SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

SeaSide Surfing on High Tide: An overview of CSIDH and its recent improvements in Post-Quantum Cryptography

av

Péter Suhajda

2026 - No M12

SeaSide Surfing on High Tide: An overview of CSIDH and its recent improvements in Post-Quantum Cryptography

Péter Suhajda

Självständigt arbete i matematik 30 högskolepoäng, avancerad nivå

Handledare: Jonas Bergström (SU) & John Preuß Mattsson (Ericsson)

2026

Abstract

As lattice-based schemes dominate standardized Key Encapsulation Mechanisms (KEMs), a viable post-quantum version of a Non-Interactive Key Exchange (NIKE) remains a challenge. This thesis aims to provide a self-contained, rigorous study of the proposed post-quantum NIKE: Commutative Supersingular Isogeny Diffie-Hellman (CSIDH), an isogeny-based cryptographic framework developed to fill this gap. We develop the complete algebraic machinery, including elliptic curve endomorphism rings, orders of imaginary quadratic fields, ideal class groups and supersingular isogeny graphs. We assess the classical and quantum security of CSIDH, and review recent structural improvements to this framework, including CTIDH, SQALE and CSURF.

Sammanfattning

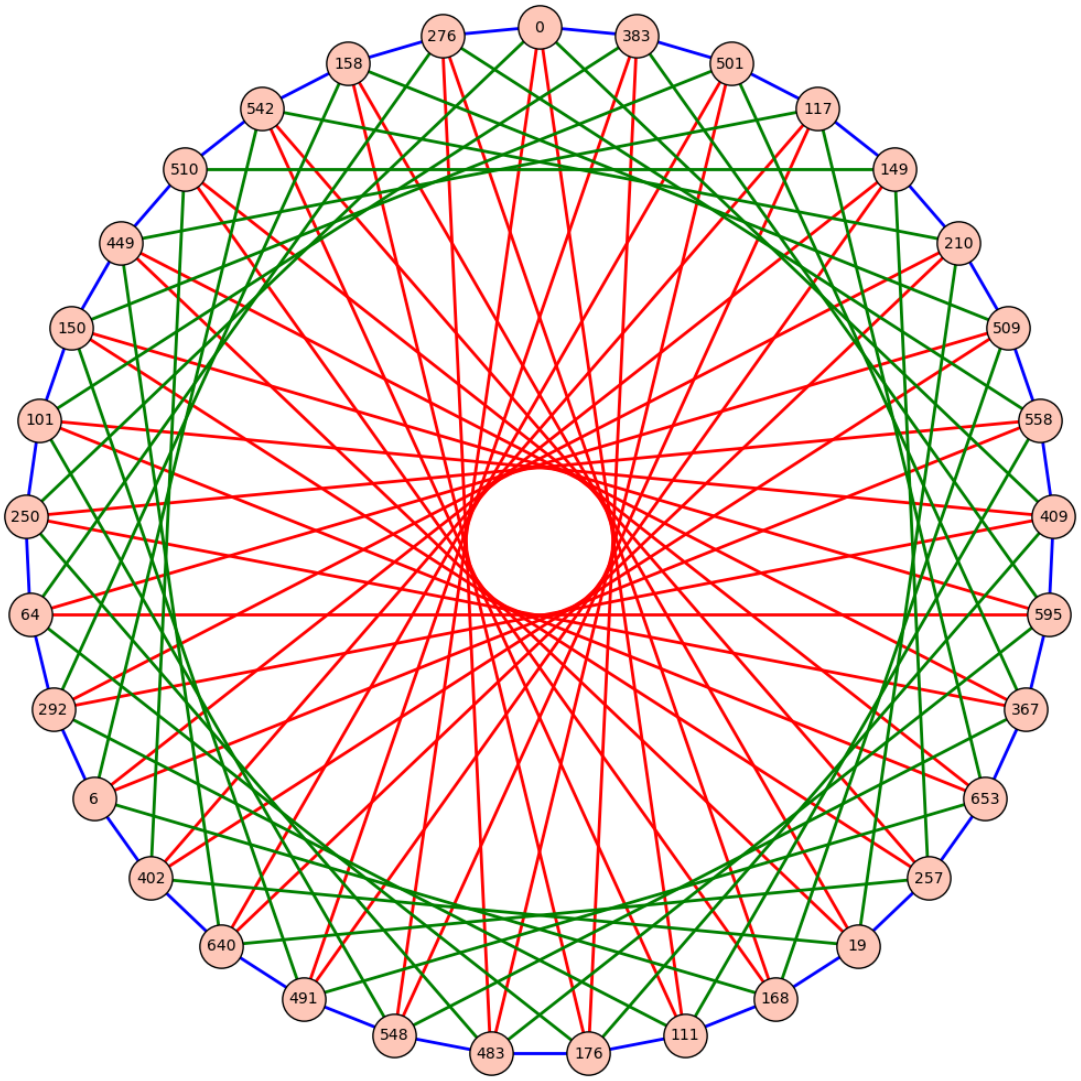
Då gitterbaserade scheman dominerar standardiserade nyckelkapslingsmekanismer (KEM:er), förblir en bärkraftig postkvantversion av ett icke-interaktivt nyckelutbyte (NIKE) en utmaning. Denna avhandling ger en fristående och stringent studie av det föreslagna postkvant-NIKE-protokollet: Kommutativ supersingulär isogeni-Diffie-Hellman (CSIDH), ett isogenibaserat kryptografiskt ramverk utvecklat för att fylla detta tomrum. Vi utvecklar den fullständiga algebraiska apparaten, inklusive endomorfiringar för elliptiska kurvor, ordningar i imaginära kvadratiska talkroppar, idealklassgrupper och supersingulära isogenigrafer. Vi utvärderar den klassiska säkerheten samt kvantsäkerheten för CSIDH, och granskar nyligen gjorda strukturella förbättringar av ramverket, däribland CTIDH, SQALE och CSURF.

Acknowledgments

First of all, I would like to thank my supervisors Jonas Bergström and John Preuß Mattson for guiding me even when I was not there. I would like to thank all the great teachers that I had the luck of meeting. Thank you Greg Arone for the lovely abstract nonsense. Thank you Sjoerd for getting me through the weeds of seminars. Thanks to my brothers and sisters in arms during lectures and seminars. Jesper, you will one day swap my bottle without me noticing. Tack Simon för att du delar bördan av isogenibaserad protokoll.

Of course many thanks go out to my family and friends as well, and also all my volleyball teammates during the years.

Thank you Arlen, I love you.



Contents

1 Introduction **1**
1.1 Structure of the Thesis . . . 1
1.2 Cryptography and the Quantum World . . . 2
1.3 CSIDH vs. SIDH . . . 4

2 Mathematical Background **5**
2.1 Elliptic Curves . . . 5
2.2 Isogenies . . . 12
2.3 Vélu's Formulae . . . 19
2.4 Orders $\mathcal{O}$ and Ideal Class-groups . . . 23
2.5 Endomorphism rings . . . 31
2.6 Isogeny Graphs . . . 41
2.7 Group Actions and Torsors . . . 49

3 The Design and Details of CSIDH **56**
3.1 CSIDH – Letter By Letter . . . 57
3.1.1 DH: Diffie-Hellman . . . 57
3.1.2 I: Isogeny-based . . . 58
3.1.3 S: Supersingular . . . 60
3.1.4 C: Commutative . . . 62
3.1.5 Which prime to choose? . . . 63
3.2 Non-Interactive Key Exchange Protocol . . . 64
3.2.1 Key Exchange Protocol . . . 65
3.2.2 Isogeny calculation . . . 65
3.2.3 Evaluation of the ideal class group action . . . 66
3.2.4 Public Key Validation . . . 66

4 Security of CSIDH **69**
4.1 Classical Security . . . 69
4.1.1 Brute force attack . . . 70
4.1.2 Meet-in-the-middle attack . . . 70
4.1.3 Pohlig-Hellman-style attacks . . . 71
4.2 Quantum Security . . . 71
4.2.1 NIST Security Levels . . . 71
4.2.2 Quantum Information Theory Basics . . . 72
4.2.3 Shor's Algorithm . . . 75
4.2.4 Grover's Algorithm . . . 77
4.2.5 Kuperberg's algorithm and Peikert's attack . . . 79

5 Recent improvements of CSIDH **82**
5.1 SQALE . . . 83
5.2 CTIDH . . . 85
5.3 CSURF . . . 89
5.4 Conclusion – Future Work . . . 92

Bibliography **94**

Chapter 1

Introduction

Perhaps the title of the thesis at first glance is a bit misleading. It is the continuation of the cryptographic legacy of using puns in titles. This thesis will try to make the intimidating symbols of $\mathrm{cl}(\mathcal{O}) \times Ell_p(\mathcal{O}, \pi)$ make perfect sense at a length shy of 100 pages. The topic of this thesis is the post-quantum cryptography protocol, called Commutative Supersingular Isogeny Diffie-Hellman, or in short, CSIDH. The explanation of the title is owed to this abbreviation, since the authors decided to pronounce it as "sea-side", since they were there when developing this framework. The rest of the words in the title refer to the various improvements made from CSIDH, one of them being CSURF, while the other being CTIDH, with the suggested pronunciations.

This thesis is going to be about Post-Quantum Cryptography (PQC), which is not to be confused with Quantum Cryptography. The latter is the study of creating cryptographic algorithms that can be run on quantum computers. PQC, on the other hand is the study of *protecting* cryptography against quantum computers.

Also, CSIDH should not be confused with SIDH either. They are both non-interactive key exchange protocols, but they have many differences, such as the basis of the isogeny graph, the fact that SIDH is non-commutative, and last but not least, SIDH has been broken three times [15, 48, 58].

The imminent arrival of quantum computers does pose a threat even to everyday life, because the vast majority of the Internet uses cryptographic algorithms that are easily breakable with enough qubits. We are going to explore the brief history of CSIDH, learn from the mistakes of the past, and stand on the shoulders of giants.

This thesis assumes basic knowledge of abstract algebra, such as group theory, ring theory and familiarity with a bit of cryptographic algorithms, but otherwise the basic necessary information is provided to help with understanding the design choices that went into the making of CSIDH. Happy surfing!

1.1 Structure of the Thesis

This Master's thesis focuses on one specific candidate for secure Post-Quantum Cryptographic Protocol, and it attempts to provide all the necessary mathematical background to engage with the mechanics of CSIDH (Commutative Supersingular Isogeny Diffie-Hellman). The thesis is divided into 5 chapters, including the introduction, which is aimed to give a layman-friendly general purpose introduction to the most popular notions within cryptography, and highlight the challenges posed by the advent of quantum computing.

Chapter 2 is completely structured around the different mathematical elements of CSIDH that are necessary to explore for a thorough understanding. Throughout this chapter, many theorems and solutions will be tailored to apply to CSIDH in some way, but it also focuses on more general constructions. Section 2.1 starts by introducing what elliptic curves are, and shows basic notions such as the group structure of elliptic curves, and isomorphisms between them. Section 2.2 goes one step further, describing the maps possible between different elliptic curves. Section 2.3 is a slight detour that focuses on different algorithmic ways of actually computing such isogenies, mentioning the classic Vélu's formula [74], but also including other version, such as Joost Renes' Montgomery model formula [59] and the high-level description of the asymptotically faster $\sqrt{é}$lu formula [3]. Section 2.4 goes into algebraic number theory to develop the basic theory of number fields, orders, prime splitting behaviors and ideal class groups necessary to understand the group that is acting on the set of isomorphism classes of elliptic curves. Section 2.5 investigates the self-maps of elliptic curves, which opens up a rich theory that connects it

with the previous section. All the possible $\mathbb{F}_p$-rational endomorphism rings of finite fields $\mathbb{F}_p$ are categorized. Section 2.6 introduces basic notions of graph theory, and explores the isogenies between curves as potential mathematical graphs, investigating its properties. A complete categorization of supersingular isogeny graphs over the finite field $\mathbb{F}_p$ is also included. Section 2.7 gives the last piece of the puzzle, the group action that forms the backbone of the CSIDH action.

Chapter 3 is mainly concerned of the 2018 paper [17] that gives the instantiation and mathematical background for the CSIDH protocol. Section 3.1 is meant to be a historical journey with relevant mathematics, simulating the steps that led to the birth of CSIDH. It highlights the path of researchers' resolute and the important of learning from others "mistakes". In this section, the Diffie-Hellman protocol is explained in several settings, provides visual examples of isogeny graphs generated by the author in SageMath, and explores the brief history of two cryptographic protocols that preceded CSIDH, namely the Couveignes-Rostovtsev-Stolbunov (CRS) protocol [27, 61] and SIDH [30]. Section 3.2 gives the full instantiation of the CSIDH protocol for the prime CSIDH-512, highlighting both the theoretical and low-level implementation decisions made.

Chapter 4 critically examines the validity of the proposed classical and quantum security of CSIDH. Section 4.1 focuses on the classical security, going through classical attacks ranging from the naïve brute force attack until the Meet-in-the-Middle style attacks. Section 4.2 concerns the quantum security of CSIDH together with the recent development that happened after the publication of the original paper. The specific quantum security requirements, as given by NIST, are listed to aid with the comparisons. A short introduction into quantum information theory is also included, to better understand the most known quantum attacks, such as Shor's [67] and Grover's [40] Algorithms. In the last part the biggest threat, Kuperberg's algorithm together with Peikert's hybrid attack is discussed in critically assessed.

Chapter 5 looks further into the isogeny-based cryptographic land after CSIDH. The general structure of the sections in this chapter follows the pattern of mentioning the shortcomings of CSIDH, highlighting the different ways the specific protocol aims the fix/achieve, then the specific instantiation is given with supporting performance tests and comparisons. The three main improvements of CSIDH highlighted are SQALE of CSIDH [20], CTIDH [1], and CSURF [14], but also relevant improvements of these are also mentioned.

1.2 Cryptography and the Quantum World

Cryptography is the science of keeping secrets. There are many reasons for why it is a good idea to keep things secret, and mathematics is just the right tool to do that with the most confidence, especially in the age of the modern world, where everything is being transmitted over the internet (or other insecure channels). Public Key Cryptography generally can be divided into three major branches. The first one is probably the most notable one, which is encryption/decryption. The point of encrypting is obviously to preserve the confidentiality of data, enabled anyone to communicate with another person without revealing the secret message / data to someone else, who might not be authorized to be listening. The second branch is digital signature schemes, which have the purpose of verifying authenticity. There might be a wrongdoer, pretending to be me. Without digital signatures, how would you make sure that it is actually me who you are talking to?

The last branch concern key exchange. In the modern world, if someone is talking to a person on the other side of the globe, they don't have the possibility of sharing a secret in person. Key exchange is trying to solve the problem of allowing two parties to agree on a common secret using public channels, without risking the potential leakage of this shared secret.

These schemes can be categorized into two big parts based on their key structure used as well, not only their purpose.

Definition 1.2.1. We call a cryptographic scheme **symmetric** if the same shared key is used for both encryption and decryption. If they differ, we call the it **asymmetric**

The textbook example for a symmetric encryption is the bitwise XOR operator $\oplus$ on the bits, which is completely reversible by applying it again to a message. Another common examples is AES, which forms the basis of cryptographic security comparisons. (see Section 4.2.1). In the asymmetric case, there

is a distinction between the public key pk and the secret/private key sk, where the public key is used for encryption, and the private key is used for decryption. Beyond that, digital signatures rely on asymmetry as well, since the owner signs a message using their private key, and anyone with the corresponding public key can verify its validity. The most important examples of asymmetric schemes are RSA and ECC or ML-KEM.

RSA, named after its inventors Rivest-Shamir-Adleman is the classical cryptosystem that relies on the hardness of factoring the product of two large prime numbers (semiprimes). It is one of the earliest widely deployed public-key algorithms, and is enjoying worldwide adaptation. It usually generates large keys, up in the 2048-4096 bit range, so it requires a lot of bandwidth compared to other algorithms. It is unfortunately the textbook example of what the quantum computers could break very fast, as the celebrated Shor's quantum algorithm (see Section 4.2.3) break the RSA cryptography system by very fast integer factorization.

ECC, standing for Elliptic Curve Cryptography is the other classical paradigm, where the algebraic structure of elliptic curves are used instead of the integers. Its computational security relies on the Elliptic Curve Discrete Logarithm Problem (ECDLP). It was considered to be a superior version of RSA, since it had much smaller keysizes with comparable security. To name an example, the 256-bit ECC key had the same classical security as the 3072-bit RSA key. It is still used widely across modern network protocols and digital signatures. The major setback is the fact that quantum computing and Shor's algorithm also breaks ECC very fast.

ML-KEM [54] is the newest standardized cryptographic scheme by the National Institute of Standards and Technology (NIST). It was specifically designed to withstand quantum algorithms as well as classical ones, which is based on the fact that the problem it relies on (Module Learning With Errors) is a high-dimensional algebraic lattice problem that are intractable for both quantum or classical algorithms (as of yet). Its key sizes are larger than ECC, but it offers highly efficient encryption and decryption, so it is currently being adopted globally to replace the classical key exchange mechanisms.

The reader must have heard many times that with the advent of quantum computers, most current cryptographic protocols, like RSA or ECC, stand no chance. In the literature, the quantum computers that are big enough to pose a risk are called Cryptographically Relevant Quantum Computers (or CRQC for short). At the time of writing this thesis, RSA-2048 and ECC-256 are estimated to be able to be broken on a CRQC with 13,255 and 11,961 *logical* qubits respectively [9]. However, there are no currently available, large enough quantum computers yet for this purpose – currently the best quantum chips have $\sim$1000 *physical* qubits.

The mathematical foundations of qubits are explained in Section 4.2.2, but it is important to have a general understanding of it. A classical bit is the fundamental unit of digital information, which has two states – 0 or 1. A qubit is the quantum computer counterpart, but it is a quantum mechanical system. It can exist in superpositions of both 0 and 1 and anything inbetween until a measurement collapses it to either 0 or 1. The classic comparison to illustrate the vast array of possibilities is that while n classical bits can encode a single n-bit binary string, an n-qubit system can take advantage of its quantum properties, and live in entangled states or a simultaneous superposition over all 2^n basis states.

However, life is not a bed of roses, and there are many engineering challenges that need to be overcome before CRQCs are available. We have italicized the words *logical* and *physical* qubits two paragraphs ago. The reality of the matter[1] is that when assessing the real threats, one must take into account the nature of noisy physical cubits, and the error-corrected fault-tolerant logical qubits. Current quantum hardware has many problems, one of it is called gate error rates, which essentially prevents the direct execution of quantum circuits. If the qubits are noisy, they obviously cannot be used for calculations, so they must be integrated into quantum error correcting structures. These structures are groupings of 100s or 1000s of physical cubits together to encode a single, fault-tolerant logical qubit. Thus the overall operational scale of CRQC needs to be assess with this ratio of phyisical-to-logical qubits ratio to suppress the error rates.

However, this does not save us from the need to go over to Post-Quantum Cryptography as soon as possible, since adversaries can steal all encrypted data *now*, and decrypt them *later* in the future when those cryptographically-relevant quantum computers are a reality.

