{1/m}))$ by acting on the coefficients of an element in $\mathbb{Q}(\zeta_m)((q^{1/m}))$. The automorphism is determined by the image of ζ_m. Say that $\psi(\zeta_m) = \zeta_m^k$ for an integer k relatively prime to m. Then using (3.6) we see that

$$\psi(j(\sigma\tau)) = \frac{\zeta_m^{-abk}}{(q^{1/m})^{a^2}} + \sum_{n=0}^{\infty} a_n \zeta_m^{abkn} (q^{1/m})^{a^2 n}.$$

Let b' be an integer such that $b' \equiv bk \bmod d$ and $0 \leq b' < d$. Then $\zeta_m^{abk} = \zeta_m^{ab'+adt} = \zeta_m^{ab'}$ because $ad = m$. Hence

$$\psi(j(\sigma\tau)) = \frac{\zeta_m^{-ab'}}{(q^{1/m})^{a^2}} + \sum_{n=0}^{\infty} a_n \zeta_m^{ab'n} (q^{1/m})^{a^2 n}.$$

Setting $\sigma' = \begin{pmatrix} a & b' \\ 0 & d \end{pmatrix} \in C(m)$ we conclude that

$$\psi(j(\sigma\tau)) = j(\sigma'\tau).$$

This means the elements of $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}))$ permute the $j(\sigma\tau)$'s so that any symmetric function on them is contained in the fixed field $\mathbb{Q}((q^{1/m}))$, as claimed. Furthermore, $s(\tau+1) = s(\tau)$ since $s(\tau)$ is Γ-invariant, whence the q-expansion of $s(\tau)$ is in $\mathbb{Q}((q))$.

We can do even better and show that $s(\tau) \in \mathbb{Z}((q))$. Indeed, (3.6) shows the coefficients of the q-expansion of $j(\sigma\tau)$ are algebraic integers for all $\sigma \in C(m)$, and therefore so are the coefficients of $s(\tau)$, which means $s(\tau) \in (\mathbb{Q} \cap \mathbb{Z})((q)) = \mathbb{Z}((q))$.

Finally, Lemma 3.17 tells us $s(\tau)$ is a polynomial in $j(\tau)$ with coefficients in the $\mathbb{Z}$-module generated by the coefficients of the q-expansion of $s(\tau)$. But since $s(\tau)$ is in $\mathbb{Z}((q))$ this $\mathbb{Z}$-module is just $\mathbb{Z}$ itself, so that $s \in \mathbb{Z}[j]$, as claimed. □

Theorem 3.19. *If m is not a square, then $\Phi_m(X,X)$ is a polynomial of degree at least 1 with leading coefficient equal to ± 1.*

Proof. Replacing $\Phi_m(X, X)$ with $\Phi_m(j(\tau), j(\tau))$, it suffices to look at the coefficient of the highest negative power of q in the q-expansion of $\Phi_m(j(\tau), j(\tau)) = \prod_{\sigma \in C(m)}(j(\tau) - j(\sigma\tau))$. Let $\sigma = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in C(m)$. Using the q-expansion (3.6) we see that

$$j(\tau) - j(\sigma\tau) = \frac{1}{q} - \frac{\zeta_m^{-ab}}{q^{a/d}} + \sum_{n=0}^{\infty} c'_n (q^{1/m})^n,$$

for some coefficients c'_n. Since m is not a perfect square, we know that $a \neq d$ because $ad = m$. Hence $a/d \neq 1$, which means that the coefficient of the highest negative power of q is a root of unity. It follows that the coefficient of highest negative power of the q-expansion for $\Phi_m(j(\tau), j(\tau))$ is also a root of unity. However, this coefficient is also an integer because the coefficients of $\Phi_m(X, j(\tau))$ have q-expansions in $\mathbb{Z}((q))$. Hence the coefficient must be ± 1. □

3.6.3 Integrality of j

It is clear from (3.7) that

$$\Phi_m(j(\sigma\tau), j(\tau)) = 0 \quad \text{for all } \sigma \in C(m). \tag{3.8}$$

We need to extend the class of matrices σ for which this equality holds in order to show that $j(\tau)$ is an algebraic integer for imaginary quadratic $\tau \in \mathbb{H}$. We claim that

$$\Phi_m(j(\alpha\tau), j(\tau)) = 0 \quad \text{for all primitive } \alpha \in D_m, \tag{3.9}$$

where

$$D_m = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix};\ a, b, c, d \in \mathbb{Z}, ad - bc = m \right\}.$$

A matrix α is said to be primitive if its entries are relatively prime integers.

It suffices to show that for any primitive matrix $\alpha \in D_m$ there is some $\gamma \in \Gamma$ such that $\gamma\alpha \in C(m)$, because then the Γ-invariance of j and (3.8) imply that

$$0 = \Phi_m(j(\gamma\alpha\tau), j(\tau)) = \Phi_m(j(\alpha\tau), j(\tau)). \tag{3.10}$$

Let $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in D_m$. First, we may bring α to upper triangular form. Indeed, let w, z be relatively prime integers such that $az + cw = 0$. Choose integers x, y such that $xw - yz = 1$. Then

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix}$$

Moreover, the matrix $\begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix}$ is also in D_m by construction. Next, note that

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} = \begin{pmatrix} a' & b' + kd' \\ 0 & d' \end{pmatrix}.$$

If we choose k_0 such that $0 \leq b' + k_0 d' < d'$ and set

$$\gamma = \begin{pmatrix} 1 & k_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix},$$

then we obtain $\gamma\alpha \in C(m)$ for $\gamma \in \Gamma$, as desired.

Theorem 3.20. *Let $\tau \in \mathbb{H}$ be such that $[\mathbb{Q}(\tau) : \mathbb{Q}] = 2$ then $j(\tau)$ is an algebraic integer.*

Proof. Set $K = \mathbb{Q}(\tau)$ and let $\mathcal{O}_K$ be the ring of integers of K. The ring $\mathcal{O}_K$ has a $\mathbb{Z}$-basis given by $[1, w_K]$, where $w_K = (d_K + \sqrt{d_K})/2$ and d_K is the discriminant of K (cf. [F–T, II.1.33]). There always exists $\lambda \in \mathcal{O}_K$ such that $N_{\mathbb{Q}}^K(\lambda)$ square-free. Indeed,

- if $K = \mathbb{Q}(i)$, take $\lambda = 1 + i$.
- Otherwise, $K = \mathbb{Q}(\sqrt{-m})$ for a square-free $m > 1$. Take $\lambda = \sqrt{-m}$ in this case.

Since $[1, w_K]$ is a $\mathbb{Z}$-basis for $\mathcal{O}_K$, there are relatively prime integers a, b, c, d such that

$$\begin{aligned} \lambda w_K &= a w_k + b, \\ \lambda &= c w_k + d, \end{aligned}$$

and $N_{\mathbb{Q}}^K(\lambda) = ad - bc =: n$. Let $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so that $w_K = \alpha w_K$ and by (3.9)

$$0 = \Phi_n(j(\alpha w_K), j(w_K)) = \Phi_n(j(w_K), j(w_K)).$$

Since n is square-free, Theorem 3.19 tells us the above is an integrality relation for $j(w_K)$. Hence $j(w_K)$ is an algebraic integer. We want to deduce from this that $j(\tau)$ is also an algebraic integer.

Since $\mathbb{Q}(w_K) = \mathbb{Q}(\tau)$ there exists a primitive 2×2 matrix β such that $\tau = \beta w_K$. If we can show that $j(\tau) = j(\beta w_K)$ is integral over $\mathbb{Z}[j(w_K)]$ then, by transitivity of integrality, $j(\tau)$ will be an algebraic integer. To see that $j(\beta w_K)$ is integral over $\mathbb{Z}[j(w_K)]$, note that by (3.9)

$$\Phi_m(j(\beta w_K), j(w_K)) = 0,$$

where $m = \det \beta$, so $j(\beta w_K)$ is a root of the monic polynomial $\Phi_m(X, j(w_K)) \in \mathbb{Z}[X, j(w_K)]$. □

Corollary 3.21. *The absolute norm $N(j_K)$ of a j-invariant corresponding to a curve in $\mathfrak{E}(\mathcal{O}_K)$ is an integer.* □

Remark 3.22. The reader may wonder where in our proof of the integrality of $j(\tau)$ we used the hypothesis that the associated elliptic curve E_Λ, $\Lambda = [1, \tau]$ is a CM-curve. By Theorem 2.11, if the curve E_Λ has complex multiplications, τ generates an imaginary quadratic field over $\mathbb{Q}$, and this is precisely the hypothesis for τ in Theorem 3.20. Ordinary elliptic curves may have transcendental j-invariant over $\mathbb{Q}$. Moreover, if the j-invariant of an elliptic curve E/K (K a number field) is not in $\mathcal{O}_K$ then $\operatorname{End} E = \mathbb{Z}$ (this is the claim Serre proved in his p-adic bad reduction argument, cf. [Sil 2, Theorem V.6.3]).

3.7 Gross–Zagier Numbers

Let K be an imaginary quadratic field. Recall $\mathfrak{E}(\mathcal{O}_K)$ is the set of elliptic curves with complex multiplication by $\mathcal{O}_K$ up to $\overline{\mathbb{Q}}$-isomorphism. We have seen that if $\{E_1, \ldots, E_h\}$ is a set of representatives for $\mathfrak{E}(\mathcal{O}_K)$ then the absolute norm of a j-invariant of $\mathfrak{E}(\mathcal{O}_K)$,

$$N(j_K) = \prod_{i=1}^{h} j(E_i)$$

is an integer (cf. Corollaries 3.14 and 3.21). We define more generally the absolute norm of the difference between two singular moduli as follows. Let K and K' be quadratic imaginary fields, and let $\{E_1, \ldots, E_{h_1}\}$ and $\{E'_1, \ldots, E'_{h'}\}$ be sets of representatives for $\mathfrak{E}(\mathcal{O}_{K_1})$ and $\mathfrak{E}(\mathcal{O}_{K_2})$, respectively. Then the norm $N(j_K - j_{K'})$ is simply

$$N(j_K - j_{K'}) = \prod_{m=1}^{h} \prod_{n=1}^{h'} \left(j(E_i) - j(E'_j)\right).$$

A number of this kind is called a *Gross–Zagier* number. Equivalently, using the isomorphism between lattices and elliptic curves over $\mathbb{C}$, we may write

$$N(j_K - j'_K) = \prod_{\bar{\mathfrak{a}}} \prod_{\bar{\mathfrak{b}}} \left(j(\mathfrak{a}) - j(\mathfrak{b})\right),$$

as $\bar{\mathfrak{a}}$ and $\bar{\mathfrak{b}}$ run through the ideal classes of $C(\mathcal{O}_K)$ and $C(\mathcal{O}_{K'})$, respectively.

Theorem 3.23. *Let K and K' be quadratic imaginary fields, and let E, E' be elliptic curves in $\mathfrak{E}(\mathcal{O}_K)$ and $\mathfrak{E}(\mathcal{O}'_K)$, respectively. Then the norm $N(j_k - j_{K'})$ is an integer.*

Proof. Since singular moduli are algebraic integers and $N(j_K - j'_K)$ is by definition the norm of any $j(E_i) - j(E'_j)$ with respect to the group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ it follows that $N(j_K - j'_K)$ is always an integer. □

Remark 3.24. Our previous definition of $N(j_K)$ is a special case of this more general definition when $K' = \mathbb{Q}(\sqrt{-3})$, for in this case $C(\mathcal{O}_{K'})$ consists of one element and $j((1 + \sqrt{-3})/2) = 0$.

The aim of this paper is to study the size of prime factors that divide Gross–Zagier numbers. The basic idea is that in order for p to divide $N(j_K - j_{K'})$, the curves E and E' must "fit together" modulo this prime. Since each of these curves has complex multiplication by different rings of integers of imaginary quadratic fields, the only way one can fit the curves together is if the reduction of E and E' modulo p is supersingular, i.e., the endomorphism ring of the reduced curve must be an order in a rational quaternion algebra. This order will (hopefully) have enough room to fit both $\mathcal{O}_K$ and $\mathcal{O}_{K'}$ inside it. Our next task then is to study supersingular curves, but before we can do so, we must understand rational quaternion algebras in depth.

Chapter 4

Rational Quaternion Algebras

In Chapter 2 we saw the endomorphism ring of an elliptic curve can sometimes be an order in a rational quaternion algebra. In the present chapter, we will study these rings in the abstract, outside the context of elliptic curves. This way we will build up an arsenal of theorems that will aid our discussion of supersingular curves in Chapters 5 and 6.

Our first task will be to study quaternion algebras over a field K with $\operatorname{char} K \neq 2$; later we will specialize to the case $K = \mathbb{Q}$ and classify rational quaternion algebras by looking at their behavior under tensor product with the real numbers and p-adic fields. Our approach will require extensive use of Hilbert symbols; these will tremendously facilitate our future computations. The reader should refer to [Lam, KKS, Vig] for further details.

4.1 Quaternion Algebras

Unless otherwise stated, we will assume all our fields have characteristic different from 2.

Definition 4.1. *A quaternion algebra $H = (a,b)_K$ over K is a 4-dimensional K-algebra whose generators $\{1,i,j,ij\}$ are subject to the relations*

$$i^2 = a, \quad j^2 = b, \quad ij = -ji,$$

for some $a, b \in K^*$.

Let $h = x + iy + jz + ijw$ be an element of H. We call $\bar{h} = x - iy - jz - ijw$ the conjugate of h, and define the *reduced trace* and the *reduced norm* of h, respectively, as

$$t(h) = h + \bar{h} = 2x, \quad n(h) = h\bar{h} = x^2 - ay^2 - bz^2 + abw^2.$$

The reduced norm is a quadratic form over the underlying K-vector space of H. It is multiplicative over H and an element in our quaternion algebra is invertible if and only if it has nonzero norm.

Example 4.1. The set $M_2(K)$ of 2×2 matrices with entries in K is a quaternion algebra over K. To see this, identify elements of K with their image in $M_2(K)$ by the homomorphism that sends $1 \in K$ to the identity matrix and set

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = i, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = j.$$

This identification is in fact a K-algebra isomorphism between $(1,-1)_K$ and $M_2(K)$. The reduced trace and norm of an element $h \in M_2(K)$ are simply the trace and determinant of the matrix, respectively.

4.1.1 Quadratic Spaces

Given a quaternion algebra H over K we may use its associated reduced trace to produce a quadratic space (V, B) on the underlying (finite dimensional) K-vector space V of H, whose symmetric bilinear form B is given by

$$B(x,y) = \frac{1}{2}(x\bar{y} + \bar{x}y) = \frac{1}{2}T(x\bar{y}), \quad x, y \in V.$$

The quadratic form $q : V \to \mathbb{R}$ associated to B is the reduced norm of H (hence the sometimes used terminology 'norm form'):

$$q(x) = B(x,x) = \frac{1}{2}t(x\bar{x}) = \frac{1}{2}t(n(x)) = n(x).$$

As is often the case in number theory, the geometric interpretation of an algebraic object, in this case a quadratic space, provides much insight into the structure and properties of the algebraic object at hand. It is therefore advantageous for us to pause briefly in our study of quaternion algebras and look closely at quadratic spaces. We recall the basic definitions and properties of these spaces. The reader interested in a more detailed study would do well to consult [Lam].

Definition 4.2. *A quadratic space is a pair (V, B), where V is a finite dimensional vector space over a field K and B is a symmetric bilinear form. The bilinear form B induces a quadratic map $q : V \to \mathbb{R}$ given by $x \mapsto B(x,x)$.*

Remark 4.1. Given a quadratic space V and its associated quadratic form q we may recover the symmetric bilinear form B through the usual polarization formula

$$B(x,y) = \frac{1}{2}(q(x+y) - q(x) - q(y)).$$

A quadratic space can therefore be specified by either B or q. Sometimes we will call V a quadratic space without specifying either B or q; it will be implicitly understood V possesses these. Conversely, if the space V is understood, we may refer to q itself as a quadratic space.

The map q is a quadratic form and consequently has a unique symmetric matrix A such that

$$q(x) = x^t \cdot A \cdot x \quad x \in V.$$

Given a basis $\{v_1, \ldots, v_n\}$ of V we may write this matrix explicitly as $A_{ij} = B(v_i, v_j)$, where A_{ij} is the (i,j)-th entry of A.

Recall that two quadratic forms Q and Q' are said to be K-*equivalent* if there exists an invertible change of basis $M \in GL(V)$ and the associated matrices of these forms are A and M^tAM, respectively. On the other hand, two quadratic spaces (V, B) and (V', B') are said to be *isometric* if there is an invertible linear map $f : V \to V'$ such that

$$B'(f(x), f(y)) = B(x,y) \quad \text{for all } x, y \in V.$$

In this way we get the one-to-one correspondence

$$\left\{\begin{matrix}\text{classes of } n\text{-ary}\\ \text{quadratic forms}\end{matrix}\right\} \longleftrightarrow \left\{\begin{matrix}\text{isometry classes of}\\ \text{quadratic spaces}\end{matrix}\right\}.$$

The *determinant* of a quadratic form q is just the determinant of the associated matrix A. It is an invariant of the equivalence class of q up to squares in K^* because

$$\det A' = \det A \cdot \det^2 M.$$

We denote this invariant by $d(V) = d(q) = \det A \cdot K^{*2}$.

It is possible to construct new quadratic spaces from old ones. For example, we may add two quadratic spaces (V, B) and (V', B') in the following sense. Set

$$V'' = V \oplus V',$$
$$B''((x, y), (x', y')) = B(x, y) + B'(x', y').$$

The pair (V'', B'') is a quadratic space called the *orthogonal sum* of (V, B) and (V', B'). It is usually denoted $V \perp V'$. The quadratic form of $V \perp V'$ is the sum of q and q' and is denoted $q \perp q'$. The following theorem of Witt makes precise the sense in which the above sum of quadratic spaces is orthogonal (cf. [Lam, Theorem I.4.2], [Se 1, § IV.1.5]).

Theorem 4.2 (Witt cancellation). *Let V, V' and V'' be three quadratic spaces such that $V \perp V' \cong V \perp V''$. Then $V' \cong V''$.* □

Remark 4.3. Every quadratic form over a field of characteristic different from 2 may be transformed into an equivalent diagonal form $\lambda_1 x_1^2 + \cdots \lambda_n x_n^2$. In other words, the matrix of the quadratic form may be diagonalized by a similarity transformation into a matrix whose diagonal consists of the numbers $\{d_1, \ldots, d_n\}$. We denote such a form by $\langle d_1, \ldots, d_n\rangle$. Note that $\langle d_1\rangle \perp \langle d_2\rangle \cong \langle d_1, d_2\rangle$.

We say a quadratic space (V, B) is *regular* if for a given $x \in V$

$$B(x, y) = 0 \text{ for all } y \in V \implies x = 0.$$

This is equivalent to the non-singularity of the matrix A of the associated quadratic form q.

Another important notion in our study of quadratic spaces is that of *isotropy.* A quadratic space (V, B) is called isotropic if there exists a nonzero vector $x \in V$ such that $B(x, x) = q(x) = 0$. A special but rather important kind of quadratic space is a two dimensional, regular, isotropic space. Such a space is called a *hyperbolic plane.* The following theorem gives three equivalent ways to think about such spaces (cf. [Lam, Theorem I.3.2]).

Theorem 4.4. *Let (V, B) be a two dimensional quadratic space over K and let q be its associated quadratic form. Then the following statements are equivalent:*

(i) V is a hyperbolic plane,

(ii) V is regular and $d(q) = -1 \cdot (K^*)^2$.

(iii) V is isometric to $\langle 1, -1\rangle$. □

Hyperbolic planes will be of great use in our study of quadratic spaces; however, most spaces we are interested in are unfortunately not 2-dimensional. We need a criterion that can tell us when a given space contains a hyperbolic plane. The following theorem provides us that criterion.

Theorem 4.5. *Let (V, B) be a regular quadratic space which is at least 2-dimensional over K. Then V contains a hyperbolic plane if and only if it is isotropic.*

Proof. Recall V is isotropic if and only if there is a nonzero vector $x \in V$ such that $B(x,x) = 0$. Take V' to be the 1-dimensional subspace of V spanned by x and let $y \in V$ be an element which is not orthogonal to x. The vectors x and y are linearly independent over K; otherwise there is a t in K^* such that $x = ty$. But then

$$0 = B(x,x) = tB(x,y)$$

which contradicts the assumption that x and y are not orthogonal. Let $P = Kx + Ky$. Then

$$d(P) = \begin{vmatrix} B(x,x) & B(x,y) \\ B(y,x) & B(y,y) \end{vmatrix} \cdot (K^*)^2 = -1 \cdot (K^*)^2.$$

Regularity of V is inherited by the subspace P, and so Theorem 4.4 tells us P is a hyperbolic plane. This completes the proof. □

Remark 4.6. The regularity of the subspace P means V can be decomposed as $P \perp P^{\perp}$, which is to say that V contains the hyperbolic plane P as an orthogonal summand (cf. [Lam, Corollary I.2.5]).

4.2 Quaternion Algebras and Quadratic Spaces

Through our knowledge of quadratic spaces we are now in a position to prove that a quaternion algebra $(a,b)_K$ is isomorphic to $M_2(K)$ when the binary quadratic form $ax^2 + by^2$ represents 1. A detailed study of this equivalent question will in turn yield the classification of quaternion algebras we seek.

We observe that the quaternion algebra $(a,b)_K$ with orthogonal basis $\{1, i, j, ij\}$ is a regular quadratic space and is isometric to $\langle 1, -a, -b, ab\rangle$; the isometry is given by the reduced norm map.

Theorem 4.7. *Let $H = (a,b)_K$ be a quaternion algebra. Then the following are equivalent:*

(i) $H \cong (1,-1)_K$ $(\cong M_2(K))$,

(ii) H is not a division algebra,

(iii) The binary quadratic form $ax^2 + by^2$ represents 1.

Proof. (i $\Longrightarrow$ ii) This is clear since $M_2(K)$ has zero divisors.

(ii $\Longrightarrow$ i) We first show that H is a simple K-algebra. Indeed, let $\overline{K}$ be an algebraic closure of K. Then $(a,b)_{\overline{K}} = H \otimes \overline{K}$. Every element of $\overline{K}$ is a square, so $(a,b)_{\overline{K}} \cong (1,-1)_{\overline{K}} \cong M_2(\overline{K})$ (cf. Example 4.1). Since $M_2(\overline{K})$ is a simple $\overline{K}$ algebra it follows that H is a simple K-algebra.

By Wedderburn's Theorem, $H \cong M_n(D)$ where D is a skew field containing K. If H is not a division algebra, we must have $n \geq 2$, but H is a 4-dimensional K-vector space, so $n = 2$ and $D = K$.

(ii $\Longrightarrow$ iii) To say H is not a division algebra means there is a nonzero element $x \in H$ which is not invertible. We've seen this is the case if and only if $N(x) = B(x,x) = 0$, which means x is an isotropic element. By Theorem 4.5 and Remark 4.6 it follows that $H = P \perp P^{\perp}$, where P is a hyperbolic plane. Hence, computing determinants we see that

$$d(H) = d(P) \cdot d(P^{\perp}) = -d(P^{\perp}).$$

On the other hand, since $H \cong \langle 1, -a, -b, ab \rangle$, we may also compute

$$d(H) = 1 \cdot (-a) \cdot (-b) \cdot (ab) \cdot (K^*)^2 = 1 \cdot (K^*)^2.$$

Hence $d(P^{\perp}) = -1 \cdot (K^*)^2$. Since $P^{\perp}$ is a regular subspace of H, Theorem 4.4 tells us $P^{\perp}$ is itself a hyperbolic plane. Therefore

$$\langle 1, -a, -b, ab \rangle \cong H \cong \langle 1, -1, 1, -1 \rangle.$$

By Witt's cancellation theorem $\langle -a, -b, ab \rangle \cong \langle -1, 1, -1 \rangle$. Adding $\langle a, b, 1 \rangle$ orthogonally to both sides of this equation we obtain

$$\begin{gathered}\langle a, -a, b, -b, 1, ab \rangle \cong \langle a, b, 1, -1, 1, -1 \rangle \\ 2P \perp \langle 1, ab \rangle \cong 2P \perp \langle a, b \rangle,\end{gathered}$$

where this last equality follows from the isomorphisms $\langle a, -a \rangle \cong \langle b, -b \rangle \cong \langle 1, -1 \rangle \cong P$. Applying Witt's cancellation theorem again we obtain $\langle 1, ab \rangle \cong \langle a, b \rangle$. Hence the form $ax^2 + by^2$ represents 1.

(iii $\Longrightarrow$ ii) Suppose there are $x, y \in K^*$ with $ax^2 + by^2 = 1$. Let $v = 1 + xi + yj$; then $N(v) = 1 - ax^2 - by^2 = 0$. Since $v \neq 0$ this means v is not invertible in H. Hence H is not a division algebra. □

4.2.1 Hilbert Symbols and Quadratic Forms over $\mathbb{Q}_p$

Given a, b in a field K, Theorem 4.7 tells us that if we can settle the question of when the quadratic form $ax^2 + by^2$ represents 1 for some $x, y \in K$ then we'll have a better understanding of the structure of the quaternion algebra $(a,b)_K$. We will settle this question for the p-adic fields $\mathbb{Q}_p$, where p is a rational prime or the 'prime at infinity,' in which case we will agree that $\mathbb{Q}_\infty = \mathbb{R}$. The Hasse–Minkowski principle will then yield information about the case $K = \mathbb{Q}$. The relevant background material on p-adic numbers can be found in [Se 1, Ch. 2]. We follow the ideas in [KKS, Ch. 2], and use some proofs presented there in our exposition.

Over the field of real numbers it is clear the quadratic form $ax^2 + by^2$ represents 1 if and only if at least one of a or b is positive. We encode this information in the symbol $(a,b)_\infty$ defined as

$$(a,b)_\infty = \begin{cases} 1 & \text{if } a \text{ or } b\ > 0, \\ -1 & \text{otherwise.} \end{cases}$$

Then, for $a, b \in \mathbb{Q}^*$

$$ax^2 + by^2 \text{ represents 1 for some } x, y \in \mathbb{R} \iff (a,b)_\infty = 1. \tag{4.1}$$

Similarly, we would like to define a symbol $(a,b)_p$ that satisfies an analogous property to (4.1) for $x, y \in Q_p$.

Definition 4.3. *Let $a, b \in \mathbb{Q}^*$ and let p be a prime number. Write*

$$a = p^i u, \quad b = p^j v \qquad i, j \in \mathbb{Z}, \quad u, v \in \mathbb{Z}_{(p)}^*,$$

where $\mathbb{Z}_{(p)}$ is the localization of $\mathbb{Z}$ at p. If $p \neq 2$ we define the rational Hilbert symbol $(a, b)_p$ *by*

$$(a, b)_p = (-1)^{ij(p-1)/2} \left(\frac{\bar{u}}{p}\right)^j \left(\frac{\bar{v}}{p}\right)^i,$$

where $\left(\frac{\bar{u}}{p}\right)$ denotes the usual Legendre symbol and $\bar{u}$ is the image of u under the homomorphism of reduction modulo p. Otherwise set

$$(a, b)_2 = (-1)^{\frac{r^2-1}{8}} \cdot (-1)^{\frac{u-1}{2} \cdot \frac{v-1}{2}},$$

where $r = (-1)^{ij} u^j v^{-i}$.

Remark 4.8. The rational Hilbert symbol can be extended naturally to a map $(,)_p : \mathbb{Q}_p \times \mathbb{Q}_p \to \{\pm 1\}$. All we have to do is replace $\mathbb{Z}_{(p)}$ by $\mathbb{Z}_p$ in the above definition.

Before proving the above definition has our desired property, we give some attributes of the Hilbert symbol that follow easily from the above definition (and which hold for the extended symbol as well cf. [KKS, Proposition 2.4]).

Theorem 4.9. *Let v be a rational prime or ∞, and let $a, b \in \mathbb{Q}^*$. Then*

(i) $(a, b)_v = (b, a)_v$.

(ii) $(-a, a)_v = 1$. *If $a \neq 1$ then* $(a, 1-a)_v = 1$.

(iii) $(a, bc)_v = (a, b)_v (a, c)_v$.

(iv) If $a, b \in (\mathbb{Z}_{(p)})^*$ and p is odd then

$$(a, b)_p = 1 \text{ and } (a, bp)_p = \left(\frac{\bar{a}}{p}\right).$$

(v) If $a, b \in (\mathbb{Z}_{(2)})^*$ then we have

$$(a, b)_2 = \begin{cases} 1 & \text{if } a \text{ or } b \text{ are congruent to } 1 \bmod 4, \\ -1 & \text{otherwise;} \end{cases}$$

$$(a, 2b)_2 = \begin{cases} 1 & \text{if } a \cong 1 \text{ or } 1 - 2b \bmod 8, \\ -1 & \text{otherwise.} \end{cases}$$

Lemma 4.10 (Squares in $\mathbb{Q}_p$). *Let $x \in \mathbb{Q}_p$ and write $x = p^i u$ with $i \in \mathbb{Z}$ and $u \in \mathbb{Z}_p^*$. Then x is the square of an element in $\mathbb{Q}_p$ if and only if i is even and* *(i) If $p \neq 2$, then $\left(\frac{\bar{u}}{p}\right) = 1$,*

(ii) if $p = 2$, then $u \equiv 1 \bmod 8\mathbb{Z}_2$. □

(cf. [KKS, Proposition 2.18])

Theorem 4.11. *Let $a, b \in \mathbb{Q}_p$. Then the quadratic form $ax^2 + by^2$ represents 1 for some $x, y \in \mathbb{Q}_p$ if and only if $(a, b)_p = 1$.*

Proof. Suppose the quadratic form $ax^2 + by^2$ represents 1 in $\mathbb{Q}_p$. If $x = 0$ then b is a square in $\mathbb{Q}_p^*$ in which case it is clear that $(a, b)_p = 1$; similarly if $y = 0$. If both x and y are nonzero then

$$(a, b)_p = (ax^2, by^2)_p = (ax^2, 1 - ax^2)_p = 1,$$

where the last equality is a consequence of Theorem 4.9 (ii).

Now suppose $(a, b)_p = 1$. The conditions $(a, b)_p = 1$ and the form $ax^2 + by^2$ representing 1 in $\mathbb{Q}_p$ are unchanged if we multiply a or b by elements of $(\mathbb{Q}_p^*)^2$. We may therefore assume without loss of generality that $a, b \in Z_p \cup p\mathbb{Z}_p$.