[1] Pun intended

1.3 CSIDH vs. SIDH

There is (or was [58]) a non-interactive key exchange protocol with a very similar name, that is SIDH [30]. The team of SIDH with its competing version ´´SIKE" were contestants in the Post-Quantum Cryptography standardization competition held by the National Institute of Standards and Technology (NIST). The name stands for "Supersingular Isogeny Diffie-Hellman", which happens to be the same as CSIDH, modulo C. It is, as the name suggests, also an isogeny-based cryptographic protocol, and it also uses supersingular elliptic curves. This exhausts all the similarities to CSIDH, because there are many structural differences that are orthogonal to each other. SIDH works over the *full* supersingular isogeny graph, which is *non-commutative*. It has a different prime setup, a different method of trespassing the isogeny graph, and it publishes extra kernel points in its public key. The protocol is detailed in more steps in Section 3.1.3, and we will see how CSIDH is completely different.

Another big difference is that SIDH has been broken by classical algorithms, in particular Castryck-Decru [15] published a paper in 2022, detailed that they were able to break the NIST competition version SIKE in an hour on their server, then further papers [48, 58] have also completely invalidated that protocol. This chapter is only here to stress the difference in spite of the similar name, and to mention that the attacks on one protocol do not carry over to the other protocol.

Chapter 2

Mathematical Background

In this chapter, we give the necessarily mathematical background for understanding the foundations and algorithms of CSIDH. Since in the field of post-quantum cryptography, isogeny-based schemes are arguably the most mathematics-heavy schemes, this chapter is designed to be self-contained. This chapter is a collection of important definitions, theorems and proofs from two books, namely Washington [75] and Silverman [69]. The main areas covered are going to be the basic theory of elliptic curves, isogenies between elliptic curves, algebraic number theory to the extent of ideal-class groups and group actions. If the reader feels comfortable in these topics, they may skim this chapter for a quick refresher, or skip altogether and move on to Chapter 3 for the design and workings of CSIDH.

2.1 Elliptic Curves

In its most general form, the definition of an elliptic curve is a *smooth abelian projective variety of dimension 1 with a distinguished point.* Whether this definition says anything is up to the discretion of the reader. However, the definition we use in this paper is the more friendly "set of points satisfying a curve" type. Furthermore, throughout the paper we will assume that if an elliptic curve is defined over a field, then the characteristic of that field is not 2 or 3, unless otherwise stated. Before giving the definition, we have to introduce what is meant by a characteristic, and the notion of projective planes and homogeneous coordinates.

Definition 2.1.1. Let k be a field. The **characteristic** of k, denoted $\mathrm{char}(k)$, is defined as the smallest positive integer n such that for all elements $a \in k$, we have that $na = 0$. If such a number does not exist, we say the field has characteristic 0.

Definition 2.1.2. The **(real) projective plane**, denoted by $\mathbb{P}^2 \in \mathbb{R}^3$, is the set of equivalence classes of $\mathbb{R}^3 \setminus \{\mathbf{0}\}$, where the equivalence $(x_1, x_2, x_3) \sim (y_1, y_2, y_3)$ holds if there is a non-zero element $\lambda \in \mathbb{R}$ such that $(y_1, y_2, y_3) = (\lambda x_1, \lambda x_2, \lambda x_3)$.
The equivalence class is represented by any of its members as $[x : y : z]$, which is called the **homogeneous coordinate**.

Remark 2.1.3. We will call points of the form $[x : y : 1]$ as *affine* points, and when $z = 0$, we call points of the form $[x : y : 0]$ the *points at infinity.*

A homogeneous point $[x : y : z]$ defines an affine point (x', y') by scaling by $1/z$ and setting $[x' : y' : 1] = [x/z : y/z : 1]$. In this paper, we freely interchange between the affine and homogeneous coordinates because the concepts later are more easily expressed in affine coordinates, but calculations themselves are vastly more efficient when done in homogeneous coordinates. This difference in efficiency is discussed further in Remark 2.3.3. We give now the two equivalent definitions of elliptic curves.

Definition 2.1.4 (Affine Elliptic Curves)**.** Let k be a field with $\mathrm{char}(k) \neq 2, 3$. Then all elliptic curves (up to isomorphism) defined over k can be expressed as the set of points satisfying

$$E(k) := \{(x, y) \in k^2 \mid y^2 = x^3 + Ax + B, \quad 4A^3 - 27B^2 \neq 0\} \cup \{\infty\},$$

where $A, B \in k$ and ∞ denotes the "point at infinity".

Remark 2.1.5. The point at infinity is usually denoted by $\mathcal{O}$, but I chose not to use this symbol to avoid clashes of notation. Later on in this paper, we will be using $\mathcal{O}$ extensively to denote the order of an imaginary quadratic field.

Definition 2.1.6 (Homogeneous Elliptic Curves)**.** Let k be a field with $\text{char}(k) \neq 2, 3$. Then all elliptic curves (up to isomorphism) defined over k can be expressed as the set of points satisfying

$$E(k) := \{[x : y : z] \in \mathbb{P}^2(k) \mid y^2z = x^3 + Axz^2 + Bz^3, \quad 4A^3 - 27B^2 \neq 0\},$$

where $A, B \in k$ and $[0 : 1 : 0]$ denotes the "point at infinity".

Remark 2.1.7. These definitions of the elliptic curve is given in the *short Weierstrass form*. One could extend this definition to fields with characteristic $2, 3$ using the *general Weierstrass form* given by

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

There is a third form of elliptic curve that is going to be used in CSIDH and consequent versions that is different from the short Weierstrass form.

Definition 2.1.8. An elliptic curve defined over a field k with $\text{char}(k) \neq 2, 3$ is said to be in **Montgomery form** if it is defined as the set of points satisfying

$$E : y^2 = x^3 + Ax^2 + x,$$

or equivalently, defined as the locus of points in $\mathbb{P}^2$ by

$$E : y^2z = x^3 + Ax^2z + xz^2,$$

for some $A \in k$. This coefficient A is called the **Montgomery coefficient**.

We will denote the set of points of E that are defined over the field k as $E(k)$. For example, if the elliptic curve is defined over the rational numbers, we write $E(\mathbb{Q})$ for their set of points, or if it is defined over a finite field with $q = p^n$ elements, we write $E(\mathbb{F}_q)$. If nothing is specified, or it is clear from the context, we simply write E.

Now we establish a simple lemma that states when a curve in Montgomery form is nonsingular.

Lemma 2.1.9. *An elliptic curve $E : y^2 = x^3 + Ax^2 + x$ in Montgomery form is nonsingular if and only if $A \neq \pm 2$.*

Proof. Any elliptic curve E defined by the equation $y^2 = f(x)$ for some cubic polynomial f is singular if and only if the discriminant Δ of $f(x)$ is zero. On a Montgomery curve, we have that

$$y^2 = f(x) = x^3 + Ax^2 + x = x(x^2 + Ax + 1).$$

Let Δ' denote the discriminant of x^2+Ax+1, then we clearly have that $\Delta' = A^2-4$, and since multiplying $x^2 + Ax + 1$ by x does not affect the discriminant of the whole curve, we have that

$$\Delta = \Delta' = A^2 - 4 = 0 \iff A = \pm 2. \qquad \square$$

Some examples of "good" (nonsingular) and "bad" (singular) elliptic curves are given below in Figures 1 and 2, taken from [69, III.1., pp. 43] with a slight correction of the first discriminant.

The reason for why there is only one point at infinity becomes clear when looking at the homogeneous case. If we set $z = 0$, we get that $0 = x^3 \implies x = 0$, therefore we have the point $[0 : y : 0]$ at infinity, which can then be rescaled by $1/y$, which stays in the same equivalence class, resulting in $[0 : 1 : 0]$ as the distinguished point.

Let $P, Q, R \in E(k)$ denote points on some elliptic curve. Then by Bezout's theorem, if we choose any two points on the elliptic curve $E(k)$, then the line connecting them will intersect the curve at a third point (which might also be at infinity). If we then take this third point, and reflect it along the x-axis, we can define this new point as the "sum" of the two other chosen points. Notationally, we can say that $P \oplus Q = -R$. If we would like to "double" a point, that is compute $[2]P = P \oplus P$, we can take the tangent of the curve at point P, and find the third intersection of that line as before. Figure 3 illustrates these two constructions geometrically, borrowed from Luca De Feo's summer school lecture notes [29].

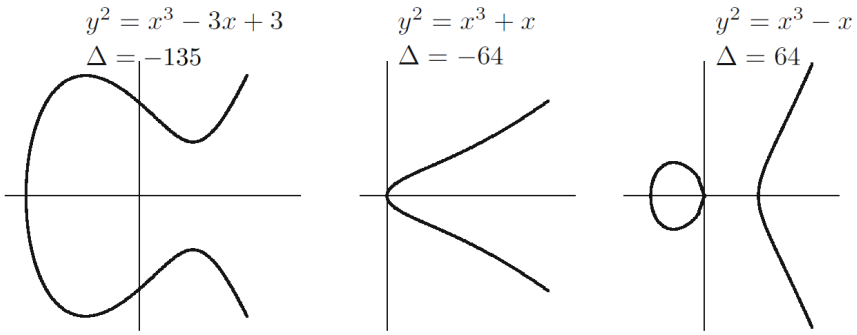


Figure 1: Three nonsingular elliptic curves

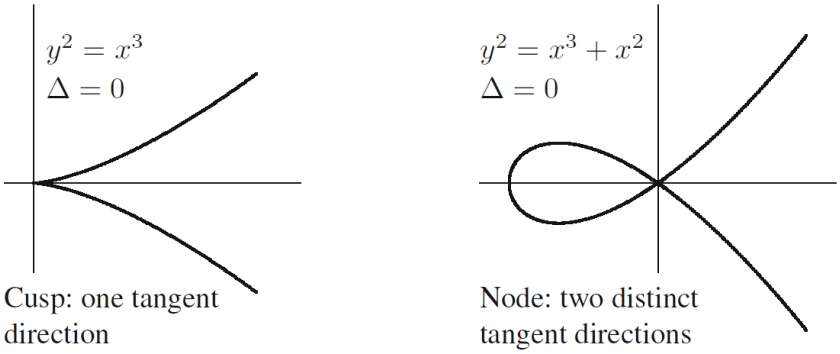


Figure 2: Two singular elliptic curves with different singularity types

Definition 2.1.10 (Addition on Elliptic Curves)**.** Let $E : y^2 = x^3 + ax + b$ be an elliptic curve, and let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be two points on E that are different from ∞. We define the addition $\oplus$ on the elliptic curve as follows:

- $P \oplus \infty = \infty \oplus P = P$ for any point $P \in E$;
- If $x_1 = x_2$ and $y_1 = -y_2$, i.e. they lie on a vertical line, then $P_1 \oplus P_2 = \infty$;
- Otherwise set

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1} & \text{if } P_1 \neq P_2, \\ \frac{3x_1^2 + a}{2y_1} & \text{if } P_1 = P_2, \end{cases}$$

then the point $(P_1 \oplus P_2) = (x_3, y_3)$ is defined by

$$x_3 = \lambda^2 - x_1 - x_2,$$
$$y_3 = \lambda(x_1 - x_3) - y_1.$$

Theorem 2.1.11. *Let E be a nonsingular elliptic curve defined over a field k. Then E, together with the point addition $\oplus$ defined above defines an abelian group structure.*

Proof. For the most part, it is clear to check the axioms of a group. We have that

- **Closure:** If we take two points P, Q on the curve, then their sum is defined to be an intersection of the curve and the line $\overline{PQ}$, so in particular, $P \oplus Q \in E$.
- **Identity:** The point at infinity, ∞, acts as the identity in the group by definition.
- **Inverse:** The inverse of a point P is denoted $-P$, and is the reflection of P along the x-axis. Then $P + (-P) = \infty$ by definition.
- **Abelian:** By the definition of the sum, it is clear that the line between P, Q and Q, P is the same, therefore $P \oplus Q = Q \oplus P = -R$.

Associativity is the last axiom that needs to be shown. This can be done by a very tedious algebraic calculation for the two sides of $(A + B) + C = A + (B + C)$ for three distinct points $A, B, C \in E$, but

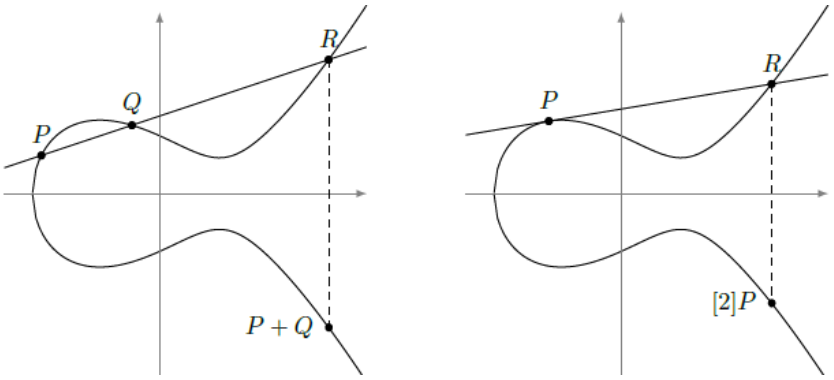


Figure 3: Addition and doubling of points on an elliptic curve

there is nothing interesting in doing so. A more interesting alternative is showing this with the help of the Cayley-Bacharach theorem, as shown by Terence Tao in [72].

Lemma 2.1.12 (Cayley-Bacharach Theorem)**.** *Let* $\gamma' = \{g(x,y) = 0\}$ *and* $\gamma'' = \{h(x,y) = 0\}$ *be two cubic curves that intersect (over some algebraically closed field* $\overline{k}$*) in precisely nine distinct points* $A_1, \ldots, A_9 \in \overline{k}^2$*. Let* f *be a cubic polynomial that vanishes on eight of these points (say* $A_1, \ldots, A_8$*). Then* f *is a linear combination of* g, h*, and in particular, vanishes on the ninth point* A_9*.*

Let us assume that we have eight distinct points on our curve:

$$\infty, \quad A, B, C, \quad A+B, B+C, -(A+B), -(B+C),$$

which are themselves distinct from both $-((A+B)+C)$ and $-(A+(B+C))$, since the elliptic curve is nonsingular by assumption. This, and the following argument is visualized in Figure 4.

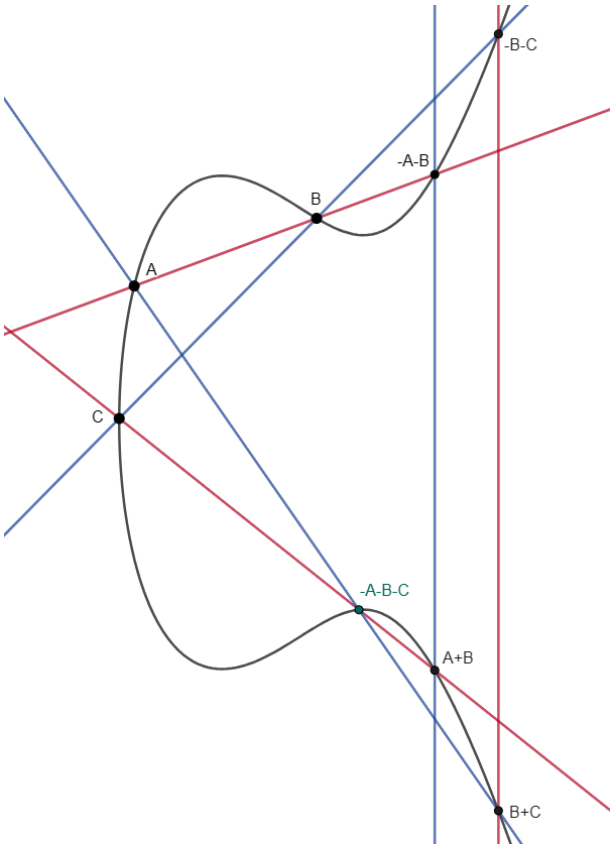


Figure 4: Illustration of the associativity of the elliptic curve addition. Figure made with GeoGebra.

Let γ be the elliptic curve E (black curve), let γ' be the union of the three red lines, and let γ'' be the union of the three blue lines in Figure 4. Now we can use the Cayley-Bacharach Theorem in the following way. First, notice that γ and γ' are cubic curves with no common component that meet at the nine distinct points shown in Figure 4. Initially, the cubic curve γ'' goes through all the points, except for $-((A+B)+C)$, but by the Cayley-Bacharach Theorem, it must also go through $-((A+B)+C)$, as that is the ninth point. This implies that the line through A and $B+C$ meets the elliptic curve γ in both $-((A+B)+C)$ and $-(A+(B+C))$, so these two points must be equal, therefore we have shown that

$$(A+B)+C = A+(B+C),$$

so $(E, \oplus)$ is indeed an abelian group, as required. □

From now on, we will drop the special notation $\oplus$ in favor of $+$, but we must remember which addition the plus sign will denote, which should be obvious from context. Furthermore, for $m \in \mathbb{Z}$ and $P \in E$, we will also denote

$$[m]P = \underbrace{P + \cdots + P}_{m \text{ times}}, \quad \text{and} \quad [-m]P = \underbrace{-P - \cdots - P}_{m \text{ times}}.$$

We have shown that elliptic curves admit a group structure, so it would be of interest to investigate what kind of maps exist between elliptic curves. But first, let's see under what circumstances we are talking about the same elliptic curves.

Definition 2.1.13. Let $E : y^2 = x^3 + Ax + B$ and $E' : y^2 = x^3 + A'x + B'$ be two elliptic curves defined over a field k, given in short Weierstrass form. These two curves are **isomorphic over** k if there exists a change of variables preserving ∞ and the form of the equation that is given by

$$(x, y) \longmapsto (u^2x', u^3y')$$

for some unit $u \in k$. In this case, we have also that the coefficients satisfy $A = u^4A'$ and $B = u^6B'$.

Definition 2.1.14. The **j-invariant** of an elliptic curve $E : y^2 = x^3 + Ax + B$ over a field k that has characteristic different from 2 or 3 is the quantity

$$j(E) = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2}.$$

Theorem 2.1.15. *[69, III.1.4b] Two elliptic curves in short Weierstrass form over a field k with char(k) ≥ 5 are isomorphic over $\overline{k}$ if and only if they both have the same j-invariant.*

Proof. To prove the forward direction, assume that E, E' are two isomorphic elliptic curves. By Definition 2.1.13, we have that the coefficients of the two curves are related by $A = u^4A'$ and $B = u^6B'$. We then get that

$$\begin{aligned} j(E) = 1728\frac{4A^3}{4A^3 + 27B^2} &= 1728\frac{4(u^4A')^3}{4(u^4A')^3 + 27(u^6B')^2} \\ &= 1728\frac{u^{12} \cdot 4(A')^3}{u^{12} \cdot (4(A')^3 + 27(B')^2)} \\ &= 1728\frac{4(A')^3}{4(A')^3 + 27(B')^2} \\ &= j(E'). \end{aligned}$$

Conversely for the other direction, assume that $j(E) = j(E')$. Then we have that

$$\frac{4A^3}{4A^3 + 27B^2} = \frac{4(A')^3}{4(A')^3 + 27(B')^2} \implies A^3(B')^2 = (A')^3B^2.$$

Remember that since these are elliptic curves, they are nonsingular, so in particular

$$\Delta = -16(4A^3 + 27B^2) \neq 0 \quad \text{and} \quad \Delta' = -16(4(A')^3 + 27(B')^2) \neq 0.$$

We look for an isomorphism of the form $(x, y) \mapsto (u^2x', u^3y')$, and consider three cases:

1. $A = 0$ ($j = 0$): Then $B \neq 0$, since $\Delta \neq 0$, so $A' = 0$, and we get an isomorphism using $u = (B/B')^{1/6}$.
2. $B = 0$ ($j = 1728$): Then $A \neq 0$, so $B' = 0$, and we geg an isomorphism using $u = (A/A')^{1/4}$.
3. $AB \neq 0$ ($j \neq 0, 1728$): Then $A'B' \neq 0$, since if one of them were 0, then both of them would be 0, but we have that $\Delta' \neq 0$. Taking either of the u's from the above two cases then gives an isomorphism. □

Remark 2.1.16. It is important to note that the isomorphism classes of elliptic curves are characterized by the j-invariant only if they are defined over an algebraically closed field $\overline{k}$. It can happen that two curves defined over k have the same j-invariant, but are not isomorphic over k, but rather over some extension of k of at most degree 6.