- $a, b \in p\mathbb{Z}_p$. In this case we replace a with $-ab^{-1}$ because on the one hand,

$$(-ab^{-1}, b)_p = (-ab^{-1}, b)_p \cdot (-b, b)_p = (a, b)_p,$$

and on the other hand,

$$\begin{aligned} -ab^{-1}x^2 + by^2 = 1 &\quad \text{has a solution in } \mathbb{Q}_p \\ \text{if and only if } -ab^{-1}x^2 + by^2 = z^2 &\quad \text{has a solution in } \mathbb{Q}_p \text{ with } (x, y, z) \neq (0, 0, 0) \\ \text{if and only if } ax^2 + bz^2 = (by)^2 &\quad \text{has a solution in } \mathbb{Q}_p \text{ with } (x, y, z) \neq (0, 0, 0) \\ \text{if and only if } ax^2 + by^2 = 1 &\quad \text{has a solution in } \mathbb{Q}_p. \end{aligned}$$

Since $-ab^{-1} \in \mathbb{Z}_p$ we are reduced to the case $a \in \mathbb{Z}_p$, $b \in p\mathbb{Z}_p$.

- $a \in \mathbb{Z}_p$, $b \in p\mathbb{Z}_p$. Suppose first that $p \neq 2$. Then by Theorem 4.9,(iv) $(a, b)_p = 1$ means the image of a under reduction mod$p$ is a square in $\mathbb{F}_p^*$. By Lemma 4.10 there is a $t \in \mathbb{Q}_p^*$ with $t^2 = a$. Then $a(1/t)^2 + b \cdot 0^2 = 1$. The case $p = 2$ is similar.

- $a, b \in \mathbb{Z}_p$. If $p \neq 2$, then $(a, b)_p = 1$ is always true (Theorem 4.9,(iv)). We use a counting argument to show $ax^2 = by^2 = 1$ has a solution in $\mathbb{Q}_p$. Let $\bar{a}, \bar{b}$ be the reduction of $a, b \bmod p$, respectively. The sets

$$\{\bar{a}u^2 \mid u \in F_p\} \quad \text{and} \quad \{1 - \bar{b}u^2 \mid u \in F_p\}$$

both have cardinality $(p + 1)/2$ and hence must overlap. So there exist $x, y \in \mathbb{Z}_p$ such that $ax^2 \equiv 1 - by^2 \bmod p\mathbb{Z}_p$. If $x \notin p\mathbb{Z}_p$ then by Lemma 4.10 there is a $t \in \mathbb{Q}_p^*$ with $t^2 = (1 - by^2)/a$; in this case $at^2 + by^2 = 1$. Otherwise we obtain $1 \equiv by^2 \bmod p\mathbb{Z}_p$ and this time Lemma 4.10 asserts there is a $t \in \mathbb{Q}_p^*$ with $t^2 = b$; hence $a \cdot 0^2 + b(1/t)^2 = 1$. The case $p = 2$ is solved using Theorem 4.9 and Lemma 4.10. □

With the aid of Hilbert symbols, which are relatively easy to compute, we can determine when the quadratic form $ax^2 + by^2$ represents 1 for given $a, b \in K$ in the cases $K = \mathbb{Q}_p$ and $K = \mathbb{R}$. The Hasse–Minkowski principle tells us that we are now in a position to answer the same question when $K = \mathbb{Q}$. For a marvelous exposition of this principle see [Con, p. 135] or [Se 1, Ch. 4].

Theorem 4.12 (Hasse–Minkowski). *Let $m \in \mathbb{Q}^*$ and let v be a rational prime or ∞. In order that a quadratic form f with $\mathbb{Q}$-coefficients represent m in $\mathbb{Q}$ it is necessary and sufficient that it does so in each $\mathbb{Q}_v$.* □

In the sequel, will shorten the phrase "v is a rational prime or ∞" by saying v is a *place.* Theorems 4.7 and 4.11, together with the Hasse–Minkowski principle yield the following important result.

Theorem 4.13. *Let $D = (a, b)_{\mathbb{Q}}$ be a rational quaternion algebra, and let v be a place. Then*

$$D \otimes \mathbb{Q}_v \cong \begin{cases} M_2(\mathbb{Q}_v) & \text{if } (a, b)_v = 1, \\ D_v & \text{otherwise,} \end{cases}$$

where D_v is a division algebra over $\mathbb{Q}_v$. We say a place v ramifies *in D if $D \otimes \mathbb{Q}_v$ is a division algebra.* □

Computationally, this result is of great use since Hilbert symbols are relatively easy to calculate using Quadratic Reciprocity. At the same time, Theorem 4.13 raises a wealth of questions worth exploring. For example, does the set of primes which ramify in D classify D in some sense? Is this set finite? Given a finite set of places S, is there a rational quaternion algebra D that only ramifies at the places contained in S? What can we say about the various division algebras D_p? (They are all isomorphic; see [Lam, Ch. 6].)

It turns out that the set S of primes that ramify in D *determines* the quaternion algebra. This is a strong result; one need only look at imaginary quadratic extensions of $\mathbb{Q}$ to find examples of distinct fields that have the same set of primes that ramify in them; for example, $\mathbb{Q}(i), \mathbb{Q}(\sqrt{-2})$ both have $S = \{2, \infty\}$.

We turn now to the question of finiteness of the set S. This set is indeed finite. Moreover, it has even cardinality. Both these claims are immediate from the following theorem for Hilbert symbols, which is but a convenient restatement of Quadratic Reciprocity.

Theorem 4.14. *Let $a, b \in \mathbb{Q}^*$. Then $(a, b)_v = 1$ except at a finite number of places and*

$$\prod_v (a, b)_v = 1,$$

where v runs through all rational primes and ∞.

Proof. First, $(a, b)_v = 1$ for all but finitely many v because $a, b \in \mathbb{Z}^*_{(p)}$ for all but finitely many v. The claim then follows from Theorem 4.9,(iv).

In view of Theorem 4.9 it suffices to show the product law in the following three cases:

- a, b are positive odd prime numbers: In this case $(a, b)_p = 1$ for $p \neq a, b, 2$. Theorem 4.9 (i),(iv),(v) then tells us

$$(a, b)_a = \left(\frac{b}{a}\right), \quad (a, b)_b = \left(\frac{a}{b}\right), \quad (a, b)_2 = (-1)^{\frac{(a-1)(b-1)}{4}}.$$

The product law is then a restatement of Quadratic Reciprocity.

- a is an odd prime, $b = -1$ or 2: In this case $(a,b)_p = 1$ for $p \neq a, 2$ and using Theorem 4.9 we calculate

$$\begin{aligned}(a,-1)_a &= \left(\frac{-1}{a}\right), & (a,-1)_2 =&(-1)^{\frac{a-1}{2}},\\ (a,2)_a &= \left(\frac{2}{a}\right), & (a,2)_2 =&(-1)^{\frac{a^2-1}{8}}.\end{aligned}$$

The product law is again a restatement of well-known properties of the Legendre symbol.

- $a = -1$, $b = -1$ or 2: In this case $(-1,2)_v = 1$ for all v so the product formula holds, and

$$(-1,-1)_v = \begin{cases} -1 & \text{if } v = 2 \text{ or } \infty,\\ 1 & \text{otherwise,}\end{cases}$$

so once again the product formula holds. □

Corollary 4.15. *The set of primes S at which a rational quaternion algebra D ramifies is finite and has even cardinality.* □

In order to calculate the Hilbert symbols at all places, it is enough to do so for all but one place, by use of the product formula. Though this may not seem like much at first, it may be an advantage since the Hilbert symbol can sometimes be hard to calculate for $p = 2$. The product rule allows us to omit one computation if we know the Hilbert symbol for a pair of nonzero rational numbers at all other places.

4.2.2 Rational Quaternion Algebras Ramified at One Finite Prime

Since a rational quaternion algebra must ramify at an even number of places, not every finite set S corresponds to a quaternion algebra the ramifies precisely at the places contained in S. In this section we will settle the particular case of the existence of a rational quaternion algebra ramified at one finite prime (i.e., $S = \{p, \infty\}$). We denote such algebras by $B_{\{p,\infty\}}$. This nontrivial case illustrates how useful Hilbert symbols are when dealing with quaternion algebras; the particular example is also of immense importance to us since the endomorphism ring of an elliptic curve with supersingular reduction at a prime p is a maximal order in such a quaternion algebra.

- $B_{\{2,\infty\}}$. In the course of proving the product formula we saw that

$$(-1,-1)_v = \begin{cases} -1 & \text{if } v = 2 \text{ or } \infty,\\ 1 & \text{otherwise,}\end{cases}$$

This means the quaternion algebra $D = (-1,-1)_{\mathbb{Q}}$ ramifies at S.

- $B_{\{p,\infty\}}, p \equiv 3 \bmod 4$. We claim the quaternion algebra $D = (-1,-p)_{\mathbb{Q}}$ ramifies precisely at S. Indeed, Theorem 4.9,(iv) tells us $(-1,-p)_v = 1$ for $v \neq 2, p, \infty$. Clearly $(-1,-p)_\infty = -1$. By Theorem 4.9,(iv)

$$(-1,-p)_p = \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = -1 \quad \text{since } p \equiv 3 \bmod 4.$$

Finally, the product formula implies that $(-1,-p)_2 = 1$.

- $B_{\{p,\infty\}}, p \equiv 5 \bmod 8$. In this case, $D = (-2,-p)_{\mathbb{Q}}$ ramifies only at S. Again, Theorem 4.9,(iv) tells us $(-1,-p)_v = 1$ for $v \neq 2, p, \infty$, and $(-2,-p)_\infty = -1$. Just for fun, let us calculate the Hilbert symbol at 2 this time:

  $$(-2,-p)_2 = (-1)^{\frac{p^2-1}{8}}(-1)^{\frac{p-1}{2}\cdot -1} = (-1)^{\frac{(p+1)(p+3)}{8}} = 1.$$

  The last equality follows because $(p+3)/8$ is an integer and $(p+1)$ is even. The product formula implies that $(-1,-p)_p = -1$.

- $B_{\{p,\infty\}}, p \equiv 1 \bmod 8$. This is the trickiest case. Let $D = (-q,-p)_{\mathbb{Q}}$, where q is a rational prime congruent to 3 mod 4 and which is not a square mod p. It follows from Theorem 4.9,(iv) that $(-q,-p)_v = 1$ for all $v \neq 2, p, q, \infty$ and it is clear that $(-q,-p)_\infty = -1$. Now

  $$(-q,-p)_p = \left(\frac{-q}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{q}{p}\right) = -1$$

  since $p \equiv 1 \bmod 4$ and q is not a square mod p. Next,

  $$(-q,-p)_q = \left(\frac{-p}{q}\right) = (-1)^{\frac{q-1}{2}}\left(\frac{q}{p}\right) = -1\cdot -1 = 1,$$

  where the last equality is a consequence of Quadratic Reciprocity. By the product formula, $(-q,-p)_2 = 1$, thus completing our verification.

Remark 4.16. In the above construction we conjured up a prime q congruent to 3 mod 4 and which is a square mod p. Why does such a prime exist? By the Chinese remainder theorem these two conditions are equivalent to a congruence for q modulo $4p$. Dirichlet's theorem on arithmetic progressions asserts there are infinitely many q's that satisfy the congruence relation. For a proof of this theorem see [Se 1, Ch. 6].

4.3 Orders

Now that we have studied how to classify quaternion algebras over $\mathbb{Q}$, it is time to turn our attention to orders inside these algebras. After all, the endomorphism ring of a supersingular elliptic curve is not a full quaternion algebra, but rather an order sitting inside one. Many of the theorems we will prove hold in much more generality, though the proofs we present may not be easy to generalize. For a complete treatment of the material presented here the reader should consult [Vig].

Let R be a Dedekind domain, K its fraction field and H a quaternion algebra over K. An *ideal* I is a subset of H which is finitely generated as an R-module and $K \otimes_R I = H$. For every ideal I of H, there exists a set $I^{-1} = \{h \in H \mid IhI \subset I\}$ which is also an ideal. The *reduced norm* of an ideal $n(I)$ is the fractional ideal of R generated by the norms of the elements of I; the reduced norm is multiplicative, i.e., $n(I)n(J) = n(IJ)$, where IJ is the set of finite sums of elements ij with $i \in I$ and $j \in J$.

An element $x \in H$ is said to be *integral* over R if $R[x]$ is a finitely generated R-module contained in H. The sum and product of two integral elements need not be integral, i.e., the integral elements of H over R do not in general form a ring.

Lemma 4.17. *An element $x \in H$ is integral over R if and only if its reduced trace and its reduced norm belong to R (cf. [Vig, Lemma I.4.1]).* □

By an *order* $\mathcal{O}$ of H we mean an ideal which is also a ring, or, equivalently, a ring $\mathcal{O}$ of R-integral elements which contains R and such that $K\mathcal{O} = H$. An order is *maximal* if it is not properly contained in any other order; every order is contained in a maximal one (cf. [Vig, Proposition I.4.2]).

4.3.1 The Different and the Reduced Discriminant

Let $\mathcal{O}$ be an order in a quaternion algebra H as above. The dual of $\mathcal{O}$ with respect to the bilinear form induced by the reduced trace of H is the set

$$\mathcal{O}^* = \{x \in H \mid t(x\mathcal{O}) \subset R\}.$$

It is an ideal. The *different* of $\mathcal{O}$, denoted by $\mathcal{O}^{*-1}$, is the inverse of this dual. It is integral over R (cf. [Vig, Lemma I.4.7]). The *reduced discriminant* of an order $\mathcal{O}$ is the reduced norm of its different:

$$\mathrm{disc}(\mathcal{O}) := n(\mathcal{O}^{*-1})$$

Theorem 4.18. *With the notation as above, if the order $\mathcal{O}$ of a quaternion algebra H is a free R-module then*

$$\mathrm{disc}(\mathcal{O})^2 = R\det(t(v_iv_j)),$$

where $\{v_i\}$ is a basis of $\mathcal{O}$ over R.

Proof. The dual basis to $\{v_i\}$, denoted $\{v_i^*\}$, defined by the property $T(v_iv_j^*) = \delta_{ij}$ is a generating set for the ideal $\mathcal{O}^*$ over R. If $v_j^* = \sum_k a_{jk}v_k$, then linearity of the trace map gives

$$t(v_iv_j^*) = \sum_k a_{jk}t(v_iv_k),$$

from which it follows that

$$1 = \det(t(v_iv_l^*)) = \det(a_{jk})\det(t(v_iv_j)). \tag{4.2}$$

We assume now the order is a principal ring (the theorem is still valid otherwise–cf. [Vig, Lemma I.4.7]). since $\mathcal{O}$ is a principal ring there is some $x \in H^*$ with $\mathcal{O}^* = \mathcal{O}x$. This means $\{v_ix\}$ is another basis for $\mathcal{O}^*$ as an R-module, so there exists a change of basis matrix $[b_{ij}]$ such that

$$v_ix = \sum_j b_{ij}v_j^* = \sum_{j,k} b_{ij}a_{jk}v_k.$$

Hence the map $H^* \to H^*$ given by $h \mapsto hx$ has matrix $[\sum_j b_{ij}a_{jk}]$ for the basis $\{v_i\}$. On the other hand one checks this map has determinant $n^2(x)$, from which we conclude that

$$\det\left(\sum_j b_{ij}a_{jk}\right) = \det(b_{ij})\cdot\det(a_{jk}) = n^2(x).$$

By (4.2) we obtain

$$1 = n^2(x)\cdot\det(b_{ij})^{-1}\det(t(v_iv_j)),$$

and since $\det(b_{ij}) \in R^*$

$$\mathrm{disc}(\mathcal{O})^2 = n^2(\mathcal{O}^{*-1}) = n^{-2}(\mathcal{O}x) = R\det(t(v_iv_j)),$$

as claimed. □

Let $\mathcal{O}$ and $\mathcal{O}'$ be two (comparable) orders such that $\mathcal{O} \subset \mathcal{O}'$. If $\{v_i\}$ and $\{w_i\}$ are bases for $\mathcal{O}$ and $\mathcal{O}'$, respectively, then $v_i = \sum_j a_{ij} w_j$ and

$$\det(t(v_i v_j)) = \det{}^2(a_{ij}) \det(t(w_i w_j)).$$

In this way we obtain the following useful corollary of Theorem 4.18.

Corollary 4.19. *If $\mathcal{O}$ and $\mathcal{O}'$ are two orders such that $\mathcal{O} \subset \mathcal{O}'$ then* $\mathrm{disc}(\mathcal{O}) \subset \mathrm{disc}(\mathcal{O}')$ *with equality if and only if $\mathcal{O} = \mathcal{O}'$.* □

Example 4.2. The order $M_2(R)$ of $M_2(K)$ is maximal since its reduced discriminant is R. In fact, when R is a principal ideal ring, the maximal orders of $M_2(K)$ are all conjugate to $M_2(R)$ (cf. [Vig, p. 28])

4.3.2 Division Quaternion Algebras over $\mathbb{Q}_p$

The reader is referred to [Vig, § II.1] for proofs (and a more general and extensive treatment) of the material in this section.

Let H be a *division* quaternion algebra over $\mathbb{Q}_p$, and let $v_p : \mathbb{Q}_p^* \to \mathbb{Z}$ be the usual p-adic valuation. We may use this map, together with the reduced norm map of H^* to construct a discrete valuation on H^*. Explicitly, the map

$$\begin{aligned} \omega : H^* &\to \mathbb{Z} \\ h &\mapsto v_p(n(h)) \end{aligned}$$

is the desired discrete valuation. The valuation ring of ω, i.e., the set $\{x \in H \mid \omega(x) \geq 0\}$ is the *unique* maximal order of H and its reduced discriminant is $p\mathbb{Z}_p$.

4.3.3 Maximal Orders in Rational Quaternion Algebras

The following theorem gives us a simple criterion to determine whether an order in a quaternion algebra over $\mathbb{Q}$ is maximal or not. This is the main result of the chapter, and will be of great use in our study of primes dividing Gross–Zagier numbers. We will need two lemmas, the first of which which says that maximality is a local property (cf. [Vig, § III.5]).

Lemma 4.20. *Let H be a rational quaternion algebra, and let $\mathcal{O}$ be an order in H. Then $\mathcal{O}$ is a maximal order if and only if $\mathcal{O}_p = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is a maximal order in $H_p = H \otimes_{\mathbb{Q}} \mathbb{Q}_p$ for each finite rational prime p.* □

Lemma 4.21. *Let I be an ideal and $\mathcal{O}$ be an order of a rational quaternion algebra H. Then*

$$n(I)_p = n(I_p) \quad \textit{and} \quad \mathrm{disc}(\mathcal{O})_p = \mathrm{disc}(\mathcal{O}_p).$$

Proof. Let $\{y_i\}$ be a (finite) system of $\mathbb{Z}$-generators for I. The first equality follows because $\{y_i \otimes 1\}$ is a system of generators for $I_P = I \otimes_{\mathbb{Z}} \mathbb{Z}_p$ over $\mathbb{Z}_p$ and the definition of the norm of an ideal. For the equality of reduced discriminants, first recall that

$$I^* = \{x \in H \mid t(xy) \in \mathbb{Z} \ \forall \ y \in I\}.$$

Using the properties of the tensor product, one checks that $(I_p)^* = (I^*)_p$. Since $\mathcal{O}^*$ is an ideal we obtain

$$\mathrm{disc}(\mathcal{O})_p = n(\mathcal{O}^{*-1})_p = [n(\mathcal{O}^*)^{-1}]_p = n(\mathcal{O}^*)_p^{-1} = n({\mathcal{O}^*}_p^{-1}) = \mathrm{disc}(\mathcal{O}_p).$$ □

Theorem 4.22. *Let H be a rational quaternion algebra, and let $\mathcal{O}$ be an order in H. Then $\mathcal{O}$ is a maximal order if and only if*

$$\operatorname{disc}(\mathcal{O}) = \prod_{\substack{p \in \operatorname{Ram} H \\ p \neq \infty}} p\mathbb{Z}, \tag{4.3}$$

where $\operatorname{Ram} H$ *is the set of places that ramify in H.*

Proof. Suppose that $\mathcal{O}$ is a maximal order. By Lemma 4.20, $\mathcal{O}_p$ is a maximal order in H_p for every finite rational prime p. We know from Theorem 4.13 that

$$H_p = H \otimes_{\mathbb{Q}} \mathbb{Q}_p \cong \begin{cases} M_2(\mathbb{Q}_p) & \text{if } p \notin \operatorname{Ram} H, \\ D_p & \text{otherwise,} \end{cases}$$

- *Case 1:* $\mathcal{O}_p$ is a maximal order in $M_2(\mathbb{Q}_p)$. Then $\mathcal{O}_p$ is conjugate to $M_2(\mathbb{Z}_p)$ (cf. Example 4.2). Set
$$v_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$
We immediately check that $\operatorname{disc}(M_2(\mathbb{Z}_p)) = \mathbb{Z}_p \det(t(v_iv_j)) = \mathbb{Z}_p$. In general, $\mathcal{O}_p = gM_2(\mathbb{Z}_p)g^{-1}$ for some $g \in M_2(\mathbb{Q}_p)$, and since
$$t(gv_ig^{-1}gv_jg^{-1}) = t(gv_iv_jg^{-1}) = t(v_iv_j),$$
it follows that $\operatorname{disc}(gM_2(\mathbb{Z}_p)g^{-1}) = \mathbb{Z}_p$ as well. In any case, $\operatorname{disc}(\mathcal{O}_p) = \mathbb{Z}_p$, so that $\operatorname{disc}(\mathcal{O})_p = \mathbb{Z}_p$ from Lemma 4.21, which is to say that $p \nmid \operatorname{disc}(\mathcal{O})$.

- *Case 2:* $H_p = D_p$ is a division algebra. Then $\mathcal{O}_p$ is the unique maximal order of H_p and $\operatorname{disc}(\mathcal{O}_p) = p\mathbb{Z}_p$ (cf. §4.3.2). Hence by Lemma 4.21 $\operatorname{disc}(\mathcal{O})_p = \operatorname{disc}(\mathcal{O}_p) = p\mathbb{Z}_p$ and therefore $p|\operatorname{disc}(\mathcal{O})$.

Collecting these results (4.3) follows immediately.

Now suppose that (4.3) holds. We want to show that $\mathcal{O}$ is a maximal order of H. By Lemma 4.20, it is enough to show that $\mathcal{O}_p$ is a maximal order of H_p for every finite rational prime p. By (4.3) and Lemma 4.21 we see that

$$\operatorname{disc}(\mathcal{O}_p) = \operatorname{disc}(\mathcal{O})_p = \begin{cases} \mathbb{Z}_p & \text{if } p \notin \operatorname{Ram} H, \\ p\mathbb{Z}_p & \text{otherwise.} \end{cases}$$

- *Case 1*: $\operatorname{disc}(\mathcal{O}_p) = \mathbb{Z}_p$. Note that in this case $H_p = M_2(\mathbb{Q}_p)$. Since every order is contained in a maximal order, $\mathcal{O}_p \subset gM_2(\mathbb{Z}_p)g^{-1}$ for some $g \in M_2(\mathbb{Q}_p)$. However, $\operatorname{disc}(\mathcal{O}_p) = \operatorname{disc}(gM_2(\mathbb{Z}_p)g^{-1}) = \mathbb{Z}_p$. By Corollary 4.19 $\mathcal{O}_p = gM_2(\mathbb{Z}_p)g^{-1}$, which is to say that $\mathcal{O}_p$ is a maximal order in H_p.

- *Case 2:* $\operatorname{disc}(\mathcal{O}_p) = p\mathbb{Z}_p$. Here H_p is a division algebra. As remarked before, such a division algebra has a unique maximal order $\mathcal{O}_{\max}$, and $\mathcal{O}_p$ must be contained in $\mathcal{O}_{\max}$. Since $\operatorname{disc}(\mathcal{O}_p) = p\mathbb{Z}_p = \operatorname{disc}(\mathcal{O}_{\max})$ it follows by Corollary 4.19 that $\mathcal{O}_p = \mathcal{O}_{\max}$, i.e., $\mathcal{O}_p$ is maximal in H_p. □

Example 4.3. Let $H = (-1, -1)_{\mathbb{Q}}$. Recall this algebra ramifies at 2 and ∞. By Theorem 4.22 a maximal order in H has reduced discriminant $2\mathbb{Z}$. The order $\mathbb{Z}[1, i, j, k]$ has reduced discriminant $4\mathbb{Z}$, as is easily checked using Theorem 4.18, and therefore is not a maximal order in H.

In general, when given a candidate order $\mathcal{O}$ in a rational quaternion algebra H, it suffices to check the following to make sure the order is maximal:

1. $\mathcal{O}$ in fact a ring.
2. Every element of the order is integral, i.e., their reduced trace and norms are in $\mathbb{Z}$. Given a set of $\mathbb{Z}$-generators for $\mathcal{O}$, it is enough to show the sum and product of two generators are integral. we must remember to check both possible products of two generators (H is not commutative).
3. The candidate order is a finitely generated $\mathbb{Z}$-module and $\mathbb{Q}\mathcal{O} = H$.
4. The reduced discriminant satisfies (4.3).

4.3.4 Maximal Orders of the Rational Quaternion Algebras $B_{\{p,\infty\}}$

With the above algorithm for checking the maximality of an order, the reader may now verify that the following are examples of maximal orders in rational quaternion algebras of the form $B_{\{p,\infty\}}$

- $B_{\{2,\infty\}}$, e.g., $(-1, -1)_{\mathbb{Q}}$. Then $\mathbb{Z}\left[1, i, j, \frac{1+i+j+ij}{2}\right]$ is maximal.
- $B_{\{p,\infty\}}$ for $p \equiv 3 \bmod 4$, e.g., $(-1, -p)_{\mathbb{Q}}$. Then $\mathbb{Z}\left[1, i, \frac{i+j}{2}, \frac{1+ij}{2}\right]$ is maximal.
- $B_{\{p,\infty\}}$ for $p \equiv 5 \bmod 8$, e.g., $(-2, -p)_{\mathbb{Q}}$. Then $\mathbb{Z}\left[1, \frac{1+i+j}{2}, j, \frac{2+i+ij}{4}\right]$ is maximal.
- $B_{\{p,\infty\}}$ for $p \equiv 1 \bmod 8$, e.g., $(-q, -p)_{\mathbb{Q}}$, for some $q \equiv 3 \bmod 4p$. Then $\mathbb{Z}\left[\frac{i+j}{2}, \frac{j-pij}{2}, ij\right]$ is maximal.

Chapter 5

Supersingular Elliptic Curves

We have seen that in characteristic zero an elliptic curve's endomorphism ring cannot be an order in a quaternion algebra (cf. Theorem 2.11). In this chapter we will focus on the study of elliptic curves over fields of positive characteristic whose endomorphism ring is an order in a rational quaternion algebra. Such curves are said to be *supersingular.* There are many ways to characterize this phenomenon. We will study a few of them in this chapter. Our exposition is synthesized from [Sil 1, V.3]; Silverman in turn bases his exposition on Deuring's comprehensive article on the subject [Deu].

After presenting a few criteria for supersingularity, we will prove $\operatorname{End} E$ is a *maximal* order in a rational quaternion algebra for curves defined over a separable closure of the field with p elements, denoted $\mathbb{F}_p$. This is a hard theorem; we will need to develop the machinery of formal group laws and invoke two difficult theorems of Tate to prove this result.

5.1 Supersingular Elliptic Curves

Let k be a perfect field of positive characteristic p.

Theorem 5.1. *Let E be an elliptic curve over k. Then the following are equivalent:*

(i) E is supersingular.

(ii) The isogeny $\hat{\phi}_r : E^{(p^r)} \to E$ dual to the p^r-power Frobenius isogeny is inseparable for one $r \geq 1$.

(iii) The isogeny $\hat{\phi}_r$ is purely inseparable for all $r \geq 1$.

(iv) $E[p^r] = 0$ for all $r \geq 1$.

(v) The map $[p] : E \to E$ is purely inseparable and $j(E) \in \mathbb{F}_{p^2}$.

Proof. (i $\Longrightarrow$ ii) Suppose that all the maps $\hat{\phi}_r$ are separable. We will show that in such a case $\operatorname{End} E$ is commutative, so it cannot be an order in a quaternion algebra.

The first step is to show the map $\operatorname{End} E \to \operatorname{End} T_p(E)$ is injective. (Note Theorem 2.14 will not suffice in this case since we assumed $l \neq p$ as part of our hypotheses.) Suppose $\psi \in \operatorname{End} E$ maps to 0 in $\operatorname{End} T_p(E)$. This means $\psi(E[p^r]) = 0$ for all $r \geq 1$.

Let $P \in E^{(p^r)}$ be an element of $\ker \hat{\phi}_r$. Since ϕ_r is a surjective map (it is a non-constant isogeny), there exists a point $Q \in E$ such that $\phi_r(Q) = P$. Recall that $\hat{\phi}_r \circ \phi = [p^r]$ on E; hence $Q \in \ker[p^r] = E[p^r] \subset \ker \psi$. Thus $P \in \phi_r(\ker \psi)$ and we have established the inclusion $\ker \hat{\phi}_r \subset \phi_r(\ker \psi)$, from which we conclude that

$$\# \ker \psi \geq \# \ker \hat{\phi}_r \quad \text{for all } r \geq 1.$$

By Theorem 2.3 we know that $\# \ker \hat{\phi}_r = \deg \hat{\phi}_r$ because $\hat{\phi}_r$ is separable. But $\deg \hat{\phi}_r = \deg \phi_r$ and the degree of the p^r-power Frobenius map is p^r. Thus

$$\# \ker \psi \geq p^r \quad \text{for all } r \geq 1,$$

and so $\psi = 0$; this shows the map $\operatorname{End} E \to \operatorname{End} T_p(E)$ is injective.

Recall that $T_p(E) \cong \{0\}$ or $\mathbb{Z}_p$. If we can show that $T_p(E) \cong \mathbb{Z}_p$ then we will have an injection

$$\operatorname{End} E \hookrightarrow \operatorname{End} T_p(E) \cong \operatorname{End} \mathbb{Z}_p \cong \mathbb{Z}_p$$

which will prove $\operatorname{End} E$ is commutative, giving our desired contradiction.

To see that $T_p(E) \cong \mathbb{Z}_p$, suppose $\hat{\phi}_1$ is separable (if it isn't we are done!). Then

$$\# E[p] = \# \ker[p] = \deg_s[p] = \deg_s \hat{\phi}_1 = \deg \hat{\phi}_1 = \deg \phi_1 > 1.$$

Hence $E[p] \neq 0$. Now, by definition of the Tate module we know

$$T_p(E)/pT_p(E) \cong E[p] \neq 0.$$

Hence $T_p(E) \neq 0$ and so $T_p(E) \cong \mathbb{Z}_p$.