Proposition 2.1.17. *Let $E = E(\overline{k})$ be an elliptic curve defined over the algebraic closure $\overline{k}$. Then we have that*

$$j(E) \in k \iff E \text{ is } \overline{k}\text{-isomorphic to } E'(k).$$

Proof. We start by a general construction that will help us in proving this proposition. Let $j \in k$ with $j \neq 0, 1728$ and let E_j be the curve defined by

$$E_j : y^2 = x^3 + \frac{3j}{1728 - j}x + \frac{2j}{1728 - j}.$$

Calculating the j-invariant of this curve yields

$$\begin{aligned} j(E_j) &= 1728\frac{4A^3}{4A^3 + 27B^2} = 1728\frac{4(\frac{3j}{1728-j})^3}{4(\frac{3j}{1728-j})^3 + 27(\frac{2j}{1728-j})^2} \\ &= 1728\frac{4 \cdot 3^3 j^3}{(1728 - j)^3\left(4 \cdot 3^3 \frac{j^3 + j^2(1728-j)}{(1728-j)^3}\right)} \\ &= 1728\frac{j^3}{j^3 + j^2(1728 - j)} = 1728\frac{j^3}{1728j^2} \\ &= j. \end{aligned}$$

Moreover, we have that the j-invariants of the curves $y^2 = x^3 + 1$ and $y^2 = x^3 + x$ are 0 and 1728 respectively.

($\Longleftarrow$) If $E(k) : y^2 = x^3 + Ax + B$ is defined over k, then by definition $A, B \in k$. Then, by the definition of the j-invariant $j(E)$, it is composed of sums, multiplications and inverses of $A, B \in k$, therefore $j(E) \in k$.

($\Longrightarrow$) If $j(E) \in k$, then $j(E) = j(E_j)$ by setting $j = j(E)$, and so $E \cong E_j$ over $\overline{k}$ by Theorem 2.1.15. Since the coefficients of E_j are composed of sums, multiplications and inverses in k, they are also in k, hence there exists an isomorphism from E to a curve E_j defined over k, therefore E is also defined over k. □

Following Remark 2.1.16, let us give a definition and some examples of what happens when two elliptic curves have the same j-invariant, but are not isomorphic over some finite field k.

Definition 2.1.18. Let E, E' be two elliptic curves defined over a field k, such that $j(E) = j(E')$. An isomorphism between E, E' that is itself not defined over k is called a **twist**.

Example 2.1.19 (Sextic Twist)**.** Let $E_1 : y^2 = x^3 + 1$ and $E_2 : y^2 = x^3 + 2$ be two curves defined over the rational numbers $\mathbb{Q}$. From Definition 2.1.14, we clearly have that $j(E_1) = j(E_2) = 0$, therefore they are isomorphic over $\overline{\mathbb{Q}}$. To find the algebraic map explicitly, we have from Definition 2.1.13 there exists a unit $u \in \overline{Q}$ such that

$$B_1 = u^6 B_2 \implies 1 = 2u^6 \implies u^6 = \frac{1}{2}.$$

Clearly there is no solution to this equation for $u \in \mathbb{Q}$, so E_1, E_2 are not isomorphic over $\mathbb{Q}$. It is not hard to see that if we instead defined the elliptic curves over the number field $\mathbb{Q}(\sqrt[6]{1/2})$, then they would become isomorphic. Since this extension has degree 6, E_2 is called the *sextic twist* of E_1.

Example 2.1.20 (Quartic twist)**.** Similarly, we could have $E_1 : y^2 = x^3 + x$ and $E_2 : y^2 = x^3 + 2x$ over $\mathbb{Q}$. Again, we have that $j(E_1) = j(E_2) = 1728$, so they are isomorphic over $\overline{\mathbb{Q}}$. Same calculation as above gives that $u^4 = \frac{1}{2}$, so the smallest number field for which this isomorphic is defined is $\mathbb{Q}(\sqrt[4]{1/2})$, which has degree 4 over $\mathbb{Q}$.

Example 2.1.21. For most elliptic curves, there are only two isomorphisms, the trivial (x, y) and the negation $(x, -y)$. The elliptic curves corresponding to $j = 0$ and $j = 1728$ are extra special, because they posses more isomorphic maps to themselves. Let ρ and i be a 3rd and 4th root of unity respectively. We then have that for $j(E) = 0$, applying the map $\rho : (x, y) \mapsto (\rho X, -Y)$ gives

$$y^2 = x^3 + B \longrightarrow (-Y)^2 = (\rho X)^3 + B \iff Y^2 = X^3 + B,$$

and similarly for $j(E) = 1728$, applying the map $i : (x, y) \mapsto (-X, iY)$ gives

$$y^2 = x^3 + Ax \longrightarrow (iY)^2 = (-X)^3 - AX \iff -Y^2 = -X^3 - Ax \iff Y^2 = X^3 + AX.$$

If ρ or i are not defined in $\mathbb{F}_q$, then these self-isomorphisms are not defined over $\mathbb{F}_q$.

Some deep theorems in the study of elliptic curves concern the number of k-rational points for some field k. For example, the famous Mordell-Weil Theorem says that $E(K)$, where K is a number field, is finitely generated, and Siegel's Theorem states that there are finitely many integral point solutions to $E(K)$. It is not that surprising though that if $E(\mathbb{F}_q)$ is defined over a finite field with $q = p^n$ elements, then the number of points is also finite. A better question is how many are there, and can we determine it fast?

Heuristically speaking, every elliptic curve in short Weierstrass form is $y^2 = f(x)$ for some degree-3 function in x. There is a k-rational point on this curve, if $f(x)$ is a square in k, and then there are 2 solutions: (x, y) and $(x, -y)$. On the other hand, if $f(x)$ is not a square, there are 0 k-rational points with coordinates (x, y). A standard theorem in elementary number theory says that half of the numbers in $\mathbb{F}_q$ are quadratic residues, that is squares. By this, we can expect $2 \cdot (q/2) + 0 \cdot (q/2) = q$ k-rational points, and since we also have a point at infinity, adding that we can expect about $q + 1$ points for $\#E(\mathbb{F}_q)$. The next theorem gives a strong bound over the actual possible values.

Theorem 2.1.22 (Hasse's Theorem)**.** *Let $E(\mathbb{F}_q)$ be an elliptic curve defined over the finite field $\mathbb{F}_q$ with $q = p^n$ and characteristic p. Then*

$$\#E(\mathbb{F}_q) = q + 1 - t,$$

where $|t| \leq 2\sqrt{q}$.

Definition 2.1.23. The t in Hasse's Theorem is called the **trace of Frobenius**.

The proof of Hasse's Theorem requires a bit more setup, and in view of the topic of this thesis, we prove it only for the case $n = 1$, i.e. $\mathbb{F}_p$, as the elliptic curves in the CSIDH protocol are defined over $\mathbb{F}_p$ only. It is good news that the number of possible points different from the expected $q + 1$ is bounded by $\pm 2\sqrt{q}$, because it makes it possible to have a specialized algorithm for calculating it. This is exactly what Schoof [62] has done in 1985, where he proposed an algorithm that computes the number of points in polynomial time $O(\log^9 q)$. At the time, this was computationally impractical, but Atkin and Elkies [63] have made some improvements to the algorithm, which now can run as fast as $O(\log^4 q)$ with fast multiplication, and is the industry standard for point counting. This is called the SEA Algorithm for Schoof-Elkies-Atkin. This algorithm is the basis for another algorithm by Morrison, Panny, Sotáková and Wills [51] to faster calculate the traces of endomorphisms of supersingular elliptic curves. What this means is discussed in Section 2.5.

If E is defined over $\mathbb{F}_p$, then there is an additional property that we can assign to the elliptic curve. This definition is related to the number of $\mathbb{F}_p$ rational points on the curve.

Definition 2.1.24 (First Definition)**.** Let $E(\mathbb{F}_p)$ be an elliptic curve defined over the prime field $\mathbb{F}_p$. The curve is called **supersingular** if $\#E(\mathbb{F}_p) = p + 1$, and is called **ordinary** otherwise.

Remark 2.1.25. [69, pp. 145] Supersingular elliptic curves are still by definition nonsingular. This confusing terminology comes from the fact that the word "singular" used to mean "unusual/rare". This was the case for elliptic curves defined over $\mathbb{C}$ that had some extra structure in its endomorphism ring (c.f. Section 2.5), but this extra structure is present for elliptic curves defined over $\overline{\mathbb{F}}_p$ too. In some rare cases, certain elliptic curves have even more structure, hence the term "supersingular" (as in "super rare").

Equivalently, we could also say that an elliptic curve is supersingular if the corresponding trace of Frobenius $t = 0$ from Hasse's Theorem. This definition is actually not the "original" definition of supersingularity, but it is used quite a lot, since it gives an easy condition to check for (and is used in the verification of the public key in the CSIDH protocol, c.f. Subsection 3.2.4). Later on we will see the original (Definition 2.2.24, but to understand that, we will need to develop more theory in the upcoming sections.

2.2 Isogenies

We have now gotten a basic understanding of what an elliptic curve is and its group structure. In this section, we focus on maps *between* elliptic curves. Since elliptic curves are both projective varieties and abelian groups, naturally we would like to focus on maps that preserve both of these properties. We have already seen an example of such a map in the previous section, namely the substitution of $(x, y) \longmapsto (u^2x', u^3y')$ for defining an isomorphism between two elliptic curves (in short Weierstrass form). We will define now in general what such structure-preserving maps are, and focus on the most relevant properties of them.

Definition 2.2.1. Let E, E' be two elliptic curves defined over k and with points at infinity ∞, ∞' respectively. A **morphism of elliptic curves** $\varphi : E \to E'$ is a morphism of algebraic varieties that

(i) is given by a substitution of rational functions in the coordinates, i.e. $(x, y) \mapsto (R_1(x, y), R_2(x, y))$ for some $R_1, R_2 \in k(x, y)$,

(ii) is defined without singularities or poles at every point of E,

(iii) preserves the point at infinity, i.e. $\varphi(\infty) = \infty'$.

Definition 2.2.2. A nonconstant morphism of elliptic curves is called an **isogeny**. Two curves are called **isogenous** if there exists an isogeny between them.

In other words, we have that if $\varphi : E \to E'$ is an isogeny, then we have that $\varphi(P + Q) = \varphi(P) + \varphi(Q)$ for all $P, Q \in E(\overline{k})$, which is the familiar form of a group homomorphism, and that there are rational function R_1, R_2 such that $\varphi(x, y) = (R_1(x, y), R_2(x, y))$. Moreover, we can further simplify how the rational functions are given in the substitution the following way:

Lemma 2.2.3. *Let $\varphi : E \to E'$ be defined by the substitution $(x, y) \mapsto (R_1(x, y), R_2(x, y))$ for some rational functions R_1, R_2. Then the isogeny can equivalently be rewritten in the form*

$$\varphi(x, y) = (r_1(x), r_2(x) \cdot y)$$

*for some other rational functions r_1, r_2 in only the x-coordinate. This form is called the **standard form**.*

Proof. [75, pp. 51] Let $R(x, y)$ be any rational function. Since $y^2 = x^3 + Ax + B$ for all $(x, y) \in E(\overline{k})$, we can use this equation as a way to replace even powers of y by a polynomial in x, and likewise replace the odd powers of y by y times another polynomial in x. Therefore we may assume that

$$R(x, y) = \frac{p_1(x) + p_2(x)y}{p_3(x) + p_4(x)y}.$$

Next, we can multiply both the numerator and denominator by $p_3(x) - p_4(x)y$, this way we can replace the y^2 part in the denominator by a polynomial in x, and get

$$R(x,y) = \frac{q_1(x) + q_2(x)y}{q_3(x)}. \tag{2.1}$$

Consider now an isogeny given by

$$\varphi(x,y) = (R_1(x,y), R_2(x,y)),$$

as before. Since φ is a group homomorphism, we have that

$$\varphi(x,-y) = \varphi(-(x,y)) = -\varphi(x,y) \implies R_1(x,-y) = R_1(x,y) \quad \text{and} \quad R_2(x,-y) = -R_2(x,y).$$

Then if R_1 and R_2 are written in the form 2.1, then we have that

$$\begin{aligned} R_1(x,-y) = R_1(x,y) &\implies \frac{q_1(x) - q_2(x)y}{q_3(x)} = \frac{q_1(x) + q_2(x)y}{q_3(x)} \\ &\implies q_2(x) = 0 \text{ or } y = 0, \\ R_2(x,-y) = -R_2(x,y) &\implies \frac{q_1(x) - q_2(x)y}{q_3(x)} = \frac{-q_1(x) - q_2(x)y}{q_3(x)} \\ &\implies q_2(x) = 0. \end{aligned}$$

This shows that indeed we can write $\varphi(x,y) = (r_1(x), r_2(x)y)$ with rational functions r_1, r_2 in only the x-coordinate, as required. □

Now we define some important properties of isogenies that will be used extensively later.

Definition 2.2.4. Let φ be an isogeny such that $\varphi(x,y) = (r_1(x), r_2(x)y)$. If the coefficients of r_1, r_2 lie in k, then we say that φ is **defined over** k.

Definition 2.2.5 (ℓ-isogeny.)**.** Let φ be an isogeny such that $\varphi(x,y) = (r_1(x), r_2(x)y)$, and write $r_1(x) = p(x)/q(x)$ with $\gcd(p(x), q(x)) = 1$. Then the **degree** of the isogeny φ is defined to be

$$\deg(\varphi) = \max\{\deg p(x), \deg q(x)\},$$

and if $\deg(\varphi) = \ell$, then we call φ an **ℓ-isogeny**.
If the derivative $r_1'(x)$ is identically 0, we say that φ is **inseparable**, otherwise we call the isogeny **separable.**

Theorem 2.2.6. *[75, Prop. 12.8] Let $\varphi : E \to E'$ be an isogeny. If φ is separable, then*

$$\deg(\varphi) = |\ker(\varphi)|.$$

Proof. Write $\varphi(x,y) = (r_1(x), r_2(x)y)$ in standard form, with $r_1(x) = p(x)/q(x)$. Then since E' is nonsingular, $r_1' \neq 0$, so $r' = p'q - pq'$ is not the zero polynomial. Let S be the set of $x \in \overline{k}$ such that

$$(pq' - p'q)(x)q(x) = 0.$$

Let $(a,b) \in E(\overline{k})$ be such that

1. $a \neq 0, b \neq 0, (a,b) \neq \infty$,
2. $\deg(p(x) - aq(x)) = \max\{\deg(p), \deg(q)\} = \deg(\varphi)$,
3. $a \notin r_1(S)$, and
4. $(a,b) \in \varphi(E(\overline{k}))$.

Since $p'q - pq'$ is not the zero polynomial, S is a finite set, hence its image under φ is finite. The function $r_1(x)$ takes on infinitely many distinct values as x runs through $\overline{k}$, so in particular, there is a point $(x, y) \in E(\overline{k})$ for each x, so $\varphi(E(\overline{k}))$ is an infinite set. Therefore, such a pair (a, b) exists.

We claim that there are exactly $\deg(\varphi)$ points $(x_1, y_1) \in E(\overline{k})$ such that $\varphi(x_1, y_1) = (a, b)$. For such a point, we have that

$$\frac{p(x_1)}{q(x_1)} = a, \quad y_1 r_2(x_1) = b.$$

Since $(a, b) \neq \infty$, we must have that $q(x_1) \neq 0$. Since $b \neq 0$ and $y_1 r_2(x_1) = b$, we must have that $y_1 = b/r_2(x_1)$, therefore we only need to count the values of x_1. By assumption (2), we have that $p(x) - aq(x) = 0$ has $\deg(\varphi)$ roots (counting multiplicities). Thus we now need to show that $p - aq$ has no multiple roots.

Suppose that x_0 is a multiple root. Then

$$p(x_0) - aq(x_0) = 0 \quad \text{and} \quad p'(x_0) - aq'(x_0) = 0.$$

Multiplying the equations $p = aq$ and $aq' = p'$ gives

$$ap(x_0)q'(x_0) = ap'(x_0)q(x_0) \overset{a \neq 0}{\implies} x_0 \in S,$$

since x_0 is a root of $pq' - p'q$. Therefore, $a = r_1(x_0) \in r_1(S)$, contradiction assumption (3). It follows that $p - aq$ has no multiple roots, so it has $\deg(\varphi)$ distinct roots. Since there are exactly $\deg(\varphi)$ points (x_1, y_1) with $\varphi(x_1, y_1) = (a, b)$, the kernel of φ has $\deg(\varphi)$ elements. □

Since φ is a homomorphism, for each $(a, b) \in \varphi(E(\overline{k}))$, there are exactly $\deg(\varphi)$ points (x_1, y_1) with $\varphi(x_1, y_1) = (a, b)$. The assumptions on (a, b) were made during the proof to obtain this result for at least one point, which suffices.

Remark 2.2.7. If φ is not separable, then $p' - aq'$ is always the zero polynomial, so $p(x) - aq(x) = 0$ always has multiple roots, so $\deg(\varphi) > |\ker(\varphi)|$.

Now we give a couple examples of important isogenies that play a crucial role in the upcoming discussion of CSIDH in Chapter 3.

Example 2.2.8 (Doubling)**.** Let us compute the doubling formula for the Montgomery curve $E_A : y^2 = x^3 + Ax^2 + x$, as an exercise to see how one can derive the formula in Definition 2.1.10. Since we want to double a point $P = (X, Y) \in E$, we have to use the slope of the tangent line at that point. Implicit derivation gives

$$y^2 = x^3 + Ax^2 + x \iff 2yy' = 3x^2 + 2Ax + 1.$$

Evaluating the derivative at $P = (X, Y)$ gives the slope

$$\lambda = \frac{3X^2 + 2AX + 1}{2Y}.$$

The equation of the line going through P with slope λ is

$$y = \lambda(x - X) + Y.$$

To find the intersection of this line with our curve E_A, we can substitute y into the curve equation:

$$x^3 + Ax^2 + x = (\lambda(x - X) + Y)^2 = \lambda^2(x - X)^2 + 2\lambda Y(x - X) + Y^2.$$

Collecting like terms gives

$$0 = x^3 + (A - \lambda^2)x^2 + (1 + 2\lambda^2 X - 2\lambda Y)x + (Y^2 - \lambda^2 X^2 - 2\lambda XY).$$

The roots of this cubic equation are the x-coordinates of the intersection points. Since the line is set up to be tangent to E_A, we know that $P = (X, Y)$ is a double root. We also know that the sum of the roots

is going to be the negative coefficient of x^2, so then if we denote $2P = (X', -Y')$, we have that

$$2X + X' = -(A - \lambda^2) = \lambda^2 - A \implies X' = \lambda^2 - A - 2X.$$

Substituting λ from before into this expression gives

$$X' = \left(\frac{3X^2 + 2AX + 1}{2Y}\right)^2 - A - 2X.$$

The y-coordinate can be computed by similarly substituting λ into the equation of the line:

$$-Y' = \lambda(X' - X) + Y \implies Y' = \lambda(X - X') - Y = \frac{3X^2 + 2AX + 1}{2Y}(X - X') - Y.$$

We note that the x-substitution map can be further reduced to

$$X' = \frac{(X^2 - 1)^2}{4X(X^2 + AX + 1)},$$

from which we can deduce that the doubling isogeny has degree 4, and is in particular separable, since it has nonzero derivative (provided that we are not in characteristic 2).