(ii $\iff$ iii) Fix r such that $\hat{\phi}_r$ is inseparable. Since $\hat{\phi}_r \circ \phi_r = [p^r]$ and the Frobenius map is purely inseparable, we have

$$\deg_s \hat{\phi}_r = \deg_s \hat{\phi}_r \deg_s \phi_r = \deg_s[p^r] = (\deg_s[p])^r = (\deg_s \hat{\phi}_1)^r \tag{5.1}$$

and so $\deg_s \hat{\phi}_r = 1$ or p^r. Since $\hat{\phi}_r$ is inseparable, it must be that $\deg_s \hat{\phi}_r = 1$ and so $\hat{\phi}_r$ is purely inseparable. Furthermore, the above chain of inequalities shows that $\deg_s \hat{\phi}_1 = 1$, so that $\deg_s \hat{\phi}_r = (\deg_s \hat{\phi}_1)^r = 1$ *for all* $r \geq 1$. The other direction is obvious.

(iii $\iff$ iv) By (5.1) we know that $\deg_s \hat{\phi}_r = 1$ or p^r and since $\hat{\phi}_r$ is purely inseparable for every $r \geq 1$, it follows that $\deg_s \hat{\phi}_r = 1$ for all $r \geq 1$. Theorem 2.3 combined with (5.1) tell us that

$$\# E[p^r] = \# \ker[p^r] = \deg_s \hat{\phi}_r,$$

hence the equivalence of (iii) and (iv).

(iii $\implies$ v) Since $\hat{\phi}_r$ is purely inseparable for all $r \geq 1$, it is in particular purely inseparable for $r = 1$. However, $[p] = \hat{\phi}_1 \circ \phi_1$ and the Frobenius map is purely inseparable; therefore $[p]$ is purely inseparable.

To see that $j(E) \in \mathbb{F}_{p^2}$ we separate the map $\hat{\phi}_1$ into its separable and purely inseparable parts:

$$\begin{tikzcd} E^{(p)} \arrow[r, "\hat{\phi}_1"] \arrow[d, "\phi'"'] & E \\ E^{(p^2)} \arrow[ur, "\psi"] & \end{tikzcd}$$

The map $\hat{\phi}_1$ has degree p, as does the Frobenius map ϕ'. Thus ψ is a map of degree 1 and is consequently an isomorphism. Since the j-invariant classifies curves up to isomorphism, we have

$$j(E) = j(E^{(p^2)}) = j(E)^{p^2}.$$

To see why the last equality is true, recall the j-invariant is a homogeneous expression on the coefficients of E; raising these coefficients to any power has the effect of raising the j-invariant to that same power because we are working over in a field of positive characteristic.

(v $\implies$ i) Suppose (i) is false. Then by Theorem 2.17 $\mathcal{K} := \operatorname{End} E \otimes \mathbb{Q}$ is either the field of rational numbers or a quadratic imaginary extension of it.

Let E' be an elliptic curve isogenous to E by a map ψ. Since $\psi \circ [p] = [p] \circ \psi$ and $[p]$ is inseparable in $\operatorname{End} E$, it follows by a degree count that $[p]$ is inseparable in $\operatorname{End} E'$. This means

$$\# E'[p] = \deg_s [p] = 1,$$

where the first equality is a consequence of Theorem 2.3. By the chain of implications $(iv) \implies (iii) \implies (v)$ it follows that $j(E') \in \mathbb{F}_{p^2}$. Hence there are finitely many isomorphism classes of curves isogenous to E.

Choose a prime $l \in \mathbb{Z}$ different from p such that l is prime in each $\operatorname{End} E'$ as E' ranges through isomorphism classes of curves isogenous to E (since there are finitely many such classes the integer l exists). Then Theorem 2.13 gives $E[l^i] \cong \mathbb{Z}/l^i\mathbb{Z} \times \mathbb{Z}/l^i\mathbb{Z}$ for all $i \geq 1$. Since each $E[l^i]$ is a subgroup of E there exists a sequence of groups of points of E

$$H_1 \subset H_2 \subset \cdots \subset E \quad \text{with } H_i \cong \mathbb{Z}/l^i\mathbb{Z}$$

By Theorem 2.4 there exists an elliptic curve E_i and an isogeny $E \to E_i$ whose kernel is H_i. Since there are finitely many isomorphism classes of curves isogenous to E there are only finitely many distinct E_i. Thus, for some pair of positive integers m, n The curves E_{m+n} and E_n are isomorphic. Let $\pi : E_m \to E_{m+n}$ be the natural projection map. The map

$$\lambda : E_m \xrightarrow{\pi} E_{m+n} \xrightarrow{\sim} E_m$$

is an endomorphism of E_m. By construction, its kernel is H_{m+n}/H_n, and is therefore cyclic of order l^n. On the other hand, $l^n = \# \ker \lambda = \deg \lambda$ and since l is prime in $\operatorname{End} E_m$ by assumption, we must conclude n is even and $\lambda = u \circ [l^{n/2}]$. However, the kernel of the map $[l^{n/2}]$ is never cyclic, and this is a contradiction. □

So far we know that in the supersingular case, the endomorphism ring of a curve is an order in a rational quaternion algebra. We will devote the rest of this chapter to making this result more precise. First, we will see which quaternion algebras correspond to supersingular elliptic curves.

Theorem 5.2. *Let E/k be a supersingular curve. Then the quaternion algebra $H = \operatorname{End} E \otimes \mathbb{Q}$ is ramified only at $p = \operatorname{char} k$ and infinity.*

Proof. Let l be a prime different from p. Recall there is an injection (cf. Theorem 2.14)

$$\operatorname{End} E \otimes \mathbb{Z}_l \hookrightarrow \operatorname{End} T_l(E).$$

For $l \neq p$ we know $T_l(E) \cong \mathbb{Z}_l \times \mathbb{Z}_l$, so $\operatorname{End} T_l(E) \cong M_2(\mathbb{Z}_l)$. The above injection gives in turn an injection of the quaternion algebra $H \otimes_{\mathbb{Q}} \mathbb{Q}_l$ into $M_2(\mathbb{Q}_l)$. A dimension count tells us

$H \otimes_{\mathbb{Q}} \mathbb{Q}_l \cong M_2(\mathbb{Q}_l)$, so l does not ramify in H. We saw in our proof to Theorem 2.17 that the quaternion algebra H is of the form $(a, b)_{\mathbb{Q}}$, where a, b are negative rational numbers. This means the Hilbert symbol $(a, b)_\infty$ is negative and by Theorem 4.13 the algebra H must ramify at infinity. Since a rational quaternion algebra must ramify at an even number of places (by the Hilbert product law), H must also ramify at p. □

Now we focus our efforts on the endomorphism ring of a supersingular curve. To show this ring is a maximal order in $B_{\{p,\infty\}}$ we will need to develop some machinery involving formal groups. We will return to our goal in §5.5.

5.2 The Formal Group Law of an Elliptic Curve

The basic theory of formal group laws presented in this section can be found in more detail in [Sil 1, Ch. IV] and [Lang, Appendix].

R will always denote a commutative ring with unit. A one-dimensional commutative formal group law over R is a power series $F(X, Y) \in R[\![X, Y]\!]$ that satisfies

(i) $F(X, Y) \equiv X + Y \mod \deg 2$.

(ii) $F(X, F(Y, Z)) = F(F(X, Y), Z)$.

(iii) $F(X, Y) = F(Y, X)$.

(iv) There is a unique power series $i(X) \in R[\![X]\!]$ such that $F(X, i(X)) = 0$.

(v) $F(X, 0) = X = F(0, X)$.

Here, two power series are said to be congruent $\mod \deg n$ if they coincide on all terms of degree strictly less than n.

Example 5.1. The formal additive group law

$$F(X, Y) = X + Y.$$

Example 5.2. The formal multiplicative group law

$$F(X, Y) = X + Y + XY.$$

Our next example is the one of interest for our purposes. We will show how to associate a formal group law to an elliptic curve. In essence, we will "steal" the abelian group law on the curve. Recall how we add two points P and Q on an elliptic curve. First, we consider the line joining P and Q (if $P = Q$ we consider the line tangent to the curve at that point). This line intersects the curve at a third auxiliary point. Next, we consider the line between this auxiliary point and the distinguished point O (the origin). This line intersects the curve at a third point, which we call $P + Q$.

Thus far, we have avoided explicit work with Weierstrass equations for elliptic curves. It would be inconvenient to continue with this approach in our discussion of formal group laws (this will become apparent soon). A *Weierstrass* model for an elliptic curve E is an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{5.2}$$

where $a_i \in \overline{K}$ for $i = 1, \ldots, 6$; if $a_i \in K$ we say E is defined over K. *Every* elliptic curve (including those over characteristic 2 or 3 fields) has a Weierstrass equation. When $\operatorname{char} K \neq 2, 3$ this equation may be simplified to one of the form given in Chapter 2.

For an elliptic curve E defined by a Weierstrass equation (5.2) we define

$$z = -\frac{x}{y}, \ w = -\frac{1}{y} \quad \text{so} \quad x = \frac{z}{w}, \ y = -\frac{1}{w}. \tag{5.3}$$

In these new variables, the Weierstrass equation (5.2) becomes

$$w = z^3 + a_1 zw + a_2 z^2 w + a_3 w^2 + a_4 zw^2 + a_6 w^3. \tag{5.4}$$

Let $f(z, w)$ denote the right hand side of (5.4). Without regards to convergence we substitute the equation into itself recursively. This way we obtain a power series expansion in z for w:

$$\begin{aligned} w &= z^3 + a_1 z^4 + (a_1^2 + a_2) z^5 + (a_1^3 + 2a_1 a_2 + a_3) z^6 && (5.5)\\ &\quad + (a_1^4 + 3a_1^2 a_2 + 3a_1 a_3 + a_2^2 + a_4) z^7 + \cdots && (5.6)\\ &= z^3 (1 + A_1 z + A_2 z^2 + \cdots), && (5.7) \end{aligned}$$

where $A_n \in \mathbb{Z}[a_1, \ldots, a_6]$ is a polynomial of weight n in the a_i. This procedure gives a power series in $\mathbb{Z}[a_1, \ldots, a_6][\![z]\!]$ that satisfies the equation

$$w(z) = f(z, w(z))$$

(cf. [Sil 1, Proposition IV.1.1]). Using (5.3) we obtain the Laurent series expansions

$$\begin{aligned} x &= \frac{1}{z^2} - \frac{a_1}{z} - a_2 - a_3 z - (a_4 + a_1 a_3) z^2 + \cdots && (5.8)\\ y &= \frac{x}{z} = -\frac{1}{z^3} + \frac{a_1}{z^2} + \frac{a_2}{z} + a_3 + \cdots . && (5.9) \end{aligned}$$

The coefficients of these expansions are in $\mathbb{Z}[a_1, \ldots, a_6]$.

To obtain a formal group law from E we consider the power series that formally gives the addition law on E. Let $P_i = (z_i, w_i(z_i))$, $i = 1, 2$. The line connecting P_1 to P_2 has slope

$$\lambda = \frac{w_2 - w_1}{z_2 - z_1} = \sum_{n=3}^{\infty} A_{n-3} \frac{z_2^n - z_1^n}{z_2 - z_1} \in \mathbb{Z}[a_1, \ldots, a_6][\![z_1, z_2]\!],$$

where the A_n are the polynomials from before. The line connecting P_1 and P_2 has equation $w = \lambda z + v$, where $v = w_1 - \lambda z_1$. Substituting this expression for w into the (z, w)-Weiertrass equation (5.4) we obtain a cubic equation in z; z_1 and z_2 give two solutions for the equation. Looking at the coefficient of the quadratic term we find the third root z_3:

$$z_3 = -z_1 - z_2 + \frac{a_1 \lambda + a_3 \lambda^2 - a_2 v - 2a_4 \lambda v - 3a_6 \lambda^2 v}{1 + a_2 \lambda + a_4 \lambda^2 + a_6 \lambda^3}.$$

The points $(z_1, w_1), (z_2, w_2)$ and (z_3, w_3) add up to O (this follows from the definition of the group law on E). Hence the sum of P_1 and P_2 is the inverse of (z_3, w_3). Using the Weierstrass equation for E one may compute the inverse of a point $(x, y) \in E$ to be $(x, -y - a_1 x - a_3)$ (cf. [Sil 1, p.58]).

Using (5.8) and (5.9), and remembering that $z = -x/y$, the z-coordinate of the inverse of (z, w) is just

$$i(z) = \frac{x}{y + a_1x + a_3} = \frac{z^{-2} - a_1z^{-1} - \cdots}{-z^{-3} + 2a_1z^{-2} + \cdots}.$$

In this way we obtain the formal group law in $\mathbb{Z}[a_1 \ldots, a_6][\![z_1, z_2]\!]$:

$$\begin{aligned} F(z_1, z_2) &= i(z_3) \\ &= z_1 + z_2 - a_1z_1z_2 - a_2(z_1^2z_2 + z_1z_2^2) \\ &\quad - (2a_3z_1^3z_2 - (a_1a_2 - 3a_3)z_1^2z_2^2 + 2a_3z_1z_2^3) + \cdots \end{aligned}$$

Since we have "stolen" the additive law of E, we easily check that $F(z_1, z_2)$ is indeed a formal group law from the corresponding properties of the former law.

A *homomorphism* between two formal group laws F and G is a power series $f(T) \in R[\![T]\!]$ that satisfies

$$f(F(X, Y)) = G(f(X), f(Y)).$$

Example 5.3. Every formal group law F comes equipped with multiplication by m maps, $m \in \mathbb{Z}$, which we denote $[m]_F$. We can define these maps inductively by setting $[0]_F(T) = 0$ and letting

$$\begin{aligned} [m + 1]_F(T) &= F([m]_F(T), T), \\ [m - 1]_F(T) &= F([m]_F(T), i(T)). \end{aligned}$$

These maps are endomorphisms of F. Using forwards and backwards induction one may show that

$$[m](T) = mT + \cdots . \tag{5.10}$$

Example 5.4. Let p be a rational prime. There is a nice way to write the multiplication by p map (cf. [Sil 1, Corollary IV.4.4]):

$$[p](T) = pf(T) + g(T^p)$$

for some power series $f, g \in R[\![T]\!]$ that vanish at 0.

The set of endomorphisms of F forms a ring which we denote $\operatorname{End} F$. We make two preliminary observations about this ring. First, we note that $\operatorname{End} F$ has no zero divisors whenever R is an integral domain. Indeed, suppose that f and g are nonzero elements of $\operatorname{End} E$, and that

$$f \equiv f_rx^r \bmod \deg(r + 1) \quad \text{and} \quad g \equiv g_sx^s \bmod \deg(s + 1),$$

where $f_r, g_s \in R$ are nonzero. Then

$$f \cdot g = f \circ g = f_rg_s^rx^{r+s} \bmod \deg(r + s + 1) \neq 0.$$

Second, if F is a formal group law over a field of positive characteristic p, then the (unique) identity preserving homomorphism $\mathbb{Z} \to \operatorname{End} F$ given in Example 5.3 can be extended to a homomorphism

$$\begin{aligned} \mathbb{Z}_p &\to \operatorname{End} F \\ m = \varprojlim_i m_i &\mapsto \varprojlim_i [m_i]_F, \end{aligned}$$

provided the inverse limit of the $[m_i]_F$ exists (if it does then it would be an endomorphism by the properties of the inverse limit). Using induction on m one may show that

$$[m+p]_F(T) \equiv [m]_F(T) \bmod \deg(m+1).$$

Since $m_{i+1} = m_i + xp^{i+1}$ it follows that

$$[m_{i+1}]_F(T) \equiv [m_i]_F(T) \bmod \deg(m_i + 1) \quad \text{for all } i,$$

whence $\varprojlim [m_i]_F$ exists. It follows that $\operatorname{End} F$ is a $\mathbb{Z}_p$-module.

The ring $\operatorname{End} F$ over a separably closed field of nonzero characteristic will be our main object of study in the next few pages. A good understanding of this ring will provide a key step in describing the endomorphism ring of a supersingular elliptic curve.

5.3 Formal Group Laws in Characteristic p

Let R be a ring of positive characteristic p, and let $f : F \to G$ a homomorphism of formal group laws over R. The *height* of f, which we denote $\operatorname{ht}(f)$ is defined as the largest integer h such that

$$f(T) = g(T^{p^h}),$$

where $g(T)$ is a power series over R. When f is the zero homomorphism we set $\operatorname{ht}(f) = \infty$. The height of the formal group law F is by definition the height of the multiplication by p map. We note that height adds over composition of formal group law homomorphisms.

For example, when m and p are relatively prime, $\operatorname{ht}([m]) = 0$ by (5.10). By Example 5.4 it follows that $[p](T) = g(T^p)$ for some power series g over R since we are working in characteristic p. Hence $\operatorname{ht}([p]) \geq 1$.

It is natural to ask what are the possible heights of the formal group associated to an elliptic curve. The following theorem will answer this question for us [Sil 1, Theorem IV.7.4].

Theorem 5.3. *Let E be an elliptic curve over a field k of positive characteristic p, with associated formal group law F. Let ϕ be an endomorphism of E and denote f its corresponding formal group law homomorphism. Then*

$$\deg_i \phi = p^{\operatorname{ht}(f)}. \qquad \square$$

Corollary 5.4. *Let E be an elliptic curve over a field k of positive characteristic p, with associated formal group law F. Then* $\operatorname{ht}(F) = 1$ *or 2.*

Proof. Let $\phi = [p]$ in Theorem 5.3. We know the map $[p]$ has degree p^2 (cf. Theorem 2.12(iii)); the Corollary follows immediately. □

The height of the formal group law associated to an elliptic curve yields a new criterion for supersingularity.

Theorem 5.5. *Let E be an elliptic curve over a field k of positive characteristic p. Then E is supersingular if and only the formal group law F associated to E has height 2.*

Proof. Let $\phi : E \to E^{(p)}$ be the p-th power Frobenius endomorphism; this map is purely inseparable and has degree p. Since $\hat{\phi} \circ \phi = [p]$ we obtain

$$\deg_i \hat{\phi} = (\deg_i(\hat{\phi} \circ \phi))/p = (\deg_i[p])/p = p^{\mathrm{ht}(F)-1},$$

where the last equality follows from Theorem 5.3. Since $\deg \hat{\phi} = p$ (cf. Theorem 2.12(iv)), and since E is supersingular if and only if $\hat{\phi}$ is purely inseparable it follows E that is supersingular if and only if $\mathrm{ht}(F) = 2$. □

5.4 The Endomorphism Ring of a Formal Group

In this section we will prove the endomorphism ring of a formal group law F over a separably closed field of characteristic $p > 0$ is a maximal order in a local division algebra. This result is central to our proof that the endomorphism ring of a supersingular elliptic curve is a *maximal* order in a rational quaternion algebra. The theorem is due to Lubin (cf. [Lub, §5.1.3]), though our exposition is closer to that in [Froh, Ch. III] or [Haz, §20]; whenever we omit the proof to a standard theorem or lemma concerning formal groups we will give appropriate references from these books.

To begin, we state a result due to Lazard (cf. [Froh, Theorem 1, § III.1]). Define the Lazard polynomials

$$B_n(X,Y) = (X+Y)^n - X^n - Y^n \quad \text{and}$$

$$C_n(X,Y) = \begin{cases} B_n(X,Y) & \text{if } n \text{ is not a prime power,} \\ \frac{1}{l}B_n(X,Y) & \text{if } n = l^r,\ l \text{ a prime number.} \end{cases}$$

Theorem 5.6 (Lazard). *Let F and G be formal group laws over a ring R such that $F \equiv G$ mod $\deg n$. Then there exists an $a \in R$ for which*

$$F \equiv G + aC_n \bmod \deg(n+1).$$ □

Let k be a separably closed field of characteristic $p > 0$, and let $q = p^h$, where h is a fixed positive integer. A formal group law F over k of finite height h is said to be in *normal form* if

(i) $[p]_F(X) = X^q$ and

(ii) $F(X,Y) = X + Y + cC_q(X,Y) \bmod \deg(q+1)$,

(the division by p in necessary to compute $C_q(X,Y)$ is a formality, i.e., first expand $B_q(X,Y)$ and take out one power of p from each coefficient; the resulting polynomial is $C_q(X,Y)$). Every formal group law over k is isomorphic to another such law in normal form (cf. [Froh, Lemma 5,§ III.2]); this proposition makes use of the separably closed hypothesis imposed on k.

Let $\mathcal{F}$ denote the set of formal group laws in normal form of height h over k, and let M be the $\mathbb{F}_p$-vector space of polynomials of the form

$$a(X) = \sum_{i=0}^{h-1} a_i X^{p^i}.$$

It is clear M has dimension h^2 over $\mathbb{F}_p$. We can make M into a ring with unit by defining multiplication through composition and then moding out by $\deg q$.

Lemma 5.7. *Let F be a formal group law in $\mathcal{F}$. Then the k-linear map*

$$Xk[\![X]\!] \to Xk[X]$$
$$f(X) = \sum_{j=1}^{\infty} f_j X^j \mapsto \bar{f}(X) = \sum_{j=1}^{q-1} f_j X^j$$

is a surjective ring homomorphism $\operatorname{End} F \to M$ *with kernel* $p(\operatorname{End} F)$.

Proof. We must first make sure that the above map makes sense, i.e., that if $f \in \operatorname{End} F$ then $\bar{f} \in M$. Since F is in normal form we know $[p]_F(X) = X^q$, from which we deduce that

$$f(X^q) = f \circ [p]_F = [p]_F \circ f = f(X)^q,$$

so f is defined over $\mathbb{F}_q$. Moreover, we know

$$f \equiv X + Y \bmod \deg q,$$

from which we obtain

$$f(X+Y) \equiv f(F(X+Y)) = F(f(X), f(Y)) \equiv f(X) + f(Y) \bmod \deg q.$$

Hence $f \bmod \deg q$ is a polynomial in X^p, which means $\bar{f} \in M$.

Next, we briefly show the map is a ring homomorphism. Compatibility of the map with multiplication (composition) is straightforward, and $\bar{x} = x$. To see compatibility with addition note that

$$(f+g)(X) = F(f(X), g(X)) \equiv f(X) + g(X) \bmod \deg q$$

because F is in normal form. Since $\deg \bar{f}, \deg \bar{g} < q$, we deduce that

$$\overline{(f+g)(X)} = \overline{f(X) + g(X)} = \bar{f}(X) + \bar{g}(X).$$

Now we prove $\operatorname{End} F \to M$ is surjective. The polynomials $a(X) \in M$ whose last coefficient a_0 is nonzero generate M as an additive group. Since the above map is a homomorphism of rings, it is enough to show each such a has a preimage $f \in \operatorname{End} F$. The idea is to construct f using the completeness of $\mathbb{F}_q[\![X]\!]$. Explicitly, we construct a sequence $\{f_n\}$ of invertible power series ($n \geq q$) such that

(i) $f_q = a$,

(ii) $f_n \circ F \equiv F \circ f_n \bmod \deg n$,

(iii) $f_{n+1} \equiv f_n \bmod \deg n$.

Suppose we have constructed f_m. Set $G = f_m^{-1} \circ F \circ f_m$. By property (ii) above $F \equiv G \bmod \deg m$, so by Theorem 5.6 there exists a constant $c \in F_q$ such that

$$F \equiv G + cC_m \bmod \deg(m+1).$$

If $m \neq p^r$ then $cC_m = bB_m$ for some $b \in F_q$. Otherwise $cC_m = 0$. In any case

$$F \equiv G + bB_m \bmod \deg(m+1) \quad \text{for some } b \in \mathbb{F}_q.$$

We claim there is an invertible power series $g(X)$ such that

$$g(X) \equiv X \bmod \deg m \quad \text{and} \quad g \circ F \circ g^{-1} \equiv G \bmod \deg(m+1).$$

Indeed, $g(X) \equiv X - bX^m \bmod \deg(m+1)$ will do the trick. Explicitly,

$$\begin{aligned} g(F(X,Y)) &\equiv F(X,Y) - b(X+Y)^m \bmod \deg(m+1) \\ &\equiv G(X,Y) + b(X+Y)^m - bX^m - bY^m - b(X+Y)^m \bmod \deg(m+1) \\ &\equiv G(X,Y) - bX^m - bY^m \bmod \deg(m+1) \\ &\equiv G(g(X), g(Y)) \bmod \deg(m+1). \end{aligned}$$

Now set $f_{m+1} = f_m \circ g$. This power series satisfies conditions (ii) and (iii). Now put $f = \lim_{n\to\infty} f_n$ (the limit exists by completeness of $\mathbb{F}_q[\![X]\!]$). By construction $\bar{f} = a$ and $f \in \operatorname{End} F$. This concludes the proof of surjectivity.

Finally we look at the kernel of the map. We claim that

$$p^n \operatorname{End} F = \{f \in \operatorname{End} F \mid \operatorname{ht}(f) \geq nh\}. \tag{5.11}$$

On the one hand, if $f = [p]_F^n \circ g$ then

$$\operatorname{ht}(f) = n \operatorname{ht}([p]_F) + \operatorname{ht}(g) \geq nh.$$

On the other hand, if $f \in \operatorname{End} F$ and $\operatorname{ht}(f) \geq nh$ then there is a power series $g(X)$ with $f(X) = g(X^{q^n})$ by definition of height (recall $q = p^h$). Hence $f = g \circ [p]_F^n$; if we can show $g \in \operatorname{End} F$ we are in good shape. By Remark 5.8 below, the formal group law F is defined over $\mathbb{F}_q$, so

$$f(F(X,Y)) = g(F(X,Y)^{q^n}) = g(F(X^{q^n}, Y^{q^n})),$$

while

$$F(f(X), f(Y)) = F(g(X^{q^n}), Y^{q^n}),$$

from which we conclude that

$$g(F(X,Y)) = F(g(X), g(Y)),$$

which is to say that g is an endomorphism of F, as required. It is now clear $p(\operatorname{End} F)$ is the kernel of the map $\operatorname{End} F \to M$. □

Remark 5.8. We have shown in the proof above that a formal group law F over k in normal form has all its endomorphims defined over $\mathbb{F}_q$. Since $[p]_F \circ F = F \circ [p]_F$ and $[p]_F = X^q$, the formal group law itself is also defined over $\mathbb{F}_q$.

We are now in a position to prove the main theorem of this section.

Theorem 5.9 (Lubin). *Let F be a formal group law of height h over a separably closed field k. Then $\operatorname{End} F$ is a free $\mathbb{Z}_p$-module of rank h^2. Furthermore, it is the maximal order in the local division algebra $D = \operatorname{End} F \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.*

Proof. We already know $\operatorname{End} F$ is a $\mathbb{Z}_p$-module without zero divisors, i.e., it is a torsion-free $\mathbb{Z}_p$-module. To show this module has rank h^2 note that Lemma 5.7 gives an isomorphism of $\mathbb{F}_p$-vector spaces $(\operatorname{End} F)/p(\operatorname{End} F) \cong M$. Hence $(\operatorname{End} F)/p(\operatorname{End} F)$ has dimension h^2 over $\mathbb{F}_p$ and by Nakayama's lemma $\operatorname{End} F$ has rank h^2 as a $\mathbb{Z}_p$-module (cf. [A–M, Proposition 2.8]).

It follows $D = \operatorname{End} F \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a local division $\mathbb{Q}_p$-algebra of dimension h^2. It remains to show $\operatorname{End} F$ is the maximal order in this algebra (note the maximal order is unique as we are working over a local field). Recall the p-adic valuation v of $\mathbb{Q}_p$ extends to a unique valuation of D, which we also denote v, and the set

$$\mathcal{O} = \{x \in D \mid v(x) \geq 0\}$$

is the maximal order of D (cf. §4.3.2–the remarks of this section apply to general local division algebras). The inclusion $\operatorname{End} F \subset \mathcal{O}$ is clear. To see the other inclusion first note that the map $f \mapsto h^{-1} \operatorname{ht} f$ is a valuation of $\operatorname{End} F$ that coincides with the usual p-adic valuation when restricted to $\mathbb{Z}_p$ (when we write $f \in Z_p$ we mean $f = a \cdot [1]_F$ for some $a \in Z_p$). By uniqueness of the extension of a valuation to D we obtain

$$\operatorname{ht} f = h \cdot v(f), \quad f \in \operatorname{End} F. \tag{5.12}$$

If $f \in \mathcal{O}$ then $p^n f \in \operatorname{End} F$ for some $n \geq 0$ because $\operatorname{End} F$ spans the algebra D over $\mathbb{Q}_p$. We know $v(p^n f) \geq n$, so from (5.12) we deduce that $\operatorname{ht} p^n f \geq nh$, whence $p^n f \in p^n \operatorname{End} F$ by (5.11). This means $f \in \operatorname{End} F$ and so $\mathcal{O} \subset \operatorname{End} F$. □

5.5 The Endomorphism Ring of a Supersingular Elliptic Curve

We are almost ready to prove the endomorphism ring of a supersingular elliptic curve is a *maximal* order in a quaternion algebra.

Let E be an curve defined over a finite field k and let l be a prime number different from $\operatorname{char} k = p$. Tate's Theorem (2.16) gives an isomorphism

$$\operatorname{End}_{\mathcal{G}} E \otimes \mathbb{Z}_l \xrightarrow{\sim} \operatorname{End}_{\mathcal{G}} T_l(E), \tag{5.13}$$

where $\operatorname{End}_{\mathcal{G}} E$ is the subring of endomorphims of E that commute with the action of $\mathcal{G} = \operatorname{Gal}(\bar{k}/k)$, similarly for $\operatorname{End}_{\mathcal{G}} T_l(E)$. This isomorphism is crucial to our proof of the maximality of $\operatorname{End} E$.

There is a p-adic analogue of the l-adic Tate module. It is called a Dieudonné module and is denoted $T_p(E)$ to emphasize the analogy. Tate proved that over a finite field k there is an isomorphism

$$\operatorname{End}_{\mathcal{G}} E \otimes \mathbb{Z}_p \xrightarrow{\sim} \operatorname{End}_{\mathcal{G}} T_p(E), \tag{5.14}$$

where as usual $\mathcal{G} = Gal(\bar{k}/k)$. When an elliptic curve E is supersingular, the Dieudonné module can be identified with the formal group F of the curve. We will not even attempt to define what a Dieudonné module is. This task, as well as a proof of both of Tate's isomorphisms, is quite beyond the level of this thesis. The interested reader may consult [W–M]. What is important is that the reader be aware that it is possible to identify $\operatorname{End}_{\mathcal{G}} E \otimes \mathbb{Z}_p$ with $\operatorname{End}_{\mathcal{G}} F$.