Definition 2.2.9 (p-Frobenius morphism)**.** Let $E(\mathbb{F}_q)$ be an elliptic curve defined over a finite field $\mathbb{F}_q$ with $q = p^n$. The **p-Frobenius morphism** is defined by

$$\begin{aligned} \pi_p : E &\longrightarrow E^{(p)}, \\ (x, y) &\longmapsto (x^p, y^p), \end{aligned}$$

where $E^{(p)} : y^2 = x^3 + A^p x + B^p$ is the **p-Frobenius twist** curve of E. If we define this map with q instead, then we get

$$\begin{aligned} \pi_q : E &\longrightarrow E, \\ (x, y) &\longmapsto (x^q, y^q) \end{aligned}$$

is a map from E to itself, since $A^q = A$ for all $A \in \mathbb{F}_q$, and this map is called the **Frobenius endomorphism**.

Example 2.2.10 (1-isogeny)**.** We have seen before that two elliptic curves in short Weierstrass form are isomorphic if there is a change of variables (or algebraic map) $\varphi(x, y) = (u^2 x, u^3 y)$ for some unit $u \in \overline{k}$. This is a nonconstant algebraic map between projective varieties, so it is also an isogeny. Here we can see that $r_1(x) = u^2 x$, $r_1'(x) = u^2 \neq 0$, so φ is separable, and $\deg(\varphi) = \deg(u^2 x) = 1$, so this defines a 1-isogeny.

This also gives an alternative definition for an isomorphism of two curves.

Definition 2.2.11. Two elliptic curves E, E' are isomorphic if there exists a 1-isogeny φ between them.

Now we will focus on the fact that any isogeny of elliptic curves over a finite field $\mathbb{F}_q$ can be decomposed into the composition of a separable isogeny and a power of the Frobenius morphism. To prove this, we will need two lemmas first concerning some consequences of inseparable isogenies.

Lemma 2.2.12. *[70, Lemma 5.1.] Let $\varphi = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)} y\right)$ be an inseparable isogeny in standard form. Then*

$$\left(\frac{u}{v}\right)' = 0 \iff u = f(x^p) \quad \text{and} \quad v = g(x^p),$$

where f, g are polynomials in $k[x]$, and $p = \operatorname{char}(k)$.

Proof. Since φ is inseparable, we have that $\left(\frac{u}{v}\right)' = 0$. Then by the quotient rule of differentiation, we get that $u'v = v'u$, but since $\gcd(u, v) = 1$, we must have that $u \mid u'$. This can only happen when $u' = 0$, because $\deg u' < \deg u$. The same argument holds for $v' = 0$ as well.

Now let $u(x) = \sum a_n x^n$. If $u'(x) = \sum n a_n x^{n-1} = 0$, then $na_n = 0$ for every n, which means that n must be a multiple of p for every nonzero a_n, so we can write u as

$$u(x) = \sum_m a_{pm}(x^p)^m = f(x^p),$$

where $f = \sum a_m x^m$. Similarly, we could write $v(x) = g(x^p)$ for some $g \in k[x]$.

Conversely, if $u(x) = f(x^p)$ and $v(x) = g(x^p)$, then

$$u'(x) = px^{p-1}f'(x^p) = 0 \quad \text{and} \quad v'(x) = px^{p-1}g'(x^p) = 0.$$

Moreover, if $u' = v' = 0$, then $u'v - uv' = 0$, so then $(\frac{u}{v})' = 0$, so φ is indeed inseparable. □

Lemma 2.2.13. *[70, Lemma 5.3.] Let $\varphi : E \to E'$ be an inseparable isogeny of elliptic curves $E : y^2 = f(x)$ and $E' : y^2 = g(x)$ over a field k of characteristic $p > 0$. Then φ can be written in the form*

$$\varphi = (a(x^p), b(x^p)y^p)$$

for some rational functions $a, b \in k(x)$.

Proof. From Lemma 2.2.12, we have that $\frac{u(x)}{v(x)} = a(x^p)$ for some $a \in k(x)$, we only need to show that $\frac{s(x)}{t(x)}y = b(x^p)y^p$ for some $b \in k(x)$ as well. If we now substitute u/v and s/t into the equation for E_2, and using the equation for E_1 to eliminate y^2 yields the equality

$$v^3 s^2 f = t^2 w,$$

where $w = v^3 f(u/v) \in k[x]$. Since φ is inseparable, we have that $u' = v' = 0 \implies w' = 0$, and therefore $(w/v^3)' = (s^2 f/t^2)' = 0$. Thus $s(x)^2 f(x) = F(x^p)$ and $t(x)^2 = G(x^p)$ for some polynomials F, G. If $x_0 \in \overline{k}$ is a root of f, then x_0^p is a root of F, so

$$(x - x_0^p) \mid F(x) \implies (x^p - x_0^p) = (x - x_0)^p \mid F(x^p).$$

The roots of f are distinct, so $f(x)^p$ divides $F(x^p)$, and $F(x^p) = h_1(x^p)f(x)^p$ for some $h_1 \in k[x]$.

We now have that $s(x)^2 = h_1(x^p)f(x)^{p-1}$. Since p is odd, $h_1(x^p)$ is a square, and in particular $h_1(x^p) = h_2(x)^2$, where $h_2 = s/f^{(p-1)/2} \in k[x]$. We have now that

$$F(x^p)' = 0 \implies H' = 0 \implies (h_2^2)' = 2h_2 h_2' = 0,$$

since h_1 cannot be zero (s is not) and $p \neq 2$. So $h_2(x) = H(x^p)$ for some $H \in k[x]$ and

$$s(x)^2 f(x) = H(x^p)^2 f(x)^p.$$

We now note that

$$(s(x)y)^2 \equiv s(x)^2 f(x) \equiv H(x^p)^2 f(x)^p \equiv (H(x^p)y^p)^2,$$

where the equivalences are modulo the curve equation for E_1. Therefore we have that

$$\left(\frac{s(x)}{t(x)}y\right)^2 \equiv \left(\frac{H(x^p)y^p}{G(x^p)}\right)^2 \implies \frac{s(x)}{t(x)}y \equiv b(x^p)y^p,$$

where $b = \pm H/G$, as required. □

Now we can move on to proving the main theorem.

Theorem 2.2.14. *[70, Corollary 5.4.] Let $\varphi = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny of elliptic curves over a field k of characteristic $p > 0$ in standard form. Then*

$$\varphi = \varphi_{sep} \circ \pi^n$$

for some separable isogeny φ_{sep} and integer $n \geq 0$, where π is the p-Frobenius morphism. Furthermore, we then have that $\deg \varphi = p^n \deg(\varphi_{sep})$.

Proof. If φ is separable, then we clearly have that $\varphi_{\text{sep}} = \varphi$ and $n = 0$, so now assume that φ is inseparable. By Lemma 2.2.13, we can write

$$\varphi = (r_1(x^p), r_2(x^p)y^p) = \varphi_1 \circ \pi,$$

with $\varphi_1 = (r_1(x), r_2(x)y)$. If this φ_1 is separable, we are done, otherwise we apply the same procedure again to φ_1, and eventually obtain $\varphi = \varphi_n \circ \pi^n$, where φ_n is a separable isogeny. This process terminates, because each step reduces the degree of φ_n by a factor of p.

If $\varphi_n = \varphi_{\text{sep}} = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ is in standard form, composing with π^n has the effect of replacing $u(x)$ by $u(x^{p^n})$ and $v(x)$ by $v(x^{p^n})$. Then we have that

$$\deg \varphi = \max\{p^n \deg u, p^n \deg v\} = p^n \max\{\deg u, \deg v\} = p^n \deg(\varphi_{\text{sep}}). \qquad \square$$

Corollary 2.2.15. *The order of the kernel of an isogeny φ equals its separable degree. In other words,*

$$|\ker \varphi| = \deg_s \varphi.$$

Proof. From Theorem 2.2.14, we have that $\varphi = \varphi_{\text{sep}} \circ \pi^n$. The corollary follows from Theorem 2.2.6 if we can show that the inseparable part has trivial kernel. This is easy to see since, expressed in homogeneous coordinates, we can have $(x^p : y^p : z^p) = (0 : 1 : 0)$ if and only if $(x : y : z) = (0 : 1 : 0)$. $\square$

Definition 2.2.16. For an isogeny $\varphi = \varphi_{\text{sep}} \circ \pi^n$ decomposed as in Theorem 2.2.14, we define the **separable degree** $\deg_s \varphi$ and the **inseparable degree** $\deg_i \varphi$ as

$$\deg_s \varphi := \deg \varphi_{\text{sep}} \quad \text{and} \quad \deg_i \varphi := p^n,$$

and we also have the relation

$$\deg \varphi = (\deg_s \varphi)(\deg_i \varphi).$$

Definition 2.2.17. If an isogeny has separable degree 1, i.e. $\deg_s \varphi = 1$, then this isogeny is called **purely inseparable**.

The degree of a purely inseparable isogeny is always a power of p, but the converse does not hold, as shown in the following example.

Example 2.2.18. Let $E(\mathbb{F}_p)$ be an elliptic curve over the finite field $\mathbb{F}_p$. Recall that for an *ordinary* curve, we have that $E[p] = \mathbb{Z}/p\mathbb{Z}$, so consider the multiplication-by-p map

$$[p] : E \longrightarrow E$$

. We have that

$$\deg([p]) = \deg_s([p]) \deg_i([p]),$$

and also that

$$\deg_s([p]) = |\ker[p]| = |E[p]| = p.$$

Since $\deg([p]) = p^2$, we clearly see that $[p]$ is neither purely separable nor purely inseparable, but its degree is a power of p.

Remark 2.2.19. Note that 1-isogenies, which are also isomorphisms as shown in Definition 2.2.11, are both separable and purely inseparable. Indeed, its separability was shown already in Example 2.2.10, and pure separability follows from

$$1 = \deg \varphi = p^n \deg(\varphi_{\text{sep}}) \implies n = 0 \quad \text{and} \quad \deg(\varphi_{\text{sep}}) = 1.$$

If we are able to go from one elliptic curve to another via an isogeny, then one might think that we should be able to go backwards as well. Not every isogeny is invertible, in the sense that composing the isogeny with its inverse gives the identity isogeny, but there is an analogous definition for when this composition yields a multiplication-by-m map.

Definition 2.2.20. Let $\varphi : E \to E'$ be a nonconstant isogeny of degree m. The **dual isogeny** to φ is the unique isogeny $\hat{\varphi} : E' \to E$ such that

$$\hat{\varphi} \circ \varphi = [m].$$

Proving the existence of the dual isogeny is outside of the scope of this paper (uses invariant differentials, for a full proof see [69, Theorem III.6.1.]), but it is easy to show uniqueness. Suppose that $\hat{\varphi}$ and $\hat{\varphi}'$ are two such isogenies. Then

$$(\hat{\varphi} - \hat{\varphi}') \circ \varphi = [m] - [m] = [0],$$

and since φ is nonconstant, $\hat{\varphi} - \hat{\varphi}'$ must be constant, so $\hat{\varphi} = \hat{\varphi}'$.

We record some relevant properties of the dual isogeny.

Theorem 2.2.21 (Properties of dual isogenies)**.** *Let $\varphi : E \to E'$ be an isogeny with degree m. Then the following properties are true:*

(i) Let $\lambda : E' \to E''$ be another isogeny. Then

$$\widehat{\lambda \circ \varphi} = \hat{\varphi} \circ \hat{\lambda}.$$

(ii) Let $\psi : E \to E'$ be another isogeny. Then

$$\widehat{\varphi + \psi} = \hat{\varphi} + \hat{\psi}.$$

(iii) For all $m \in \mathbb{Z}$

$$\widehat{[m]} = [m] \quad \text{and} \quad \deg[m] = m^2.$$

(iv) $\deg \hat{\varphi} = \deg \varphi$.

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

Proof. With the exception of (ii), the proofs of this theorem are not too hard to show.

(i) Let $n = \deg \lambda$, we then have

$$(\hat{\varphi} \circ \hat{\lambda}) \circ (\lambda \circ \varphi) = \hat{\varphi} \circ [n] \circ \varphi = [n] \circ \hat{\varphi} \circ \varphi = [nm].$$

The uniqueness in the statement of Definition 2.2.20 implies that $\widehat{\lambda \circ \varphi} = \hat{\varphi} \circ \hat{\lambda}$.

(ii) Outside of the scope (can be done with divisors or the Weil pairing), see [69, Theorem III.6.2c]

(iii) The statement is clearly true for $m = 0, 1$. Using (ii) with $\varphi = [m]$ and $\psi = [1]$, we get

$$\widehat{[m+1]} = \widehat{[m]} + \widehat{[1]},$$

and ascending and descending induction shows that $\widehat{[m]} = [m]$ for all m. Now let $d = \deg[m]$ and consider the multiplication-by-d map. Then

$$[d] = \widehat{[m]} \circ [m] = [m^2] \implies [d - m^2] = [0] \implies d = m^2.$$

(iv) Using (iii) and the definition of the dual isogeny, we have that

$$m^2 = \deg[m] = \deg(\varphi \circ \hat{\varphi}) = (\deg \varphi)(\deg \hat{\varphi}) = m \cdot \deg \hat{\varphi} \implies m = \deg \hat{\varphi}.$$

(v) Again, using the definition, (i) and (iii) gives

$$\hat{\varphi} \circ \varphi = [m] = \widehat{[m]} = \widehat{\hat{\varphi} \circ \varphi} = \hat{\varphi} \circ \hat{\hat{\varphi}} \implies \varphi = \hat{\hat{\varphi}}. \qquad \square$$

Now we can focus a bit more on the multiplication-by-m maps, and their kernels in particular. This investigation will let us give a new definition for supersingular curves, which will be used later in this thesis to ease some proofs. Recall that any finite abelian group can be written as a direct sum of cyclic groups of prime power order (this is the fundamental theorem of abelian groups). We start with a definition.

Definition 2.2.22. Let E be an elliptic curve and $[m]$ the multiplication-by-m map on E. Then the **m-torsion part** of E is defined as

$$E[m] = \ker[m] = \{P \in E \mid [m]P = \infty\}.$$

Theorem 2.2.23. *[70, Theorem 6.1.] Let $E(k)$ be an elliptic curve defined over the field k, and let $char(k) = p$. For each prime ℓ, we have*

$$E[\ell^r] \cong \begin{cases} \mathbb{Z}/\ell^r\mathbb{Z} \oplus \mathbb{Z}/\ell^r\mathbb{Z} & \text{if } \ell \neq p, \\ \mathbb{Z}/\ell^r\mathbb{Z} \text{ or } \{\infty\} & \text{if } \ell = p. \end{cases}$$

Proof. We start by considering $\ell \neq p$. By [69, Corollary III.5.4], the multiplication-by-ℓ map is separable when $\ell \nmid p$, so $\#E[\ell] = \ker[\ell] = \deg[\ell] = \ell^2$ by 2.2.21(iii). Every element of $E[\ell]$ has order ℓ, so with size ℓ^2, we must have that the torsion subgroup decomposes to $E[\ell] = \mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}$. If $E[\ell^r]$ decomposes to $\langle P_1 \rangle \oplus \cdots \oplus \langle P_n \rangle$ with each $P_i \in E(\overline{k})$ of order $\ell^{e_i} > 1$, then

$$E[\ell] \cong \langle \ell^{e_1 - 1} P_1 \rangle \oplus \cdots \oplus \langle \ell^{e_n - 1} P_n \rangle \cong (\mathbb{Z}/\ell\mathbb{Z})^n,$$

in which case $n = 2$. Again, since $\#E[\ell^r] = \ell^{2r}$, by a similar argument we must have that $E[\ell^r] = \mathbb{Z}/\ell^r\mathbb{Z} \oplus \mathbb{Z}/\ell^r\mathbb{Z}$.

Now suppose $\ell = p$. We have that $\deg[\ell] = \deg_s[\ell] \deg_i[\ell] = \ell^2$ by Definition 2.2.16. We also have that $\deg_s[\ell] > 1$, so clearly it must be either 1 or ℓ, which means $E[\ell]$ is isomorphic to $\mathbb{Z}/\ell\mathbb{Z}$ or $\{\infty\}$. In the latter case, we trivially have that $E[\ell^r] = \{\infty\}$.

Assume now that $E[\ell] \cong \mathbb{Z}/\ell\mathbb{Z}$. Then by the same argument as before, $E[\ell^r] = \langle P \rangle$ with P having order $\ell^e \neq \ell^r$. Since $[\ell]$ is an isogeny, it is surjective, so $P = \ell Q$ for some $Q \in E(\overline{k})$, and $P \neq 0$ implies Q has order ℓ^{e+1}, and in particular it's not an element of $E[\ell^r]$. This is possible only if $e = r$, so we indeed have $E[\ell^r] \cong \mathbb{Z}/\ell^r\mathbb{Z}$ as required. $\square$

Definition 2.2.24 (Second definition)**.** In the case of $E[p^e] = \mathbb{Z}/p^e\mathbb{Z}$, we call the elliptic curve **ordinary**, and in the case of $E[p] = \{\infty\}$, we call the elliptic curve **supersingular**.

The fact that this definition and Definition 2.1.24 are equal is not trivial, but it requires a bit more knowledge, which we are going to gain in the coming sections. Their equivalence is proved later in Theorem 2.5.9, but before that, let us see if we can use these ℓ-torsion points to define isogenies. This would help us generate isogenies easier, paving to way to isogeny-based cryptography.

2.3 Vélu's Formulae

From the previous section, we have seen that the kernel of any isogeny $\varphi : E \to E'$ is finite (Corollary 2.2.15), and in particular it is a finite subgroup of $E(\overline{k})$. It is natural to ask the harder question of whether by choosing a finite subgroup $G \subset E(\overline{k})$ it is possible to define another isogeny $\psi : E \to E'$ such that $\ker \psi = G$. This is indeed true, and furthermore this isogeny would be uniquely defined by our chosen subgroup G up to isomorphism. The proof of this is, however, outside of the scope of this paper since it uses elements from algebraic geometry, such as the Hurwitz genus formula, but we record the statement of the theorem, as it is one of the pillars of isogeny-based cryptography.