Theorem 5.10. *Let E be a supersingular elliptic curve over a separable closure k of $\mathbb{F}_p$. Then* $\operatorname{End} E$ *is a* maximal *order in a rational quaternion algebra ramified at p and ∞.*

Proof. We know $\operatorname{End} E$ is an order in $B_{\{p,\infty\}}$; it remains to show it is maximal. It is enough to prove maximality of the order locally (cf. Lemma 4.20). As a preliminary remark, we note that since k is a separable closure of a perfect field it is also an algebraic closure of $\mathbb{F}_p$.

First, we consider a prime $l \neq p$. We saw the j-invariant of E satisfies $j(E) = j(E)^{p^2}$, so E descends to a curve over the field $\mathbb{F}_{p^2}$. The group $\operatorname{Gal}(\overline{\mathbb{F}}_{p^2}/\mathbb{F}_{p^2}) = \operatorname{Gal}(k/\mathbb{F}_{p^2})$ is generated by the square of the Frobenius map $\phi : E \to E^{(p)}$ (denoted ϕ_2), which on E is an automorphism of the curve composed with multiplication by p. By the definition of isomorphism of two elliptic curves, it is not hard to see that automorphisms can only have finite order equal to 1, 2, 3, 4 or 6. This means that over an extension $\mathbb{F}$ of $\mathbb{F}_{p^2}$ of degree equal to the order of the pertinent automorphism, the Frobenius map is a power of p. Hence the group $\operatorname{Gal}(k/\mathbb{F})$ acts on $T_l(E/\mathbb{F})$ by scalars. Since

$$T_l(E/\mathbb{F}) \cong \mathbb{Z}_l \times \mathbb{Z}_l \quad \text{whenever } l \neq p,$$

it follows that

$$\operatorname{End}_{\operatorname{Gal}(k/\mathbb{F})} T_l(E/\mathbb{F}) \cong M_2(\mathbb{Z}_l)$$

because every element of $\operatorname{End} T_l(E/\mathbb{F})$ commutes with scalar matrices. As we are now working over a finite field, Tate's isomorphism implies

$$\operatorname{End}_{\operatorname{Gal}(k/\mathbb{F})} E/\mathbb{F} \otimes \mathbb{Z}_l \cong M_2(\mathbb{Z}_l),$$

which is to say $\operatorname{End}_{\operatorname{Gal}(k/\mathbb{F})} E/\mathbb{F} \otimes \mathbb{Z}_l$ is a maximal order in the quaternion algebra $\operatorname{End}_{\operatorname{Gal}(k/\mathbb{F})} E/\mathbb{F} \otimes_{\mathbb{Q}} \mathbb{Q}_l$.

Any endomorphism of E is defined over some finite field $\mathbb{F}'$, which we assume contains $\mathbb{F}$. Since $\overline{\mathbb{F}'} = k$, the group $\operatorname{Gal}(\overline{\mathbb{F}'}/\mathbb{F}')$ is contained in the group $\operatorname{Gal}(k/\mathbb{F})$; consequently the former Galois group also acts by scalars on the Tate module $T_l(E)$ and hence $\operatorname{End}_{\operatorname{Gal}(\overline{\mathbb{F}'}/\mathbb{F}')} E \otimes \mathbb{Z}_l \cong M_2(\mathbb{Z}_l)$ as before. As this is true for any endomorphism, it follows that $\operatorname{End} E \otimes \mathbb{Z}_l \cong M_2(Z_l)$, i.e., $\operatorname{End} E \otimes \mathbb{Z}_l$ is a maximal order in the quaternion algebra $\operatorname{End} E \otimes \mathbb{Q}_l$.

Next, to show $\operatorname{End} E \otimes \mathbb{Z}_p$ is a maximal order in the local division algebra $D = \operatorname{End} E \otimes \mathbb{Q}_p$ we use our work on formal group laws. Let F be the formal group law associated to the curve E. Since E is supersingular we know F has height 2, and consequently $\operatorname{End} F$ is a $\mathbb{Z}_p$-module of rank 4 and is the maximal order $\mathcal{O}$ in the local (quaternion) division algebra $\operatorname{End} F \otimes \mathbb{Q}_p$ (cf. Theorem 5.9).

As before, any endomorphism of E is defined over some finite field $\mathbb{F}$. The isomorphim 5.14 together with the identification of $T_p(E)$ with F in the supersingular case tell us that

$$\operatorname{End}_{\operatorname{Gal} k/\mathbb{F}} E/\mathbb{F} \otimes \mathbb{Z}_p \cong \mathcal{O}.$$

Since this is true for any endomorphism, $\operatorname{End} E \otimes \mathbb{Z}_p$ is the maximal order of D. □

There are other known proofs of this result, due to Deuring [Deu, §2.4] and Waterhouse [Wat, Theorem 4.2]. Cornut found a proof recently [Cor, Proposition 2.1] which is unfortunately beyond the level of this paper.

Chapter 6

Reduction of CM–Elliptic Curves and Gross–Zagier Numbers

In this Chapter we will study the reduction of CM–Elliptic curves and the behavior of their endomorphism rings under reduction. As an application, we will explain the high divisibility of Gross–Zagier Numbers. As in Chapter 3, the curves we consider will have complex multiplication by the ring of integers of an imaginary quadratic field. The exposition of the theory of reduction is influenced by [Sil 2] and [Lang, Ch. 13].

Remark. In our study of elliptic curves in Chapter 2 we assumed the field we worked over had characteristic different from 2 or 3. We did this for pedagogical reasons, as this assumption greatly simplified the presentation of the material without sacrificing the generality of all the theorems presented therein. This approach backfires, however, when one considers reduction mod 2 or 3. The reader interested in seeing how the material from Chapter 2 applies to fields with these two characteristics should consult [Sil 1, Ch. III and Appendix A].

6.1 Good Reduction of Elliptic Curves over Number Fields

We begin with the concept of reduction at a prime $\mathfrak{P}$ of a local field L.

Let L be a local field, $\mathcal{O}_L$ its ring of integers and let $\mathfrak{P}$ be a prime in $\mathcal{O}_L$. Given an elliptic curve E/L with equation

$$y^2 = 4x^3 - g_2 x - g_3,$$

if we replace (x, y) by $(u^{-2}x, u^{-3}y)$, we find $g_2 \mapsto u^4 g_2$ and $g_3 \mapsto u^6 g_3$, so if we take u divisible by a large power of $\mathfrak{P}$ we obtain an equation for E such that $v_{\mathfrak{P}}(\Delta(E)) \geq 0$, where $v_{\mathfrak{P}}$ denotes the usual valuation, and Δ is the discriminant of E (cf. §2.1). Among all such equations, there is at least one that minimizes the quantity $v_{\mathfrak{P}}(\Delta(E)) \geq 0$. We call this a *minimal equation* for E at $\mathfrak{P}$. It is unique up to an isomorphism $(x, y) \mapsto (u^{-2}x, u^{-3}y)$ with $u \in \mathcal{O}_L$.

Given a minimal equation for an elliptic curve E at $\mathfrak{P}$, we can reduce its coefficients modulo $\mathfrak{P}$ to obtain a possibly singular curve with coefficients in the field $k := \mathcal{O}_L/\mathfrak{P}\mathcal{O}_L$,

$$\widetilde{E} : y^2 = \tilde{4}x^3 - \widetilde{g_2}x - \widetilde{g_3}.$$

We say $\widetilde{E}/k$ has *good reduction* if it is nonsingular.

Now let K be a number field, and let $v_{\mathfrak{P}}$ be the discrete valuation of K associated to the prime $\mathfrak{P} \in \mathcal{O}_K$. We say that an elliptic curve E/K has good reduction at $\mathfrak{P}$ if E has good reduction when considered as a curve over the completion $K_{v_{\mathfrak{P}}}$ (a local field).

The following lemma, whose proof can be found in [Sil 1, Proposition VIII.1.4] or in [Lang, §13.4], will provide a key ingredient in showing that isogenies are well behaved under reduction.

Lemma 6.1. *Let E/K be an elliptic curve over a number field K, and let l be a rational prime relatively prime to the characteristic of k_v (the residue field of the local field K_v). If $\widetilde{E}/k_v$ is nonsingular then*

$$(E/K)[l^n] \cong (\widetilde{E}/k_v)[l^n] \quad \textit{for all } n \in \mathbb{Z}_{\geq 1}. \qquad \square$$

By the definition of the Tate Module (cf. §2.5) it follows that, under the conditions of the Lemma, $T_l(E) \cong T_l(\widetilde{E})$.

Theorem 6.2. *Let E_1/K and E_2/K be two elliptic curves over a number field K with good reduction at a prime ideal $\mathfrak{P} \in \mathcal{O}_K$. Denote $\widetilde{E_1}$ and $\widetilde{E_2}$ their reductions at $\mathfrak{P}$, respectively. Then the natural reduction map*

$$\begin{aligned} \operatorname{Hom}(E_1, E_2) &\longrightarrow \operatorname{Hom}(\widetilde{E_1}, \widetilde{E_2}) \\ \phi &\longmapsto \widetilde{\phi} \end{aligned}$$

is injective, and $\deg(\phi) = \deg \widetilde{\phi}$. □

Proof. We follow the approach in [Sil 2, Proposition II.4.4]. The injectivity of the map above follows from the equality of degrees, since the only degree zero map is the trivial one.

To show the equality of degrees we will use the definition and properties of the Weil pairing (cf. Theorem 2.19). Let l be a rational prime relatively prime to $\mathfrak{P}$. We will prove that

$$e_{\widetilde{E_1}}(\tilde{x}, \tilde{y})^{\deg \phi} = e_{\widetilde{E_1}}(\tilde{x}, \tilde{y})^{\deg \widetilde{\phi}} \quad \text{for all } \tilde{x}, \tilde{y} \in T_l(\widetilde{E_1}). \tag{6.1}$$

Since the Weil pairing is nondegenerate, the desired equality of degrees follows from (6.1).

Let E/K be a curve with good reduction at $\mathfrak{P}$. One checks using the definition of the pairing $e_E : T_l(E) \times T_l(E) \to T_l(\mu)$ that

$$\widetilde{e_E(x, y)} = e_{\widetilde{E}}(\tilde{x}, \tilde{y}) \quad \text{for all } x, y \in T_l(E).$$

Let $x, y \in T_l(E_1)$. Then

$$e_{E_1}(x, y)^{\deg \phi} = e_{E_1}((\deg \phi)x, y) = e_{E_1}(\hat{\phi}\phi x, y) = e_{E_2}(\phi x, \phi y),$$

where the last equality follows from the adjointness of $\phi : E_1 \to E_2$ and $\hat{\phi} : E_1 \to E_2$ for the pairing (cf. Theorem 2.20). Analogously, we show that

$$e_{\widetilde{E_1}}(\tilde{x}, \tilde{y})^{\deg \widetilde{\phi}} = e_{\widetilde{E_2}}(\widetilde{\phi}\tilde{x}, \widetilde{\phi}\tilde{y}).$$

Hence, for any $x, y \in T_l(E)$

$$\begin{aligned} e_{\widetilde{E_1}}(\tilde{x}, \tilde{y})^{\deg \phi} &= \widetilde{e_{E_1}(x, y)}^{\deg \phi} = \widetilde{e_{E_2}(\phi x, \phi y)} \\ &= e_{\widetilde{E_2}}(\widetilde{\phi}\tilde{x}, \widetilde{\phi}\tilde{y}) = e_{\widetilde{E_1}}(\tilde{x}, \tilde{y})^{\deg \widetilde{\phi}}. \end{aligned}$$

Since the above holds for all $x, y \in T_l(E)$, it also is true for any $\tilde{x}, \tilde{y} \in T_l(\widetilde{E_1})$. This proves (6.1). □

An immediate corollary of this theorem is that the endomorphism ring of an elliptic curve E over a number field K injects into the endomorphism ring of the reduction $\widetilde{E}/k_v$ at some prime $\mathfrak{P} \in \mathcal{O}_K$. In particular, since an elliptic curve $E/\mathbb{C}$ with complex multiplication by $\mathcal{O}_K$ is defined over the number field $\mathbb{Q}(j(E))$ (cf. Theorem 3.7) (and consequently the endomorphisms of E are defined over the compositum $L = K\mathbb{Q}(j(E))$—cf. Theorem 3.7 given a prime $\mathfrak{P}$ in $\mathcal{O}_L$ we obtain an injection

$$\operatorname{End} E \hookrightarrow \operatorname{End} \widetilde{E} \tag{6.2}$$

where $\widetilde{E}$ is the reduction of E at $\mathfrak{P}$.

The above injection need not be a surjection. Indeed, there are CM elliptic curves over $\mathbb{C}$ whose reduced endomorphism ring is an order in a quaternion algebra (note this is possible because the reduced curve is defined over a field of nonzero characteristic). Such curves are said to have *supersingular reduction* at the residue field characteristic p. Otherwise we say the curve has *ordinary reduction.*

Theorem 6.3. *Let E/L be an elliptic curve over a number field L. Assume E has complex multiplication by the ring of integers of a quadratic imaginary field K, and that E has good reduction $\widetilde{E}$ at a prime $\mathfrak{P} \in \mathcal{O}_L$ which lies over the rational prime p. Then $\widetilde{E}$ is supersingular if and only if p ramifies or remains inert in K.*

Proof. Suppose p splits completely in K as $p\mathcal{O}_K = \mathfrak{p}\mathfrak{p}'$. To show E has ordinary reduction at $\mathfrak{P}$ we prove $\widetilde{E}[p] \neq O$ (cf. Theorem 5.1(iv)). Let

$$\Phi : K \hookrightarrow \operatorname{End} E \otimes \mathbb{Q}$$

be the extension by scalars of normalized embedding[1] $\mathcal{O}_K \hookrightarrow \operatorname{End} E$. Let m be a positive integer such that $\mathfrak{p}^m$ and $\mathfrak{p}'^m$ are principal ideals, say $\mathfrak{p}^m = \alpha\mathcal{O}_K$ and $\mathfrak{p}'^m = \beta\mathcal{O}_K$ (hence $\alpha\beta = p^m$). Since Φ is normalized and $\alpha \notin \mathfrak{p}'$ it follows that

$$\Phi^*(\alpha)\omega = \alpha\omega \neq 0 \bmod \mathfrak{p}'$$

for $\omega \in \Omega_E$. This means $\widetilde{\Phi(\alpha)}$ is a separable map (cf. Theorem 2.5). We know from Theorem 6.2 that

$$\deg(\widetilde{\Phi(\alpha)}) = \deg(\Phi(\alpha)) = p^r.$$

Since $\widetilde{\Phi(\alpha)}$ is a separable map it follows that

$$\#\ker(\widetilde{\Phi(\alpha)}) = \deg(\widetilde{\Phi(\alpha)}) = p^r.$$

We conclude that $\widetilde{E}$ has a torsion point of order p—any nontrivial element of order p in the group $\ker(\widetilde{\Phi(\alpha)})$ will do. Hence $\widetilde{E}[p] \neq O$, as desired.

We will not prove the reverse direction, as a complete exposition of the required machinery necessitates more room than we have to develop (in any case, we will not use this part of the theorem in the rest of the paper). The interested reader is referred to [Lang, Theorem 13.4.12] for a proof. □

[1] We defined the normalized embedding for an elliptic curve over $\mathbb{C}$. To see how one defines the normalized embedding over a number field L see [Lang, §9.4]

6.2 Primes Dividing Gross–Zagier Numbers

We defined Gross–Zagier numbers in §3.7; let us recall the definition.