Theorem 2.3.1. *[70, Theorem 5.11] Let $E(k)$ be an elliptic curve and let G be a finite subgroup of $E(\overline{k})$. There exists an elliptic curve E' and a separable isogeny $\varphi : E \to E'$ with $\ker \varphi = G$. The curve E' and the isogeny φ are defined over a finite extension of k and unique up to isomorphism.*

Proof. See in Silverman's book [69, Proposition III.4.12], which uses Galois theory, function field pullbacks and the aforementioned Hurwitz genus formula. □

It is nice to have a theoretical proof of existence, but so far it is nonconstructive. Thankfully, Vélu has given us an explicit way of constructing the required isogeny in [74]. We now give the general construction with the general Weierstrass equation in affine coordinates, as given in [75, Theorem 12.16]. The proof shall be omitted as it is mostly algebraic manipulation, and consequently not very enlightening.

Theorem 2.3.2. *Let E be an elliptic curve given by*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

with $a_i \in k$. Let $G \subset E(\overline{k})$ be a finite subgroup. For a point $Q = (x_Q, y_Q) \in G$ with $Q \neq \infty$, define

$$\begin{aligned} g_Q^x &= 3x_Q^2 + 2a_2x_Q + a_4 - a_1y_Q; \\ g_Q^y &= -2y_Q - a_1x_Q - a_3; \\ v_Q &= \begin{cases} g_Q^x & \text{if } 2Q = \infty, \\ 2g_Q^x - a_1g_Q^y & \text{if } 2Q \neq \infty; \end{cases} \\ u_Q &= (g_Q^y)^2. \end{aligned}$$

Let G_2 be the points of order 2 in G. Choose $R \subset G$ such that for each pair of non-2-torsion points $P, -P \in G$, put exactly one of them in R. We then have a disjoint union

$$G = \{\infty\} \cup G_2 \cup R \cup (-R).$$

Now let $S = R \cup G_2$ and set

$$v = \sum_{Q \in S} v_Q, \quad w = \sum_{Q \in S} (u_Q + x_Q v_Q),$$

then E' has the equation

$$Y^2 + A_1XY + A_3Y = X^3 + A_2X^2 + A_4X + A_6,$$

where

$$\begin{aligned} &A_1 = a_1, \quad A_2 = a_2, \quad A_3 = a_3, \\ &A_4 = a_4 - 5v, \\ &A_6 = a_6 - (a_1^2 + 4a_2)v - 7w. \end{aligned}$$

The isogeny defined by the kernel G is then given by

$$\begin{aligned} X &= x + \sum_{Q \in S} \left(\frac{v_Q}{x - x_Q} + \frac{u_Q}{(x - x_Q)^2} \right), \\ Y &= y - \sum_{Q \in S} \left(u_Q \frac{2y + a_1x + a_3}{(x - x_Q)^3} + v_Q \frac{a_1(x - x_Q) + y - y_Q}{(x - x_Q)^2} + \frac{a_1u_Q - g_Q^x g_Q^y}{(x - x_Q)^2} \right). \end{aligned}$$

Remark 2.3.3. As seen in the statement of Theorem 2.3.2, if one wants to calculate an evaluation of an isogeny, especially if it has large degree, then there are a lot of divisions in the formula. Divisions, or inversions in a finite field, are very costly algorithmically compared to multiplication, squaring or addition. This is where swapping to homogeneous coordinates really shine, because if we go from (x, y) to $[x : y : 1]$, then we can keep track of all the "denominators" in the algorithm inside the z-coordinate.

This results in a substantial speed-up, since the many inversions are changed to other, cheaper arithmetic operations, and we only need to calculate two inversions at the end to get back the affine coordinates by $[x : y : z] \mapsto (x/z, y/z)$.

Let us illustrate this rather complicated formula by computing a 2-isogeny from an elliptic curve in our CSIDH-relevant (and favorite) form, the Montgomery form.

Example 2.3.4. [75, Example 12.14] Let $E : y^2 = x^3 + Ax^2 + x$ be an elliptic curve in Montgomery form. We would like a separable 2-isogeny, so we need to find a subgroup G such that $|G| = 2$. To this end, we can take the point $(0, 0)$, which then gives $G = \{\infty, (0, 0)\}$. The set S has only one point $(0, 0)$, so we only have to calculate the intermediate values for $Q = (0, 0)$. We have $v_Q = 1$ and $u_Q = 0$, so we have

$$X = x + \frac{1}{x}, \quad Y = y - \frac{y}{x^2}.$$

The target curve E_2 is then given by

$$Y^2 = X^3 + AX^2 - 4X - 4A.$$

Define a new change of variables by

$$X_3 = X + A = x + \frac{1}{x} + A = x + \frac{Ax + 1}{x} = \frac{y^2}{x^2},$$
$$Y_3 = Y = y - \frac{y}{x^2} = y\frac{x^2 - 1}{x^2}.$$

Then we obtain the elliptic curve E_3 given by

$$Y_3^2 = X_3^3 - 2AX_3^2 + (A^2 - 4)X_3.$$

The map $\varphi : E \to E_3$ given by $\varphi(x, y) = (\frac{x^2+Ax+1}{x}, y\frac{x^2-1}{x^2})$ in standard form shows that this indeed is a 2-isogeny. Note that E_3 is *not* in Montgomery form, but it could be transformed into one by an isomorphism $(X_3, Y_3) \to (X_3/\sqrt{A^2 - 4}, Y_3/(A^2 - 4)^{3/4})$.

The elliptic curve E_3 again has $(0, 0)$ as a point with order 2, so performing the same process gives

$$E_4 : Y_4^2 = X_4^3 + 4AX_4^2 + 16X_4$$

as a result of the change of variables

$$X_4 = X_3 + \frac{-2AX_3 + A^2 - 4}{X_3^2}, \quad Y_4 = Y_3 - \frac{(A^2 - 4)Y_3}{X_3^2}.$$

If we let $X_5 = X_4/4$, and $Y_5 = Y_4/8$, then we get

$$Y_5^2 = X_5^3 + AX_5^2 + X_5,$$

which is the equation of our original curve E. A (rather tedious) calculation shows that in the map $E \to E$, we have the x-coordinate substitution

$$x \mapsto X_5 = \left(\frac{3x^2 + 2Ax + 1}{2y}\right)^2 - A - 2x.$$

Doing the same calculation for the y-coordinate reveals that this map is the multiplication-by-2 map from Example 2.2.8. We can also see that the degree of this map is 4, since the leading term of the numerator after expanding the parenthesis is x^4, as expected from Theorem 2.2.21(iii).

Vélu's formulae are very useful indeed, but since ℓ-isogenies have kernels of size ℓ, and Vélu uses every point in the defined kernel, it has complexity $O(\ell)$, that is the amount of steps needed is linear in ℓ. Cryptographically, this can become a real problem when it comes to calculating larger isogenies, since

it takes linearly more time. It is no use if a cryptographic protocol is quantum-secure, but takes 10 minutes to calculate. To try to fix this problem, Bernstein, De Feo, Leroux and Smith [3] have given an algorithm to compute the ℓ-isogeny in only $O(\sqrt{\ell})$ steps. They have amazingly named this algorithm "$\sqrt{}$élu formula".

Giving the whole algorithm for this thesis is out of the scope, because it is not a straightforward formula, but rather a new way of looking at Vélu's formula from an abstract standpoint. The theoretical foundation of $\sqrt{}$élu is the following. The main computational cost in Vélu's algorithm is evaluating the polynomial

$$h_S(X) = \prod_{i \in I} (X - x([i]P)), \tag{2.2}$$

where I is some indexing set, and $x(\cdot)$ denotes the usual x-coordinate of the point. In Vélu, this indexing set is all the multiples of the kernel generator point, going up to $(\ell-1)/2$, since we're taking a symmetric set. Together this costs 6ℓ field multiplications, which is what the $\sqrt{}$élu team has improved on by instead of treating this polynomial as a linear algorithm, they evaluate it using a baby-step-giant-step strategy. They take the set of odd integers up until $\ell - 2$, then they divide this set into two "sets of indices" I, J such that $I \pm sJ$ covers S, where s is a fixed integer, representing the size of the giant steps. Then they rewrite 2.2 as

$$h_S(X) = \prod_{i \in I} \prod_{j \in J} \big(X - x([i + sj]P)\big)\big(X - x([i - sj]P)\big).$$

This allows them to compute this function efficiently by resultants of two polynomials, resulting in the exponential speedup.

The above theorems define the isogeny for elliptic curves in Weierstrass forms, but let us also record the construction of the similar Vélu-style isogeny of odd degree for an elliptic curve in Montgomery form. The isogeny construction for $E : By^2 = x^3 + Ax^2 + x$ was given first by Costello and Hisil [25], which was later generalized for $B = 1$ by Renes [59], which is the form that is used exclusively in CSIDH, and thus is also included in this thesis. The construction of the isogeny is done in the following way.

Theorem 2.3.5. *Let k be a field with char(k) $\neq 2$, and $E(k) : y^2 = x^3 + ax^2 + x$ be an elliptic curve in Montgomery form. Let $G \in E(\overline{k})$ be a finite subgroup such that $(0,0) \notin G$, and let φ be a separable isogeny such that $\ker(\varphi) = G$. Then there exists a curve $E'(k) : y^2 = x^3 + Ax^2 + x$ such that, up to post-composition with an isomorphism,*

$$\begin{aligned} \varphi : E &\longrightarrow E' \\ (x, y) &\longmapsto (f(x), c_0 y f'(x)), \end{aligned}$$

where

$$f(x) = x \prod_{T \in G \setminus \{\infty\}} \frac{x x_T - 1}{x - x_T}.$$

Moreover, writing

$$\pi = \prod_{T \in G \setminus \{\infty\}} x_T, \quad \text{and} \quad \sigma = \sum_{T \in G \setminus \{\infty\}} \left(x_T - \frac{1}{x_T} \right),$$

we have that $A = \pi(a - 3\sigma)$ and $c_0^2 = \pi$.

Proof. See [59, Proposition 1] □

This is the affine form of the algorithm, but of course, during implementation, all the calculation are done in projective coordinates to avoid a lot of expensive field inversions. Renes' isogeny generator is "closed under Montgomery curves", that is it is guaranteed to end up on a Montgomery curve. Vélu's formula is also great, but they target curves in short Weierstrass form, so an extra transformation is required to get to the Montgomery curve model. The next example illustrates this difference.

Example 2.3.6. Let $E_0 : y^2 = x^3 + x$ be the starting curve defined over the prime field $\mathbb{F}_{11}$. Let us calculate a 3-isogeny from this curve, which exists because E_0 is supersingular, so $\#E(\mathbb{F}_{11}) = 12 = 3 \cdot 4$.

Let $P = (5, 8)$, which is a point of order 3, so we can use it as the cyclic kernel for our isogeny calculation as $\langle P \rangle = \{ (5, 8), (5, -8), \infty \}$.

Let us start with Renes' formula. We have that $x_T = 5$ for both points, so we have that

$$\pi = 5^2 \equiv 3 \pmod{11}, \quad \text{and} \quad \sigma = 2(5 - 5^{-1}) \equiv 2(5 + 2) \equiv 3 \pmod{11}.$$

Then we get $A = \pi(a - 3\sigma) = 3 \cdot (0 - 3 \cdot 3) \equiv 6 \pmod{11}$, so we get that the codomain of Renes' formula is $E' : y^2 = x^3 + 6x^2 + x$.

In Vélu's formula setup, we only have $x_Q = 5$, $y_Q = 8$ and $a_4 = 1$, everything else is 0. Then we have $g_Q^x = -1$, $g_Q^y = 6$, $v_Q = -2$ and $u_Q = 3$. Next, we calculate $v = -2$ and $w = 3 + 5 \cdot (-2) \equiv 4 \pmod{11}$, so then we get the curve

$$E' : y^2 = x^3 + 5.$$

Clearly this is not the same that we got with Renes, but it is still correct, because these curves are isomorphic over $\mathbb{F}_{11}$.

2.4 Orders $\mathcal{O}$ and Ideal Class-groups

In this section, we mostly define a lot of necessary terminology for the coming sections that are from algebraic number theory. To ease the burden of the increased amount of definitions, we illustrate them occasionally by examples. We are also going to prove the most important theorem that are *integral* to understanding CSIDH, so they will mostly concern relevant cases, such as focusing on imaginary quadratic fields and $p \equiv 3 \pmod{4}$. This section is mostly based on [70] and [28]. First we recall some basic definitions and properties of ideals.

Definition 2.4.1. An **ideal** I of a ring R is a non-empty subset $I \subseteq R$ such that I is a group under addition, and we have that

$$\sum_{i=1}^{m} a_i r_i \in I$$

for any $m \in \mathbb{N}$, $a_i \in I$ and $r_i \in R$. An ideal I is called **finitely generated** if there is a finite subset $a_1, \ldots, a_n \in R$ for which $I = \{ \sum_{i=1}^{n} a_i r_j \mid r_j \in R \}$. We denote this ideal as $I = (a_1, \ldots, a_n)$. If the ideal is generated by one element $I = (a)$, we say that the ideal is **principal**.

Definition 2.4.2. An ideal is **proper** if $I \neq R$. A proper ideal I is called **prime** if for any $a, b \in R$, we have that

$$ab \in I \implies a \in I \quad \text{or} \quad b \in I.$$

The **product** of two ideals I and J is defined as

$$IJ = \Big\{ \sum_{i,j} a_i b_j \mid a_i \in I, b_j \in J \Big\}.$$

We continue with the defining modules, giving two equivalent definitions of algebras, and pivoting towards orders.

Definition 2.4.3. Let R be a commutative ring with unity. The **R-module** M is an abelian group, together with a set of axioms

(i) $(r + s)m = rm + sm$,
(ii) $r(m + n) = rm + rn$,
(iii) $r(sm) = (rs)m$,
(iv) $1 \cdot m = m$

for all $r, s \in R$ and $m, n \in M$. A module is called **finitely generated** if there is a finite subset $S = \{m_1, \ldots, m_n\} \in M$ such that every element $m \in M$ can be written as an R-linear combination of S, that is

$$m = r_1 m_1 + r_2 m_2 + \cdots + r_n m_n.$$

A module is called **free** if it has a basis, i.e. its generating set is also linearly independent. If $M \cong R^n$, then n is the **rank** of the module. A subset N of an R-module M is a **submodule** if it is non-empty, and is closed under the same addition and scalar multiplication rule as M.

Definition 2.4.4 (Definition 1)**.** Let R be a commutative ring with unity. An R**-algebra** A is an R-module that also satisfies

$$r \cdot (ab) = (r \cdot a) \cdot b = a \cdot (r \cdot b)$$

for all $a, b \in A$ and $r \in R$.

Definition 2.4.5 (Definition 2)**.** Let R be a commutative ring with unity. An R**-algebra** is a ring A together with a ring homomorphism $f : R \to A$ such that the image $f(R)$ is in the center of A, i.e. the image of f commutes with all elements of A.

Definition 2.4.6. Let D be a squarefree integer, satisfying a monic quadratic equation with integer coefficients. Then the **quadratic field extension** $\mathbb{Q}(\sqrt{D})$ is defined by

$$\mathbb{Q}(\sqrt{D}) := \{\, a + b\sqrt{D} \mid a, b \in \mathbb{Q} \,\}.$$

If $D < 0$, we say that the $\mathbb{Q}(\sqrt{D})$ is an **imaginary quadratic field**. In general, if K is a finite extension of $\mathbb{Q}$, then we say that K is a **number field**.

Definition 2.4.7. Let $K = \mathbb{Q}^n$ be a $\mathbb{Q}$-algebra of finite dimension n. An **order** $\mathcal{O}$ in K is a subring of K that is also a free $\mathbb{Z}$-module of rank n. Additionally, it also satisfies that $K = \mathcal{O} \otimes_Z \mathbb{Q}$.

Example 2.4.8. Let $K = \mathbb{Q}$. The unique example of an order is $\mathbb{Z}$, but it is worth highlighting some nonexamples as well. The ring of even integers $2\mathbb{Z}$ is a principal ideal of $\mathbb{Z}$ generated by (2), but it's not a subring of $\mathbb{Q}$, since it does not contain the multiplicative identity 1. Another nonexample is the set of dyadic rationals $\{\, a/2^n \mid a, n \in \mathbb{Z} \,\}$. One can check that this is a ring, but it is *not* a free $\mathbb{Z}$-module, because it is not finitely generated.

Now we can begin introducing the more interesting parts of algebraic number theory. We begin by giving some more definitions, but these will lead to a theorem that is an essential part in the rest of this thesis.

Definition 2.4.9. An **algebraic integer** is a complex number that is a root of a monic polynomial with coefficients in $\mathbb{Z}$.

Definition 2.4.10. Let K be a number field. The **ring of integers** $\mathcal{O}_K$ is the ring of all algebraic integers contained in K.

Theorem 2.4.11. *[70, Theorem 12.26] Let K be a number field of degree $n = [K : \mathbb{Q}]$. The ring of integers $\mathcal{O}_K$ is the unique maximal order of K.*

Proof. We have two things to prove, that $\mathcal{O}_K$ is an order in K, and that it is maximal. The first part is mostly algebraic manipulation, we refer to the proofs in [50, Theorem 2.1, Corollary 2.30]. We proceed now with showing maximality.

We need to show that every order $\mathcal{O}$ in K is contained in $\mathcal{O}_K$. This is equivalent to showing that every $\alpha \in \mathcal{O}$ is an algebraic integer by the definition of $\mathcal{O}_K$. Since $\mathcal{O}$ is an order, by Definition 2.4.7, We can view it as $\mathbb{Z}^n$, and then take the submodule generated by all powers of α. Let $[\beta_1, \ldots, \beta_n]$ denote the basis for this for this submodule, where the basis elements are $\mathbb{Z}$-linear combinations of powers of α. Set $r > 0$ to be an integer that is larger than any basis elements' exponent of the powers of α, then $\alpha^r = c_1\beta_1 + \cdots + c_r\beta_r$ for some $c_i \in \mathbb{Z}$, since β_i are a basis. Then we have that α is a root of $0 = -\alpha^n + c_1\beta_1 + \cdots + c_n\beta_n$, which is a monic polynomial with integer coefficients. By Definition 2.4.9, α is an algebraic integer. □

Theorem 2.4.12. *[70, Theorem 12.27] Let K be a quadratic field, and $\mathcal{O}_K$ its ring of integers. The orders $\mathcal{O}$ of K are precisely the subrings $\mathbb{Z} + f\mathcal{O}_K$, where f is any positive integer.*

Proof. Since K is quadratic, $\mathcal{O}_K \cong \mathbb{Z}^2$. Denote the basis by $[1, \tau]$ for some $\tau \notin \mathbb{Z}$ and let $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$ for some $f \in \mathbb{Z}_{>0}$. By its definition, $\mathcal{O}$ it is clear that it has rank 2, and that is a subset of the ring $\mathcal{O}_K$, containing 1 (from $\mathbb{Z}$). To show that $\mathcal{O}$ is a ring, it suffices to show that it is closed under multiplication, so let $a + f\alpha$ and $b + f\beta$ be arbitrary elements of $\mathcal{O}$ with $a, b \in \mathbb{Z}$ and $\alpha, \beta \in \mathcal{O}_K$. Their product is then

$$(a + f\alpha)(b + f\beta) = ab + af\beta + bf\alpha + f^2\alpha\beta = ab + f(a\beta + b\alpha + f\alpha\beta) \in \mathcal{O}.$$

By Definition 2.4.7, since $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$ is a subring of K and is a rank 2 $\mathbb{Z}$-module, so it is indeed an order.