Let K and K' be quadratic imaginary fields, and let $\{E_1, \dots, E_{h_1}\}$ and $\{E'_1, \dots, E'_{h'}\}$ be sets of representatives for $\mathfrak{E}(\mathcal{O}_{K_1})$ and $\mathfrak{E}(\mathcal{O}_{K_2})$, respectively (remember $\mathfrak{E}(\mathcal{O})$ is the set of elliptic curve with complex multiplication by *Oint* up to $\overline{\mathbb{Q}}$-isomorphism). Then the norm

$$N(j_K - j'_K) = \prod_{m=1}^{h} \prod_{n=1}^{h'} \big(j(E_i) - j(E'_j)\big)$$

is a Gross–Zagier number.

We now give a bound for the primes that divide $N(j_K - j'_K)$. Let D and D' be the discriminants of K and K', respectively. We will assume D and D' are relatively prime. Let p be a rational prime and suppose $p \big| N(j_K - j'_K)$. Then

$$j(E_m) - j(E'_n) \equiv 0 \bmod \mathfrak{P}$$

for some prime $\mathfrak{P}$ of the number field $\mathbb{Q}(j(E_m), j(E'_n))$ that lies over p (note both E_m and E'_n are defined over this number field). This means the reductions $\widetilde{E_m}$ and $\widetilde{E'_n}$ at $\mathfrak{P}$ are isomorphic over the algebraic closure of the field with p elements. Let $\widetilde{E}$ denote a curve in the same isomorphism class. Then by Theorem 6.2 and the discussion that follows it we obtain injections[2]

$$\begin{tikzcd}
\mathcal{O}_K \cong \operatorname{End} E_m \arrow[dr] & \\
& \operatorname{End} \widetilde{E} \\
\mathcal{O}_{K'} \cong \operatorname{End} E'_n \arrow[ur] &
\end{tikzcd} \tag{6.3}$$

However, this can only happen if $\operatorname{End} \widetilde{E}$ is a maximal order in the quaternion algebra $B_{\{p,\infty\}}$. Otherwise, $\operatorname{End} \widetilde{E}$ would be an order $\mathcal{O}$ in some quadratic imaginary field K'' and then passing to the fraction fields in (6.3) would give injections

$$K \hookrightarrow K'' \quad \text{and} \quad K' \hookrightarrow K''$$

Since each of these fields is a two-dimensional vector space over $\mathbb{Q}$, the above injections are actually isomorphisms, in which case we conclude $K \cong K'$, which is absurd.

Theorem 6.4 (Gross–Zagier). *Let D and D' be relatively prime discriminants corresponding to the imaginary quadratic fields K and K'. Let p be a rational prime dividing $N(j_K - j_{K'})$. Then $p \le DD'/4$.*

Proof. Our considerations so far show that $\mathcal{O}_K$ and $\mathcal{O}_{K'}$ inject into a *maximal* order R of a rational quaternion algebra $B_{\{p,\infty\}}$ ramified at p and infinity (cf. Theorem 5.10). It is well known that

$$\mathcal{O}_K = \mathbb{Z} + \left(\frac{D + \sqrt{D}}{2}\right)\mathbb{Z} \quad \text{and} \quad \mathcal{O}_{K'} = \mathbb{Z} + \left(\frac{D' + \sqrt{D'}}{2}\right)\mathbb{Z},$$

[2] We assume the curves E_m and E'_n have good reduction at $\mathfrak{P}$. This is possible by a Theorem of Serre and Tate which goes beyond the scope of this paper (cf. [S–T]).

see for example [F–T, II.1.33]. It follows that the order

$$S = \mathbb{Z} + \left(\frac{D+\sqrt{D}}{2}\right)\mathbb{Z} + \left(\frac{D'+\sqrt{D'}}{2}\right)\mathbb{Z} + \left(\frac{D+\sqrt{D}}{2}\cdot\frac{D'+\sqrt{D'}}{2}\right)\mathbb{Z}$$

is contained in R. By Corollary 4.19 we get the containment $\operatorname{disc}(S) \subset \operatorname{disc}(R)$. We use Theorem 4.18 to compute the reduced discriminant of S. Let x denote the reduced trace of $\sqrt{DD'}$. Then

$$t\left(\frac{D+\sqrt{D}}{2}\right) = D, \quad t\left(\frac{D'+\sqrt{D'}}{2}\right) = D', \quad t\left(\frac{D+\sqrt{D}}{2}\cdot\frac{D'+\sqrt{D'}}{2}\right) = \frac{DD'+x}{2},$$

from which it follows that

$$\operatorname{disc}^2(S) = \begin{vmatrix} 2 & D & D' & (DD'+x)/2 \\ D & D(D+1)/2 & (DD'+x)/2 & 0 \\ D' & (DD'+x)/2 & D'(D'+1)/2 & 0 \\ (DD'+x)/2 & 0 & 0 & X \end{vmatrix}\mathbb{Z}$$

where $X = (DD'+D)(DD'+D')/8 + DD'x/2$. Expanding the determinant and taking square roots we obtain

$$\operatorname{disc}(S) = \left(\frac{DD'-x^2}{4}\right)\mathbb{Z}.$$

Since R is a maximal order Theorem 4.22 gives $\operatorname{disc}(R) = p\mathbb{Z}$. Since $\operatorname{disc}(S) \subset \operatorname{disc}(R)$ we conclude that

$$p \left| \frac{DD'-x^2}{4} \right. \quad \Longrightarrow \quad p \le \frac{DD'}{4}. \qquad \square$$

Corollary 6.5. *Let K be an imaginary quadratic field of discriminant D. If $3 \nmid D$ and p is a prime dividing $N(j_K)$ then $p \le 3|D|/4$.*

Proof. Let $K' = \mathbb{Q}(\sqrt{-3})$. Then $C(\mathcal{O}_{K'})$ consists of one element and $j_{K'} = j((1+\sqrt{-3})/2) = 0$, hence

$$N(j_K - j_{K'}) = N(j_K - 0) = N(j_K).$$

Suppose p divides $N(j_K)$. Since $3 \nmid D$, Theorem 6.4 implies $p \le -3 \cdot D/4 = 3|D|/4$, as claimed. $\square$

Example 6.1. Let $K = \mathbb{Q}(\sqrt{(}-133))$. We know from §3.3 that

$$N(j_K) = -(2^8 \cdot 3^4 \cdot 5^4 \cdot 11^2 \cdot 23^2 \cdot 29^2 \cdot 383)^3.$$

Theorem 6.4 gives the bound $p \le 399$ for a prime dividing $N(j_K)$. Note how good the bound is (recall $N(j_K)$ is a 53-digit number in this case).

Example 6.2. Let $K = \mathbb{Q}(\sqrt{(}-7))$ and $K' = \mathbb{Q}(\sqrt{-19})$. These imaginary quadratic fields both have class number 1, and so the singular moduli j_K and $j_{K'}$ are already Gross–Zagier numbers, as is their difference, which is

$$j\left(\frac{1+\sqrt{-7}}{2}\right) - j\left(\frac{1+\sqrt{-19}}{2}\right) = -3^3 \cdot 5^3 + 2^{15} \cdot 3^3 = 3^7 \cdot 13 \cdot 31.$$

Theorem 6.4 gives the bound $p < 34$.

6.3 Embeddings into the Rational Quaternion Algebras $B_{\{p,\infty\}}$

Let K_1 and K_2 be imaginary quadratic fields with discriminants D_1 and D_2, which we assume are relatively prime. In the course of proving Theorem 6.4 we showed that if p divides $N(j_{K_1} - j_{K_2})$ then both $\mathcal{O}_{K_1}$ and $\mathcal{O}_{K_2}$ must inject into a maximal order in a rational quaternion algebra ramified at p and infinity. In this section we will show how to embed *one* ring $\mathcal{O}_{K_1}$ into an algebra of the form $B_{\{p,\infty\}}$. Showing how two rings of integers embed into a $B_{\{p,\infty\}}$-algebra is already a very difficult task, let alone showing explicitly an embedding into a maximal order. The purpose is of this computation is two-fold: the embedding provides a kind of 'sanity check' to our earlier work and at the same time it shows how useful Hilbert symbols are.

Suppose p is odd and divides $N(j_{K_1} - j_{K_2})$. Then Theorem 6.3 tells us p is either inert or ramified in both $\mathcal{O}_{K_1}$ and $\mathcal{O}_{K_2}$. We will assume p is an inert prime in $\mathcal{O}_{K_1}$.

Theorem 6.6. *Let K be an imaginary quadratic field of odd discriminant D, and let p be a rational prime which is inert in $\mathcal{O}_K$. Then there exists an embedding $\mathcal{O}_K \hookrightarrow B_{\{p,\infty\}}$.*

Proof. We claim there is a prime q for which $\mathcal{O}_K \hookrightarrow (D, -pq)_{\mathbb{Q}}$. Choose q so that p, q and D are pairwise relatively prime (we know p and D are relatively prime since p is inert in $\mathcal{O}_K$). Since $D < 0$ it is clear that $(D, -pq)_\infty = -1$. Using Theorem 4.9(iv) we compute

$$(D, -pq)_p = \left(\frac{D}{p}\right) = -1,$$

where the last equality follows because p is inert in $\mathcal{O}_K$. The algebra $(D, -pq)_{\mathbb{Q}}$ is ramified at p and infinity. We now impose congruence conditions on q so that it doesn't ramify at any other place.

The only other possible places at which $(D, -pq)_{\mathbb{Q}}$ ramifies are $2, q$ and odd primes l that divide D. The Hilbert symbol at q is

$$(D, -pq)_q = \left(\frac{D}{q}\right).$$

If q splits in $\mathcal{O}_K$ we will obtain $(D, -pq)_q = 1$. For odd primes l dividing D we compute

$$(D, -pq)_l = \left(\frac{-pq}{l}\right).$$

To obtain $(D, -pq)_l = 1$ we must choose q so that

$$\left(\frac{q}{l}\right) = \left(\frac{-p}{l}\right) \quad \text{for all odd } l \text{ dividing } D. \tag{6.3}$$

Each equation in (6.3) gives (at least) one congruence condition for q modulo l. By the Chinese remainder theorem, these conditions can be put together into a single congruence condition for q. Dirichlet's theorem on arithmetic progressions guarantees the existence of at least one prime satisfying this global condition. We have to make sure that q can split in $\mathcal{O}_K$. Jacobi reciprocity tells us that

$$\left(\frac{D}{q}\right)\left(\frac{q}{D}\right) = (-1)^{\frac{(D-1)(q-1)}{4}}$$

Multiplying all the equations in (6.3) we obtain

$$\left(\frac{q}{D}\right) = \left(\frac{-p}{D}\right).$$

Hence, if q splits in $\mathcal{O}_K$ it must be true that

$$(-1)^{\frac{(D-1)(q-1)}{4}} = \left(\frac{D}{q}\right)\left(\frac{q}{D}\right) = \left(\frac{q}{D}\right) = \left(\frac{-p}{D}\right).$$

This gives a congruence condition for q modulo 4. The global condition we had before was a congruence modulo an odd number. A new application of the Chinese remainder theorem and Dirichlet's theorem guarantees the existence of a prime q such that $(D, -pq)_{\mathbb{Q}}$ is unramified at every place except possibly $2, p, \infty$. However, we already know $(D, -pq)_{\mathbb{Q}}$ ramifies at p and infinity, so the Hilbert product law implies $(D, -pq)_{\mathbb{Q}}$ is unramified at 2. Hence $(D, -pq)_{\mathbb{Q}}$ is of the form $B_{\{p,\infty\}}$. The inclusion $\mathcal{O}_K \hookrightarrow (D, -pq)_{\mathbb{Q}}$ then gives the desired injection. □

Bibliography

[A–M] Atiyah, M.F. & Macdonald, I.G. *Introduction to Commutative Algebra*, Addison-Wesley, Reading, MA, 1969.

[Con] Conway, J. H., *The Sensual (Quadratic) Form* AMS, Providence, 1997.

[Cor] Cornut, C., Mazur's conjecture on higher Heegner points *Invent. math.* **148** (2002), 495–523.

[Cox] Cox, D. A. *Primes of the form* $x^2 + ny^2$ John Wiley & Sons, New York, 1989.

[Deu] Deuring, M., Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abh. Math. Sem. Univ. Hamburg* **14** (1941), 197–272.

[Es] Escofier, J.-P., *Galois Theory* Springer, New York, 2001.

[F–T] Fröhlich, A. & Taylor, M. J. *Algebraic Number Theory* Cambridge University Press, Cambridge, 1991.

[Froh] Fröhlich, A. *Formal Groups* LNM **74**, Springer, New York, 1968.

[G–Z] Gross, B. H. & Zagier, D. On Singular Moduli *J. reine angew. Math* **355** (1985), 191–220.

[H–W] Hardy, G. H., & Wright, E. M. *An Introduction to the Theory of Numbers* Clarendon Press, Oxford, 1979.

[Haz] Hazewinkel, M. *Formal Groups and Applications* Academic Press, New York, 1978.

[Hus] D. Husemöller, *Elliptic Curves* Springer, New York, 1987.

[KKS] Kato, K., Kurokawa, N & Saito, T. *Number Theory 1: Fermat's Dream* AMS, Providence, 1996.

[Ked] K. S. Kedlaya, *Complex Multiplication and Explicit Class Field Theory*, Senior Honors Thesis, Harvard University, 1996.

[Lam] Lam, T. Y. *The Algebraic Theory of Quadratic Forms* W. A. Benjamin, Reading MA, 1973.

[Lang] Lang, S. *Elliptic Functions* Springer, New York, 1987.

[Lub] Lubin, J., One parameter formal Lie groups over p-adic integer rings. *Ann. Math.* **80** (1964), 464–484.

[Mar] Marcus, D. A. *Number Fields* Springer, New York, 1977.

[Se 1] Serre, J.-P. *A Course In Arithmetic* Springer, New York, 1985.

[Se 2] Serre, J.-P. Complex multiplication, in J.W.S. Cassels and A. Fröhlich (eds.), *Algebraic Number Theory*, Academic Press, London, 1967.

[S–T] Serre, J.-P. & Tate, J., Good reduction of abelian varieties, *Ann. Math.* **88** (1968), 492–517.

[Shi] Shimura, G. *An Introduction to the Arithmetic Theory of Automorphic Functions* Princeton University Press, Princeton, 1971.

[Sil 1] Silverman, J. *The Arithmetic of Elliptic Curves* Springer, New York, 1986.

[Sil 2] Silverman, J. *Advanced Topics in the Arithmetic of Elliptic Curves* Springer, New York, 1994.

[Tate] Tate, J., Endomorphims of abelian varieties over finite fields *Invent. Math.* **2** (1966), 134–144.

[Vig] Vignéras, M-F. *Arithmétique des Algèbres de Quaternions* LNM **800**, Springer, New York, 1980.

[Wat] Waterhouse W. C. Abelian varieties over finite fields, *Ann. scient. Éc. Norm. Sup.* 4$^{\text{e}}$ série **2** (1969), 521–560.

[W–M] Waterhouse W. C. & Milne, J. S., Abelian varieties over finite fields, *Proc. Symp. Pure Math.* **20** (1971), 53–64.