To show the converse, i.e. that all orders are present in that form, let $\mathcal{O}$ be any order in K. By definition and Theorem 2.4.11, it is a rank 2 submodule of $\mathcal{O}_K = [1, \tau]$ containing 1 (since it's also a subring of K). This means in turn that $\mathcal{O}$ must contain an integer multiple of τ. Let f be the least positive integer for which this is true, so $f\tau \in \mathcal{O}$. Then the $\mathbb{Z}$-module basis $[1, f\tau]$ lies in $\mathcal{O}$, so $[1, f\tau] \subseteq \mathcal{O}$. To show the converse $\mathcal{O} \subseteq [1, f\tau]$, take any element $\alpha \in \mathcal{O}$. This element must lie in $\mathcal{O}_K$, so it can be expressed in the basis $\alpha = a + b\tau$ for some $a, b \in \mathbb{Z}$. The element $b\tau = \alpha - a$ then lies in $\mathcal{O}$, and the minimality of f from our assumption means that f divides b, so indeed $\mathcal{O} \subseteq [1, f\tau]$. Since both inclusions have been shown, we have that $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$. □

Definition 2.4.13. The integer f in Theorem 2.4.12 is called the **conductor** of the order $\mathcal{O}$. It is also equal to the index $[\mathcal{O}_K : \mathcal{O}]$.

Theorem 2.4.14. *[28, Proposition 2.5.1] Every quadratic field K is of the form $\mathbb{Q}(\sqrt{\Delta})$, where Δ is a square-free integer.*

Proof. Since K is a quadratic field, any element $\tau \in K \backslash \mathbb{Q}$ is of degree 2, and thus is a primitive element of K and thus $K = \mathbb{Q}[\tau]$. Let $m_\tau = x^2 + bx + c$ with $b, c \in \mathbb{Q}$ be the minimal polynomial of τ. Solving this equation gives $2\tau = -b \pm \sqrt{b^2 - 4c}$, so $K = \mathbb{Q}(\sqrt{b^2 - 4c})$. Now $b^2 - 4c$ is a rational number $u/v = uv/v^2$ for some $u, v \in \mathbb{Z}$. We then have that $K = \mathbb{Q}(\sqrt{b^2 - 4c}) = \mathbb{Q}(\sqrt{uv})$, which can reduced to a square-free integer Δ without loss of generality by quotienting out any squares. □

Definition 2.4.15. Let K be a quadratic field and $\mathcal{O} = [1, \tau]$ an order. The **discriminant of** $\mathcal{O}$, denoted by $\Delta = b^2 - 4c$, is the discriminant of the non-integer basis element τ's minimal polynomial $m_\tau = x^2 + bx + c$. If the order is the ring of integers $\mathcal{O}_K$, then we denote it by Δ_K and call it the **fundamental discriminant**, and we have the relation $\Delta = f^2\Delta_K$, where f is the conductor of $\mathcal{O}$.

The following theorem gives the basis for the claims in Section 2.6.

Theorem 2.4.16. *Let $K = \mathbb{Q}(\sqrt{-p})$ be a quadratic field with p prime. The we have that the ring of integers*

$$\mathcal{O}_K \cong \begin{cases} \mathbb{Z}[\sqrt{-p}] & \text{if } p \equiv 1 \pmod 4, \\ \mathbb{Z}[\frac{1+\sqrt{-p}}{2}] & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

Proof. Every element of $\mathbb{Q}(\sqrt{-p})$ can be expressed as $a + b\sqrt{-p}$ for some $a, b \in \mathbb{Q}$, so let $\alpha = a + b\sqrt{-p}$, and let $\overline{\alpha}$ denote the complex conjugate. We then have

$$\alpha + \overline{\alpha} = 2a \in \mathbb{Q} \quad \text{and} \quad \alpha\overline{\alpha} = a^2 + pb^2 \in \mathbb{Q}.$$

By basic number theory, we know that $\mathbb{Z}$ is integrally closed, so we must have also that $2a \in \mathbb{Z}$ and $a^2 + pb^2 \in \mathbb{Z}$. Note that α and $\overline{\alpha}$ are roots of $f(x) = (x - \alpha)(x - \overline{\alpha}) = x^2 - 2ax + a^2 + pb^2 = 0$, so in particular α is an algebraic integer, and hence is in $\mathcal{O}_K$. We have the following chain of implications

$$a^2 + pb^2 \in \mathbb{Z} \implies 4(a^2 + pb^2) \in \mathbb{Z} \implies (2a)^2 + p(2b)^2 \in \mathbb{Z} \implies p(2b)^2 \in \mathbb{Z},$$

where the last implication holds since $2a \in \mathbb{Z}$.

We now show that $2b \in \mathbb{Z}$ as well. We proceed by contradiction, and assume that $2b \notin \mathbb{Z}$. Then $2b = b'/q$ with q prime, and $(2b)^2 = b'^2/q^2$. However, we know that $p(2b)^2 \in \mathbb{Z}$, but pb'^2/q^2 is not an

integer because the denominator cannot be cleared. We can conclude that $2b \in \mathbb{Z}$. Now let $2a = u$ and $2b = v$, with $u, v \in \mathbb{Z}$. The condition from before becomes then

$$u^2 + pv^2 \in 4\mathbb{Z}.$$

We have now two cases: If v is even, then u is too. In this case $a, b \in \mathbb{Z}$. If v is odd, then $v^2 \equiv 1 \pmod 4$, and the only possibilities for u^2 are 0 or 1, since these are the only squares modulo 4. Since p is square-free, it is not a multiple of 4, so necessarily $u^2 \equiv 1 \pmod 4$ and $p \equiv 3 \pmod 4$. This shows that $a = b = 1/2$, proving the theorem. □

Let us investigate an interesting property of quadratic fields, which is illustrated by a classic example.

Example 2.4.17. Let $K = \mathbb{Q}(\sqrt{-5})$, and let's see what happens in its ring of integers $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$. We are used to the fact that $\mathbb{Z}$ is a unique factorization domain, which means that every number can be *uniquely* decomposed into prime powers. This, however, is not true for many orders (actually only 9). To see why, note that 6 has the decompositions

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

It is easy to see that $2, 3$ and $1 + \sqrt{-5}, 1 - \sqrt{-5}$ cannot be expressed as multiples of each other, so this is indeed 2 distinct ways to decompose the integer 6.

There is a way to get back unique factorization, but it cannot be at the level of integers, but rather at the level of ideals. But first, let us describe how to measure the sizes of $\mathcal{O}$-ideals

Definition 2.4.18. Let $\mathcal{O}$ be an order in a number field K. An **$\mathcal{O}$-ideal** $\mathfrak{a}$ is an ideal of $\mathcal{O}$, so in particular, for all $\alpha \in \mathfrak{a}$ and $b \in \mathcal{O}$, we have that $b\alpha \in \mathfrak{a}$. The **norm** of $\mathfrak{a}$ is defined by

$$\mathrm{N}(\mathfrak{a}) = [\mathcal{O} : \mathfrak{a}] = \gcd(\{\, \mathrm{N}(\alpha) \mid \alpha \in \mathfrak{a} \,\}),$$

where the norm of an element $\alpha \in \mathfrak{a}$ is $\det(T_\alpha)$ the determinant of the element as a linear operator $T_\alpha : K \to K$ sending $x \mapsto \alpha x$.

Remark 2.4.19. We might assume from our experience with norms that this norm would also be multiplicative, however, this is not always true. For example, if we take the order $\mathcal{O} = [1, 2i]$ and the ideal $\mathfrak{a} = [2, 2i]$, then we have that $\mathrm{N}(\mathfrak{a}) = [\mathcal{O} : \mathfrak{a}] = 2$. However, $\mathfrak{a}^2 = [4, 4i]$ with $\mathrm{N}\left(\mathfrak{a}^2\right) = 8$, so we have that

$$\mathrm{N}\left(\mathfrak{a}^2\right) = 8 \neq 2^2 = \mathrm{N}(\mathfrak{a})^2.$$

For the norm to be multiplicative, we need some extra property of the $\mathcal{O}$-ideals, but the norm is multiplicative for scalar multiplication.

Lemma 2.4.20. *[70, Corollary 17.7.] Let $\mathcal{O}$ be an order in a number field, let $\lambda \in \mathcal{O}$ be nonzero, and let $\mathfrak{a}$ be a nonzero $\mathcal{O}$-ideal. Then*

$$\mathrm{N}(\lambda\mathfrak{a}) = \mathrm{N}(\lambda)\,\mathrm{N}(\mathfrak{a}).$$

Proof.

$$\mathrm{N}(\lambda\mathfrak{a}) = [\mathcal{O} : \lambda\mathfrak{a}] = [\mathcal{O} : \mathfrak{a}][\mathfrak{a} : \lambda\mathfrak{a}] = [\mathcal{O} : \mathfrak{a}][\mathcal{O} : \lambda\mathcal{O}] = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\lambda) = \mathrm{N}(\lambda)\,\mathrm{N}(\mathfrak{a}).$$

□

Definition 2.4.21. A **fractional $\mathcal{O}$-ideal** of $\mathcal{O}$ is an $\mathcal{O}$-submodule of K of the form $\mathfrak{b} = \lambda\mathfrak{a} = \{\, \lambda\alpha \mid \alpha \in \mathfrak{a} \,\}$, where $\lambda \in K^\times$ is nonzero, and $\mathfrak{a}$ is an $\mathcal{O}$-ideal. We say that a fractional $\mathcal{O}$-ideal $\mathfrak{b}$ is **invertible** if there exists a fractional $\mathcal{O}$-ideal $\mathfrak{b}^{-1}$ such that $\mathfrak{b}\mathfrak{b}^{-1} = \mathcal{O}$.

Lemma 2.4.22. *Let $\mathcal{O}$ be an order in a number field. The norm of a fractional $\mathcal{O}$-ideal $\mathfrak{b} = \frac{1}{b}\mathfrak{a}$ with some $b \in \mathbb{Z}_{>0}$ is*

$$\mathrm{N}(\mathfrak{b}) := \frac{\mathrm{N}(\mathfrak{a})}{\mathrm{N}(b)} \in \mathbb{Q}^\times_{>0}.$$

Proof. Apply Lemma 2.4.20, which also ensures that the norm of $\mathfrak{b}$ does not depend on the choice of $\mathfrak{a}$ or b. □

Fractional ideals can be multiplied as one might expect, that is $\mathfrak{b}\mathfrak{b}' = (\lambda\lambda')\mathfrak{a}\mathfrak{a}'$, where $\mathfrak{a}\mathfrak{a}'$ (and in the statement of the definitions as well) is the product of ideals as defined in Definition 2.4.2. Without loss of generality for number fields K, one can also assume that $\lambda = 1/b$ for some $b \in \mathbb{Z}$. The following example shows that not every fractional $\mathcal{O}$-ideal is invertible.

Example 2.4.23. Let $p \equiv 3 \pmod 4$, and take $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$. Let $\mathfrak{l}_2 = (2, 1+\sqrt{-p})$ be an ideal in $\mathcal{O}$. It is obvious that $\mathfrak{l}_2 \neq 2\mathcal{O} = (2)$. In light of Remark 2.4.19, let us see what happens when we square $\mathfrak{l}_2$:

$$\begin{aligned}\mathfrak{l}_2^2 &= (2, 1+\sqrt{-p}) \cdot (2, 1+\sqrt{-p}) \\ &= (2\cdot 2, \quad 2(1+\sqrt{-p}), \quad (1+\sqrt{-p})^2) \\ &= (4, \quad 2(1+\sqrt{-p}), \quad 1-p+2\sqrt{-p}).\end{aligned}$$

Since $p \equiv 3 \pmod 4$, we can write $p = 4k+3$ for some integer k. Substituting this into the third generator gives

$$1 - p + 2\sqrt{-p} = 1 - (4k+3) + 2\sqrt{-p} = -4k - 2 + 2\sqrt{-p}.$$

Notice that this can be expressed by the other two generators as

$$-4k - 2 + 2\sqrt{-p} = -4k - 4 + 2 + 2\sqrt{-p} = 4(k-1) + 2 + 2\sqrt{-p},$$

so we have that

$$\mathfrak{l}_2^2 = (4, 2(1+\sqrt{-p})) = 2(2, 1+\sqrt{-p}) = 2\mathfrak{l}_2.$$

If $\mathfrak{l}_2$ is invertible, then we would have that

$$\mathfrak{l}_2^2\mathfrak{l}_2^{-1} = 2\mathfrak{l}_2\mathfrak{l}_2^{-1} \implies \mathfrak{l}_2 = 2\mathcal{O},$$

which is a contradiction, therefore $\mathfrak{l}_2$ is not an invertible fractional ideal in $\mathcal{O}$.

Definition 2.4.24. Let $\mathcal{O}$ be an order in an imaginary quadratic field K with $\alpha \in K$. Then a fractional $\mathcal{O}$-ideal $\mathfrak{b}$ is called **proper**, if it holds that

$$\mathcal{O}(\mathfrak{b}) := \{\, \alpha \mid \alpha\mathfrak{b} \subseteq \mathfrak{b} \,\} = \mathcal{O}.$$

Lemma 2.4.25. *[70, Lemma 17.9] Let $\mathfrak{b} = \lambda\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal. Then $\mathfrak{a}$ is proper if and only if $\mathfrak{b}$ is proper, and $\mathfrak{a}$ is invertible if and only if $\mathfrak{b}$ is invertible.*

Proof. For the first statement, we have that

$$\{\, \alpha \mid \alpha\mathfrak{b} \subseteq \mathfrak{b} \,\} = \{\, \alpha \mid \alpha\lambda\mathfrak{a} \subseteq \lambda\mathfrak{a} \,\} = \{\, \alpha \mid \alpha\mathfrak{a} \subseteq \mathfrak{a} \,\}.$$

For the second statement, if $\mathfrak{a}$ is invertible, then $\mathfrak{b}^{-1} = \lambda^{-1}\mathfrak{a}^{-1}$, and if $\mathfrak{b}$ is invertible then $\mathfrak{a}^{-1} = \lambda\mathfrak{b}^{-1}$, since $\mathfrak{a}\mathfrak{a}^{-1} = \mathfrak{a}\lambda\mathfrak{b}^{-1} = \mathfrak{b}\mathfrak{b}^{-1} = \mathcal{O}$. □

We can now prove that invertible $\mathcal{O}$-ideals are equivalent to proper $\mathcal{O}$-ideals, and then we will also see that the norm is multiplicative for proper $\mathcal{O}$-ideals.

Theorem 2.4.26. *[70, Theorem 17.10] Let $\mathcal{O}$ be an order in an imaginary quadratic field K and let $\mathfrak{a} = [\alpha, \beta]$ be an $\mathcal{O}$-ideal. Then $\mathfrak{a}$ is proper if and only if $\mathfrak{a}$ is invertible. If $\mathfrak{a}$ is invertible, then the inverse of $\mathfrak{a}$ is the fractional $\mathcal{O}$-ideal $\mathfrak{a}^{-1} = \frac{1}{\mathrm{N}(\mathfrak{a})}\overline{\mathfrak{a}}$, where $\overline{\mathfrak{a}} = [\overline{\alpha}, \overline{\beta}]$ denotes the complex conjugate.*

Proof. ($\Longleftarrow$) Assume $\mathfrak{a}$ is invertible, then for any $\gamma \in K$, we have

$$\gamma\mathfrak{a} \subseteq \mathfrak{a} \implies \gamma\mathfrak{a}\mathfrak{a}^{-1} \subseteq \mathfrak{a}\mathfrak{a}^{-1} \implies \gamma\mathcal{O} \subseteq \mathcal{O} \implies \gamma \in \mathcal{O},$$

so we have that $\mathcal{O}(\mathfrak{a}) \subseteq \mathcal{O}$, and we always have that $\mathcal{O} \subseteq \mathcal{O}(\mathfrak{a})$, which means that $\mathfrak{a}$ is a proper $\mathcal{O}$-ideal by Definition 2.4.24.
($\implies$) Assume that $\mathfrak{a} = [\alpha, \beta]$ is proper. Take $\tau = \beta/\alpha$, so that we can express $\mathfrak{a}$ as fractional $\mathcal{O}$-ideal $\alpha[1, \tau]$, and let $m_\tau(x) = ax^2 + bx + c \in \mathbb{Z}[x]$ be the minimal polynomial of τ. Then, by Lemma 2.4.25, we have that $[1, \tau]$ is proper as well, so if we let $\mathcal{O} = [1, \omega]$ be another $\mathcal{O}$-ideal, then $\omega \in [1, \tau]$ and $\omega = m + n\tau$ for some $m, n \in \mathbb{Z}$. By translating ω, we may assume $\omega = n\tau$. We also have that $\omega\tau = n\tau^2 \in [1, \tau]$, since $[1, \tau]$ is an $\mathcal{O}$-module. This implies $a|n$ in the minimal polynomial m_τ, since a is chosen as minimal. Furthermore, we also have that $a\tau[1, \tau] \subseteq [1, \tau]$, so

$$a\tau \in \mathcal{O}([1, \tau]) = \mathcal{O} = [1, \omega] = [1, n\tau] \implies \mathcal{O} = [1, a\tau],$$

because we must have $n = a$. Then we see that

$$\mathrm{N}(\mathfrak{a}) = [\mathcal{O} : \mathfrak{a}] = \big[[1, a\tau] : \alpha[1, \tau]\big] = \frac{1}{a}\big[[1, a\tau] : \alpha[1, a\tau]\big] = \frac{1}{a}[\mathcal{O} : \alpha\mathcal{O}] = \frac{\mathrm{N}(\alpha)}{a}.$$

Next, we investigate what happens when we multiply $\mathfrak{a}$ by its conjugate.

$$\mathfrak{a}\overline{\mathfrak{a}} = \alpha[1, \tau]\overline{\alpha}[1, \overline{\tau}] = \mathrm{N}(\alpha)\,[1, \tau, \overline{\tau}, \tau\overline{\tau}].$$

From the minimal polynomial $m_\tau(x) = ax^2 + bx + c$, we have that $\tau + \overline{\tau} = -b/a$ and $\tau\overline{\tau} = c/a$, so we have

$$\mathfrak{a}\overline{\mathfrak{a}} = \mathrm{N}(\alpha)\,[1, \tau, \overline{\tau}, \tau\overline{\tau}] = \frac{\mathrm{N}(\alpha)}{a}[a, a\tau, -b, c] = \mathrm{N}(\mathfrak{a})\,[1, a\tau] = \mathrm{N}(\mathfrak{a})\,\mathcal{O} = (\mathrm{N}(\mathfrak{a})),$$

where $(\mathrm{N}(\mathfrak{a}))$ is the principal $\mathcal{O}$-ideal generated by the integer $\mathrm{N}(\mathfrak{a})$. Rearranging this equality gives the statement of the theorem. □

Corollary 2.4.27. *[70, Corollary 17.12.] Let $\mathcal{O}$ be an order in an imaginary quadratic field, and let $\mathfrak{a}, \mathfrak{b}$ be two proper (or invertible) fractional $\mathcal{O}$-ideals. Then*

$$\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b}).$$

Proof. Let $\mathfrak{a} = \frac{1}{a}\mathfrak{a}'$ and $\mathfrak{b} = \frac{1}{b}\mathfrak{b}'$ with $a, b \in \mathbb{Z}_{>0}$ and $\mathfrak{a}', \mathfrak{b}' \subseteq \mathcal{O}$. Then, by Lemma 2.4.20, we have that

$$\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \frac{\mathrm{N}(\mathfrak{a}'\mathfrak{b}')}{\mathrm{N}(a)\,\mathrm{N}(b)}.$$

Therefore without loss of generality, it is enough to show the equality for invertible $\mathcal{O}$-ideals. By Theorem 2.4.26, we then have

$$(\mathrm{N}(\mathfrak{a}\mathfrak{b})) = \mathfrak{a}\mathfrak{b}\overline{\mathfrak{a}\mathfrak{b}} = \mathfrak{a}\mathfrak{b}\overline{\mathfrak{a}}\,\overline{\mathfrak{b}} = \mathfrak{a}\overline{\mathfrak{a}}\mathfrak{b}\overline{\mathfrak{b}} = (\mathrm{N}(\mathfrak{a}))(\mathrm{N}(\mathfrak{b})),$$

from which it follows that $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$ since they are all integers. □

Now we can give the definition of the class group, which is the most important part of the mechanism of CSIDH, which will be explored later in Section 2.7.

Definition 2.4.28. Let $\mathcal{I}(\mathcal{O})$ denote the group of invertible fractional $\mathcal{O}$-ideals, and $\mathcal{P}(\mathcal{O})$ denote the subgroup of $\mathcal{I}(\mathcal{O})$ that is the group of principal fractional $\mathcal{O}$-ideals. The **ideal class group** $\mathrm{cl}(\mathcal{O})$ is the quotient of these groups:

$$\mathrm{cl}(\mathcal{O}) = \mathcal{I}(\mathcal{O})/\mathcal{P}(\mathcal{O}).$$

The size of the ideal class group is finite, and is denoted by $h(\mathcal{O})$, and it is called the **class number** of $\mathcal{O}$.

This definition could be a theorem as well, because maybe it is not obvious that this modulo operation is even well defined. However, it is easy to see that $\mathcal{I}(\mathcal{O})$ is an abelian group based on the fact that the elements are invertible, $\mathcal{O}$ acts as the neutral element, and associativity + commutativity comes from the Definition 2.4.2 of ideal multiplication. Then from basic group theory, we have that every subgroup of an abelian group is normal, so the quotient group exists and is well-defined.

Let us now investigate what happens to ideals in general inside orders of number fields, as that is the topic of our thesis, and we have already seen that unique factorization is not guaranteed.

Theorem 2.4.29. *Let K be a number field. Then any nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$ can be written as a product*

$$\mathfrak{a} = \mathfrak{p}_1 \dots \mathfrak{p}_r$$

of prime ideals, and the decomposition is unique up to order. Furthermore, the $\mathfrak{p}_i$'s are exactly the prime ideals of $\mathcal{O}_K$ containing $\mathfrak{a}$.

Proof. [28, Corollary 5.6] □

Example 2.4.30. Continuing Example 2.4.17, let us see if (2) is a prime ideal in $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$. From ring theory, one way to check is to see if the quotient of $\mathcal{O}_K/(2)$ is an integral domain. We can represent $\mathcal{O}_K$ in a polynomial ring gives

$$\mathbb{Z}[\sqrt{-p}] \cong \mathbb{Z}[x]/(x^2+5),$$

then quotienting out by the ideal (2) gives

$$\mathbb{Z}[x]/(2, x^2+5) \cong \mathbb{Z}_2[x]/(x^2+5) \cong \mathbb{Z}_2[x]/(x+1)^2.$$

Taking the nonzero element $x+1$, and squaring it gives $(x+1)^2 = 0$ because it is in the kernel, so we have found a zero divisor, therefore the quotient ring is not an integral domain, so (2) is not a prime ideal. In fact, since x in the polynomial ring corresponds to $\sqrt{-5}$, we have actually found that

$$(2) = \mathfrak{p}^2 = (2, 1+\sqrt{-5})$$

ramifies into the prime ideal $\mathfrak{p}$. Note that $\mathfrak{p}$ contains (2) because of Theorem 2.4.29.

In Example 2.4.17 and this example, we have seen that 6 could be decomposed into two distinct products, and that (2) also splits into two *nonprincipal* prime ideals. This shows that $\mathcal{O} = \mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain. The ideal class group is the perfect tool for these situations, because the class number $h(\mathcal{O})$ measures the failure of unique factorization.

Example 2.4.31. Let us determine the class number of $\mathcal{O} = \mathbb{Z}[\sqrt{-5}]$. We always have the identity class $[\mathcal{O}]$ of all principal ideals. We have seen in Example 2.4.23 that $(2) = \mathfrak{p}^2$ with $\mathfrak{p} = (2, 1+\sqrt{-5})$, since $-5 \equiv 3 \pmod 4$.

To see that it's not principal, assume $\mathfrak{p}_2 = (a + b\sqrt{-5})$ for some $a, b \in \mathbb{Z}$. Taking the norms gives us

$$\mathrm{N}(\mathfrak{p}_2) = \mathrm{N}\left(a + b\sqrt{-5}\right) = a^2 + 5b^2 \quad \text{and} \quad \mathrm{N}(\mathfrak{p}) = 2.$$

This means that we must have $a^2 + 5b^2 = 2$, but this equation has no integer solutions, so $\mathfrak{p}$ is indeed not a principal ideal. As a result, we have that $\{\,[\mathcal{O}], [\mathfrak{p}]\,\} \subseteq \mathrm{cl}(\mathcal{O})$. The order of $\mathfrak{p}$ is 2 in $\mathrm{cl}(\mathcal{O})$ since $\mathfrak{p}^2 = (2)$, which is in the trivial class of principal ideals. Further direct calculation would show that these are the only two options, or one can use upper bounds (such as the Minkowski bound) to limit the norm of prime ideals considered. Either way, we would arrive at the conclusion that the ideal class group has only these two elements, thus $h(\mathcal{O}) = h(\mathbb{Z}[\sqrt{-p}]) = 2$.

Remark 2.4.32. The record biggest class number and class group structure calculated is actually the class group of the CSIDH-512 prime. The authors of the digital signature CSI-FiSh [5] have calculated that the prime ideal $\mathfrak{l}_3 = (3, \sqrt{-p}+1)$ is a generator of $\mathrm{cl}(\mathcal{O})$, and the size of it is

$$h(\mathcal{O}) = 254652442229484275177030186010639202161620514305486423592570860975597611726191,$$

or equivalently $\sim 2^{257.1}$.

We will finish this section by recording two useful theorems that will be relevant in the coming discussions. But first, recall the definition of the Legendre symbol:

Definition 2.4.33. The **Legendre symbol** for an odd prime p is defined as

$$\left(\frac{x}{p}\right) = \begin{cases} +1 & \text{if } t^2 \equiv x \pmod{p} \text{ has a solution } t \not\equiv 0 \pmod{p}, \\ -1 & \text{if } t^2 \equiv x \pmod{p} \text{ has no solution } t, \\ 0 & \text{if } x \equiv 0 \pmod{p}. \end{cases}$$

This symbol is only defined for odd primes in the bottom slot, but it can be extended to all integers, which is called the Kronecker symbol.

Definition 2.4.34. The **Kronecker symbol** for odd primes p is defined exactly the same way as the Legendre symbol. The remaining cases are

$$\left(\frac{a}{2}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{2}, \\ 1 & \text{if } a \equiv \pm 1 \pmod{8}, \\ -1 & \text{otherwise.} \end{cases} \quad \left(\frac{a}{-1}\right) = \begin{cases} 1 & \text{if } a \geq 0, \\ -1 & \text{if } a < 0, \end{cases} \quad \text{and} \quad \left(\frac{a}{0}\right) = \begin{cases} 1 & \text{if } a = \pm 1, \\ 01 & \text{otherwise.} \end{cases}$$

Let us now see a couple important theorems that use the Kronecker symbol, but one that also gives an understanding for why the polynomial ring method worked in Example 2.4.30.

Theorem 2.4.35. *Let K be a quadratic field of discriminant Δ_K, and let p be a prime in $\mathbb{Z}$ and $\left(\frac{a}{n}\right)$ denote the Kronecker symbol. Then we have that*

(i) If $\left(\frac{\Delta_K}{p}\right) = 0$, (i.e. $p \mid \Delta_K$), then $p\mathcal{O}_K = \mathfrak{p}^2$ for some prime ideal $\mathfrak{p} \in \mathcal{O}_K$.
(ii) If $\left(\frac{\Delta_K}{p}\right) = 1$, then $p\mathcal{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$, where $\mathfrak{p} \neq \overline{\mathfrak{p}}$ are prime in $\mathcal{O}_K$.
(iii) If $\left(\frac{\Delta_K}{p}\right) = -1$, then $p\mathcal{O}_K$ is prime in $\mathcal{O}_K$.

Proof. See [28] □

There is a bijection between the behaviour of the Kronecker symbol $\left(\frac{\Delta}{p}\right)$ and how the minimal polynomial of the order behaves in modulo p. This is proved in [28, Proposition 5.11], and is generally known as the Kummer-Dedekind Theorem. This gives rise to the following definition.

Definition 2.4.36. Let $K = \mathbb{Q}(\sqrt{\Delta})$ be an imaginary quadratic field. Then we say that a prime p

- **splits** if $x^2 - \Delta \equiv (x-a)(x-b) \pmod{p}$ for some $\alpha \neq \beta \in \mathbb{F}_p$;
- **ramifies** if $x^2 - \Delta \equiv (x-a)^2 \pmod{p}$ for some $a \in \mathbb{F}_p$;
- is **inert** if $x^2 - \Delta$ is irreducible over $\mathbb{F}_p[x]$.

We continue with some important theorems, but since they are quite involved, we only refer to the proofs to Cox's book, if the reader is interested in learning more.

Theorem 2.4.37. *Let K be an imaginary quadratic field with fundamental discriminant Δ_K, and let $\mathcal{O}$ be an order with conductor f. Then the class number of the non-maximal order $h(\mathcal{O})$ can be related to the class number of the ring of integers $h(\mathcal{O}_K)$ by*

$$h(\mathcal{O}) = \frac{h(\mathcal{O}_K) f}{[\mathcal{O}_K^\times : \mathcal{O}^\times]} \prod_{p \mid f} \left(1 - \left(\frac{\Delta_K}{p}\right)\frac{1}{p}\right),$$

where $[\mathcal{O}_K^\times : \mathcal{O}^\times]$ denotes the index of the group of units, and $\left(\frac{\Delta_K}{p}\right)$ is the Kronecker symbol.

Proof. See [28, Theorem 7.24] □

Corollary 2.4.38. *Let $p \equiv 3 \pmod{8}$ with $p > 3$, and let $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$. Then we have that*

$$h(\mathcal{O}) = 3h(\mathcal{O}_K).$$

Proof. We will apply 2.4.37, so let us record the things we know. From Theorem 2.4.16, we see that the ring of integers is $\mathbb{Z}[\frac{1+\sqrt{-p}}{2}]$ for our case, so $f = [\mathcal{O}_K : \mathcal{O}] = 2$, and $\Delta = 4\Delta_K$. Since we have $p > 3$, the units of both orders are $\{\pm 1\}$, so the index of them is 1. Since the minimal polynomial of $\sqrt{-p} \in \mathcal{O}$ is $x^2 + p$, we have that $\Delta = -4p$, therefore $\Delta_K = -p$. Since the conductor is 2, that is the only prime to check in the Kronecker symbol. We have thus

$$\left(\frac{\Delta_K}{2}\right) = \left(\frac{-p}{2}\right) = \left(\frac{-3 \pmod 8}{2}\right) = -1.$$

Now we have all the ingredients, so we can substitute everything into the formula to get

$$h(\mathcal{O}) = \frac{h(\mathcal{O}_K) \cdot 2}{1}\left(1 - (-1)\frac{1}{2}\right) = 3h(\mathcal{O}_K)$$

□

Theorem 2.4.39 (Siegel's approximation)**.** *Let K be an imaginary quadratic field with fundamental discriminant Δ_K. then we have that*

$$\lim_{\Delta_K \to -\infty} \frac{\log h(\Delta_K)}{\log |\Delta_K|} = \frac{1}{2},$$

or in other words,

$$h(\mathcal{O}_K) \approx \sqrt{|\Delta_K|} \quad \text{as} \quad \Delta_K \longrightarrow -\infty.$$

Proof. See [68]. (Note! It's in German.) □

Now that we have gotten more familiar with how orders work and what the class group is, we can relate them to elliptic curves, and get one step closer to understanding CSIDH.

2.5 Endomorphism rings

In the previous sections, we have developed the basic theory of isogenies between two different elliptic curves, and a seemingly unrelated set of tools from algebraic number theory. Up until this point, sometimes we used the same curve as both the domain and image of an isogeny. In this section, we will investigate the structure and properties of isogenies that are defined on one elliptic curve, and find some deeper connections with number fields. It is important to note that the size of the set of endomorphism (and its properties) clearly depends on the field that we are defining both the elliptic curve and the endomorphisms over. These differences are discussed in the second half of this section, together with a complete characterization of the $\mathbb{F}_p$-rational endomorphism rings for the finite field $\mathbb{F}_p$. This section is mostly based on [69, 70].

Let us start by defining what an endomorphism is in the context of elliptic curves, and some other notions that will be relevant later.

Definition 2.5.1. An isogeny $\varphi : E(k) \to E(k)$ defined over the field k is called an **endomorphism**, denoted by $\mathrm{End}_k(E)$. An endomorphism that is also an isomorphism is called an **automorphism**, denoted by $\mathrm{Aut}_k(E)$.

Definition 2.5.2. Let $\alpha \in \mathrm{End}(E)$ be an endomorphism. The **norm** and **trace** of α are respectively defined as

$$\mathrm{N}(\alpha) = \alpha \cdot \hat{\alpha} \quad \text{and} \quad \mathrm{tr}(\alpha) = \alpha + \hat{\alpha}.$$

For ease of notation, we will simply denote the set of endomorphism defined over the algebraic closure of k as $\mathrm{End}(E)$ instead of $\mathrm{End}_{\overline{k}}(E)$, and denote the $\mathbb{F}_p$-rational endomorphism set over $\mathbb{F}_p$ by $\mathrm{End}_p(E)$. We have already seen two examples of endomorphisms: the multiplication-by-m map, in particular the doubling map $m = 2$ in Example 2.2.8, and in the case of finite fields, the Frobenius endomorphism in Definition 2.2.9. Note that we sometimes write n instead of $[n]$. This decision will become apparent (and proved) in Lemma 2.5.16.

We continue by recording some easier properties of the above definitions.

Proposition 2.5.3. *Let E be an elliptic curve and $\alpha \in \mathrm{End}(E)$ an endomorphism. The norm and trace of α are multiplication-by-m maps for some $m \in \mathbb{Z}$, and in particular*

$$\mathrm{N}(\alpha) = \deg \alpha \quad \text{and} \quad \mathrm{tr}(\alpha) = 1 + \deg \alpha - \deg(1 - \alpha).$$

Proof. By definition, we have that

$$\mathrm{N}(\alpha) = \alpha\hat{\alpha} = \alpha \circ \hat{\alpha} = [\deg \alpha],$$

where the last equation follows from Definition 2.2.20, and clearly $\deg \alpha \in \mathbb{Z}$. Next, we investigate what happens when we look at the degree of the endomorphism $[1] - \alpha$.

$$\begin{aligned} \deg([1] - \alpha) = (\widehat{[1] - \alpha})([1] - \alpha) &= [1] - \alpha - \hat{\alpha} + \alpha\hat{\alpha} \\ &= [1] - \mathrm{tr}(\alpha) + \mathrm{N}(\alpha). \end{aligned}$$

We have shown just now that $\mathrm{N}(\alpha) = \deg \alpha$, so reordering the equation gives the desired equality $\mathrm{tr}(\alpha) = 1 + \deg \alpha - \deg(1 - \alpha)$. □

Corollary 2.5.4. *Let $\pi \in \mathrm{End}_p(E)$ be the Frobenius endomorphism of E defined over the finite field $\mathbb{F}_p$. Then*

$$t := \mathrm{tr}(\pi) = 1 + p - \#E(\mathbb{F}_p).$$

Proof. Recall that $\deg \pi = p$ (Definition 2.2.9) From Proposition 2.5.3, we have that

$$\mathrm{tr}(\pi) = 1 + \deg \pi - \deg(1 - \pi) = 1 + p - \deg(1 - \pi).$$

Next, we investigate the endomorphism $[1] - \pi$. If we look at the kernel, it becomes clear that

$$\ker(1 - \pi) = \{P \in E(\overline{\mathbb{F}}_p) \mid P - \pi(P) = \infty\} = \{P \in E(\overline{\mathbb{F}}_p) \mid P = \pi(P)\} = \#E(\mathbb{F}_p),$$

where the last equation holds since in characteristic p, we have that $x^p = x$ implies $x \in \mathbb{F}_p$. Now all we need to show is that $\deg(1 - \pi) = \ker(1 - \pi)$, which is the same as showing that the endomorphism $1 - \pi$ is separable.

We proceed by contradiction. Assume that $1 - \pi$ is inseparable. From Definition 2.2.9, we know that π is purely inseparable, which means that by Theorem 2.2.14, we have that

$$[1] = \pi + (1 - \pi) = 1 \circ \pi + (1 - \pi)_{\mathrm{sep}} \circ \pi^n = (1 + (1 - \pi)_{\mathrm{sep}} \circ \pi^{n-1}) \circ \pi,$$

for some integer $n > 0$. This is a contradiction, since the above equation says that $[1]$ is an inseparable isogeny, but we have shown in Example 2.2.10 that 1-isogenies are separable. As a result, we conclude that $1 - \pi$ is separable, and thus

$$\mathrm{tr}(\pi) = 1 + p - \deg(1 - \pi) = 1 + p - \ker(1 - \pi) = 1 + p - \#E(\mathbb{F}_p).$$ □

Remark 2.5.5. From now on, the letter t will be exclusively reserved for the trace of the p-Frobenius endomorphism $\pi = (x^p, y^p)$, as alluded to in the statement of Corollary 2.5.4.

Proposition 2.5.6. *Let $\alpha \in \mathrm{End}(E)$ be an endomorphism. Then both α and its dual $\hat{\alpha}$ satisfies the following polynomial:*

$$x^2 - \mathrm{tr}(\alpha)\, x + \deg \alpha = 0.$$

Proof. Substituting in α for x gives

$$\alpha^2 - \mathrm{tr}(\alpha)\, \alpha + \deg \alpha = \alpha^2 - (\alpha + \hat{\alpha})\alpha + \alpha\hat{\alpha} = 0.$$

The same clearly holds for $\hat{\alpha}$ as well. □

Now we have seen some properties of endomorphisms and their traces, which is enough now to prove an important theorem in the study of elliptic curves, Hasse's theorem, which was already stated in Section 2.1. The theorem itself holds over $\mathbb{F}_q$ where $q = p^n$ is some prime power of the characteristic, but in light of CSIDH utilizing the base prime field $\mathbb{F}_p$, we prove it specifically for that.

Theorem 2.5.7 (Hasse's Theorem for $\mathbb{F}_p$)**.** *Let $E(\mathbb{F}_p)$ be a an elliptic curve defined over the prime field $\mathbb{F}_p$, and t be the trace of Frobenius endomorphism* $\operatorname{tr}(\pi)$*. Then*

$$\#E(\mathbb{F}_p) = p + 1 - t,$$

where $|t| \leq 2\sqrt{p}$.

Proof. We have already shown the first part of the statement in Corollary 2.5.4, so the only part we have left is proving that $|t| \leq 2\sqrt{p}$. We have also seen in the proof of Corollary 2.5.4, that $\#E(\mathbb{F}_p) = \ker(\pi - 1)$, so let us investigate what happens in general if we take the degree of the general endomorphism $r\pi - s$ for some $r, s \in \mathbb{Z}$. Since every endomorphism commutes with $\mathbb{Z}$, we have

$$\begin{aligned}\deg(r\pi - s) = \widehat{(r\pi - s)}(r\pi - s) &= (\hat{\pi}\hat{r} - \hat{s})(r\pi - s) \\ &= r^2 \deg \pi - rs \cdot (\hat{\pi} + \pi) + s^2 \\ &= r^2 p - rs \cdot t + s^2.\end{aligned}$$

Since the degree map is positive definite on $\operatorname{End}(E)$, we have that $\deg \alpha \geq 0$ for all $\alpha \in \operatorname{End}(E)$, so in particular we have that

$$\begin{aligned}&0 \leq r^2 p - rs \cdot t + s^2 \\ \implies &0 \leq \left(\frac{r}{s}\right)^2 p - \left(\frac{r}{s}\right) t + 1.\end{aligned}$$

Since this equation is true for all $r, s \in \mathbb{Z}$, and $\mathbb{Q}$ is dense in $\mathbb{R}$, we can extend this inequality to $\mathbb{R}$ by

$$0 \leq px^2 - tx + 1$$

for all $x \in \mathbb{R}$. This formula being nonnegative for all variables is equivalent to the discriminant being nonpositive, since then the function has no real roots. The discriminant is $t^2 - 4p$, so we get

$$0 \geq t^2 - 4p \implies |t| \leq 2\sqrt{p}. \qquad \square$$

As mentioned in a previous discussion, the heuristically expected number of k-rational points of $E(\mathbb{F}_p)$ is $p + 1$. This turns out to be the same as expecting any curve $E(\mathbb{F}_p)$ to be supersingular. In fact, if we take a large prime p, and simulate randomly picking two Weierstrass coefficients $A, B \in \mathbb{F}_p$, then recording the value of the trace of Frobenius t, then we will see that the distribution of t will look like half a circle, see Figure 5. As $p \to \infty$, this distribution is approaching a distribution called the *Sato-Tate distribution*, and constitutes the basis of the Sato-Tate conjecture, but that conjecture is slightly different to what is depicted in Figure 5. This other ("vertical") version was proven by Birch [7] already in 1968, while the Sato-Tate conjecture had to wait until 2008.

Now we will explore some consequences of Hasse's Theorem, specifically the validity of Definition 2.1.24 that states that an elliptic curve over $\mathbb{F}_p$ is supersingular if and only if it has exactly $p + 1$ points, or in other words, the trace of Frobenius is 0. To get there, we will need a helping lemma first.

Lemma 2.5.8. *[75, Theorem 4.12] Let $E(\mathbb{F}_p)$ be an elliptic curve defined over $\mathbb{F}_p$. Let $\#E(\mathbb{F}_p) = p+1-t$, and write $x^2 - tx + p = (x - \alpha)(x - \beta)$. Then for all $n \geq 1$, we have that*

$$\#E(\mathbb{F}_p) = p^n + 1 - (\alpha^n + \beta^n).$$

Proof. Let $s_n = \alpha^n + \beta^n$, and let us investigate what recurrence relation it satisfies (if any). Since α and β are roots of $x^2 - tx + p = 0$, multiplying the relations $\alpha^2 - t\alpha + p = 0$ and $\beta^2 - t\beta + p = 0$ by α^{n-1}

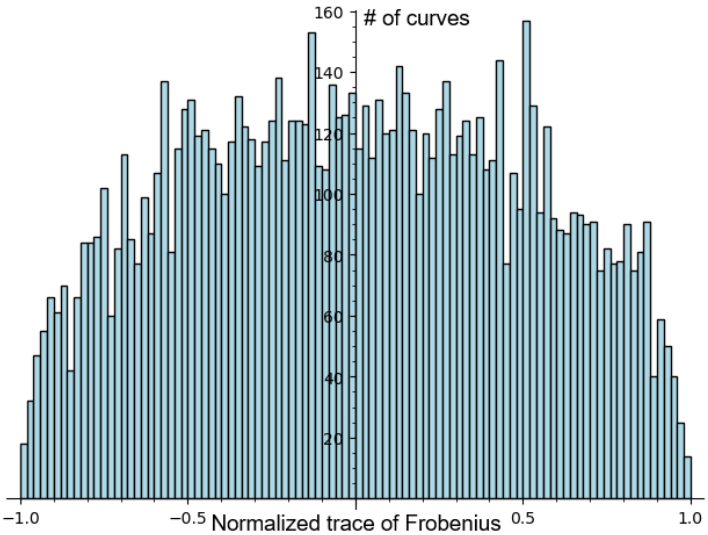


Figure 5: Histogram of traces of Frobenius for $10,000$ random elliptic curves over $\mathbb{F}_p$ with $p = 1,000,003$. The x-axis is normalized by dividing the trace of Frobenius by Hasse's bound, $\lfloor 2\sqrt{p} \rfloor$. Figure created with SageMath [39].

and β^{n-1} respectively yields

$$\begin{aligned} &\alpha^{n+1} = t\alpha^n - p\alpha^{n-1} \quad \text{and} \quad \beta^{n+1} = t\beta^n - p\beta^{n-1} \\ \implies &s_{n+1} = ts_n - ps_{n-1}, \quad s_0 = \alpha^0 + \beta^0 = 2 \quad \text{and} \quad s_1 = \alpha + \beta = t, \end{aligned} \tag{2.3}$$

where $\alpha + \beta = t$ follows immediately from the factorization of $x^2 - tx + p$. From this recurrence relation, it is clear that s_n is an integer for all $n \geq 0$. Now let

$$f(x) = (x^n - \alpha^n)(x^n - \beta^n) = x^{2n} - (\alpha^n + \beta^n)x^n + p^n.$$

Then we have that $x^2 - tx + p = (x - \alpha)(x - \beta)$ divides $f(x)$, and the quotient polynomial $Q(x)$ also has integer coefficients after the polynomial division (because the leading coefficient of $x^2 - tx + p$ is 1 and both polynomials have integer coefficients). Substituting in the Frobenius endomorphism π into $f(x)$ and utilizing the fact that $\pi^2 - t\pi + p = 0$ as endomorphisms on E by Proposition 2.5.6, we get

$$f(\pi) = (\pi^n)^2 - (\alpha^n + \beta^n)\pi^n + p^n = Q(\pi)(\pi^2 - t\pi + p) = 0.$$

Note that

$$\pi^n(x, y) = (\pi \circ \pi \circ \dots \pi)(x, y) = (x^{p^n}, y^{p^n}) = \pi_{p^n},$$

and by an extension of Corollary 2.5.4 to $\mathbb{F}_{p^n}$ and Proposition 2.5.6 again, we also have that

$$\pi_{p^n}^2 - \operatorname{tr}(\pi_{p^n})\,\pi_{p^n} + p^n = 0 \quad \text{and} \quad \operatorname{tr}(\pi_{p^n}) = p^n + 1 - \#E(\mathbb{F}_{p^n}),$$

thus combining this and $f(\pi)$ gives that

$$\begin{aligned} \alpha^n + \beta^n = \operatorname{tr}(\pi_{p^n}) &= p^n + 1 - \#E(\mathbb{F}_{p^n}) \\ \implies \#E(\mathbb{F}_{p^n}) &= p^n + 1 - (\alpha^n + \beta^n). \end{aligned}$$

□

The advantage of Lemma 2.5.8 is that if there is an elliptic curve which is defined over a small finite field, we are able to calculate the number of points very quickly, even quicker than Schoof's or the SEA algorithm mentioned in 2.1. This advantage is unfortunately irrelevant for the topic of this thesis, since CSIDH uses a 512-bit prime number (at least). Alas, it still helps in proving the following theorem.

Theorem 2.5.9. *[75, Proposition 4.31] Let $p \geq 5$ be a prime, $E(\mathbb{F}_p)$ be an elliptic curve over the finite*

field $\mathbb{F}_p$ and t be the trace of Frobenius. Then

$$E \text{ is supersingular } \iff t = 0 \iff \#E(\mathbb{F}_p) = p + 1.$$

Proof. The second if and only if statement holds trivially by Corollary 2.5.4, so we only need to focus on the first statement. To this end, we start by first factoring the characteristic polynomial of the Frobenius endomorphism as $x^2 - tx + p = (x - \alpha)(x - \beta)$. Then by Lemma 2.5.8, we have that

$$\#E(\mathbb{F}_{p^n}) = p^n + 1 - (\alpha^n + \beta^n),$$

and by Equation 2.3 that $s_n = \alpha^n + \beta^n$ satisfies the recurrence relation

$$s_0 = 2, \quad s_1 = t, \quad s_{n+1} = ts_n - ps_{n-1}.$$

We begin by a weaker assumption, so assume $t \equiv 0 \pmod{p}$. Then we have that $s_1 = t \equiv 0 \pmod{p}$, and $s_{n+1} \equiv 0 \pmod{p}$ for all $n \geq 1$ by the recurrence relation. This in turn means that

$$\#E(\mathbb{F}_p) = p^n + 1 - s_n \equiv 1 \pmod{p}$$

, which means that there are no points of order p in $E(\mathbb{F}_{p^n})$ for any $n \geq 1$. By definition of the algebraic closure of finite fields $\overline{\mathbb{F}}_p = \cup_{n \geq 1} \mathbb{F}_{p^n}$, so it follows that there are no points of order p in $E(\overline{\mathbb{F}}_p)$ either. This is exactly what supersingularity is by Definition 2.2.24.

Assume now that $t \not\equiv 0 \pmod{p}$, then the recurrence implies that $s_{n+1} \equiv ts_n \pmod{p}$ for all $n \geq 1$, and since $s_1 = t$, we have $s_n \equiv t^n \pmod{p}$ for all $n \geq 1$. Therefore

$$\#E(\mathbb{F}_{p^n}) = p^n + 1 - s_n \equiv 1 - t^n \pmod{p}.$$

By Fermat's little theorem, $a^{p-1} \equiv 1 \pmod{p}$, therefore $E(\mathbb{F}_{p^{p-1}})$ has an order divisible by p, so it contains a point of order p. This then means that E is not supersingular.

We have shown so far that E is supersingular if and only if $t \equiv 0 \pmod{p}$. To get to the specific statement of the theorem for $t = 0$, suppose E is supersingular but $t \neq 0$. Then $t \equiv 0 \pmod{p}$, so $|t| \geq p$. By Hasse's theorem (Theorem 2.5.7), we have that $|t| \leq 2\sqrt{p}$, so $p \leq 2\sqrt{p}$, which is only possible for $p \leq 4$. □

Remark 2.5.10. Corollary 2.5.4 and Theorem 2.5.7 work in general for $q = p^n$ as well, and Theorem 2.5.9's statement can be expanded to E being supersingular if and only if $t \equiv 0 \pmod{p}$ and $\#E(\mathbb{F}_q) = p + 1 - t \pmod{p}$ for $\mathbb{F}_q$. These are the versions which are usually proven, but in this thesis, I wanted to focus on the specific results that are relevant for CSIDH, which only uses the prime field $\mathbb{F}_p$.

We have now proven that Definitions 2.2.24 and 2.1.24 are equivalent. This is going to be used in the next lemma to derive another condition for supersingularity using polynomial expansion. This makes it quite easy to prove for which primes p the elliptic curve $E_0 : y^2 = x^3 + x$ is supersingular in $\mathbb{F}_p$. This is relevant for CSIDH, since E_0 is the starting curve in its protocol, and we require this curve to be supersingular for reasons described in Section 2.6.

Lemma 2.5.11. *[69, Theorem V.4.1(a)] Let $E(\mathbb{F}_p)$ be an elliptic curve given by a Weierstrass equation*

$$E : y^2 = f(x),$$

where $f(x) \in \mathbb{F}_p[x]$ is a cubic polynomial with distinct roots in $\overline{\mathbb{F}}_p$ with $p \geq 5$. Then E is supersingular if and only if the coefficient of x^{p-1} in $f(x)^{(p-1)/2}$ is zero.

Proof. Recall Definition 2.4.33 of the Legendre symbol, and count the number of solutions to $y^2 = f(x) = x^3 + Ax + B$. If $x_0^3 + Ax_0 + B$ is a solution to $f(x)$ that is a square in $\mathbb{F}_p$, then there are two solutions (x_0, y_0) and $(x_0, -y_0)$. If $f(x_0)$ is not a square, there are no solutions, and if it's 0, then there is one point on the curve $(0, 0)$. Note that these options are exactly the Legendre symbol of $f(x_0)$ with respect

to p, so if we sum over all $x \in \mathbb{F}_p$ and add the point at infinity, we get

$$\begin{aligned}\#E(\mathbb{F}_p) &= 1 + \sum_{x \in \mathbb{F}_p} \left(1 + \left(\frac{f(x)}{p}\right)\right) \\ &= 1 + p + \sum_{x \in \mathbb{F}_p} \left(\frac{f(x)}{p}\right).\end{aligned}$$

Since $\mathbb{F}_p$ is cyclic of order $p-1$, by Euler's criterion (which is also the basis for the Legendre symbol), we have that $(\frac{a}{p}) = a^{(p-1)/2} \pmod{p}$, therefore we get that

$$\#E(\mathbb{F}_p) \equiv 1 + \sum_{x \in \mathbb{F}_p} f(x)^{(p-1)/2} \pmod{p}.$$

Note that the initial p has disappeared since we are calculating modulo p now. Next, we have the result from elementary number theory that

$$\sum_{x \in \mathbb{F}_p} x^i = \begin{cases} -1 & \text{if } p-1 \mid i, \\ 0 & \text{if } p-1 \nmid i. \end{cases}$$

Since $f(x)$ has degree 3, raising it to the $(p-1)/2$ power means that the possible exponents appearing in its expansion is x^n for $0 \leq n \leq \frac{3}{2}(p-1)$. The only nonzero summand appearing is thus the coefficient of x^{p-1}. Let A_p be this coefficient, so then we can simply write

$$\#E(\mathbb{F}_p) \equiv 1 - A_p \pmod{p}.$$

So far we have proven the formula for modulo p, and to show exactly that $A_p = 0$, we can use Corollary 2.5.4, which states that

$$\#E(\mathbb{F}_p) = p + 1 - t \equiv 1 - t \pmod{p},$$

where t is the trace of Frobenius. Equating this with the previous formula, we clearly have that $t = A_p \pmod{p}$, but as a result of Hasse's theorem and the assumption on $p \geq 5$, we have that $t \equiv 0 \pmod{p} \implies t = 0$. Thus we have the chain of implications

$$A_p = 0 \iff t \equiv 0 \pmod{p} \overset{\text{Hasse}}{\iff} t = 0 \iff E \text{ is supersingular},$$

where the last implication holds by Theorem 2.5.9. □

Remark 2.5.12. As shown in the proof of this theorem, we have a new way to express the trace of Frobenius t:

$$t = -\sum_{x \in \mathbb{F}_p} \left(\frac{f(x)}{p}\right).$$

By the definition of the quadratic twist, we multiply y^2 by a non-square u, so the relationship between the number of specific points per evaluation flips. This results in the fact that if E^t is the twisted curve, we have that

$$\#E^t(\mathbb{F}_p) = 1 + \sum_{x \in \mathbb{F}_p} \left(1 - \left(\frac{f(x)}{p}\right)\right) = p + 1 + t.$$

In particular, this shows that when E is supersingular, so is its twist E^t.

Theorem 2.5.13. *Let $E_0(\mathbb{F}_p) : y^2 = x^3 + x$ be an elliptic curve defined over the finite field $\mathbb{F}_p$ with $p \geq 3$. Then*

$$E_0 \text{ is supersingular} \iff p \equiv 3 \pmod{4}.$$

Proof. Let $f(x) = x^3 + x$, and as seen in Lemma 2.5.11, we can check the coefficient of x^{p-1} in the expansion of $f(x)^{(p-1)/2}$. This coefficient is equivalent to the coefficient of $x^{(p-1)/2}$ in the polynomial

$(x^2+1)^{(p-1)/2}$. Using the binomial expansion formula, we have

$$(x^2+1)^{(p-1)/2} = \sum_{k=0}^{(p-1)/2} \binom{(p-1)/2}{k} x^{2k}.$$

The coefficient of $x^{(p-1)/2}$ is then given by $k = (p-1)/4$, which is an integer only if $p \equiv 1 \pmod 4$. Therefore we have that

$$A_p = \begin{cases} \binom{(p-1)/2}{(p-1)/4} & \text{if } p \equiv 1 \pmod 4, \\ 0 & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

By Lemma 2.5.11 again, we can conclude that if $p \equiv 3 \pmod 4$, then $A_p = 0$ and thus E_0 is supersingular. □

We have now seen a bit more about the connections between different definitions of supersingularity, which constitutes one basis for the CSIDH protocol. Another basis is the structure of the set of endomorphisms for the elliptic curves. If we allow the multiplication-by-0 map $[0]$ to be an endomorphism as well, defined as sending all points to the point at infinity, then $\mathrm{End}(E)$ is not merely a set, but it also has the structure of a ring.

Proposition 2.5.14. *Let $E(k)$ be an elliptic curve defined over the field k. Then the set of endomorphisms $\mathrm{End}(E)$ forms a ring.*

Proof. Let $\varphi, \psi : E \to E$ be endomorphisms. A ring needs an "addition" and "multiplication". We will define addition of endomorphisms by piece-wise addition, and multiplication by function composition:

$$(\varphi+\psi)(x) := \varphi(x) + \psi(x) \quad \text{and} \quad (\varphi \cdot \psi)(x) := (\varphi \circ \psi)(x) = \varphi(\psi(x)).$$

Note that in the definition of addition, the second plus-sign refers to the addition of points on the elliptic curve E. Since E forms an abelian group with respect to addition (Theorem 2.1.11), we have the abelian group basis for our ring with the addition of the "zero map", i.e. the multiplication-by-0 map. It is also easy to see that multiplication is associative, and the identity map $[1] : P \mapsto P$ acts as the multiplicative identity. We just need to show that multiplication is distributive with respect to addition.

Let $\alpha : E \to E$ be a third endomorphism. Then by the fact that all endomorphisms are in particular group homomorphisms as well, we have that

$$\begin{aligned} ((\varphi+\psi)\cdot\alpha)(x) &= (\varphi+\psi)(\alpha(x)) = \varphi(\alpha(x)) + \psi(\alpha(x)) = (\psi\cdot\alpha + \psi\cdot\alpha)(x), \\ (\alpha\cdot(\varphi+\psi))(x) &= \alpha((\varphi+\psi)(x)) = \alpha(\varphi(x)) + \alpha(\psi(x)) = (\alpha\cdot\varphi + \alpha\cdot\psi)(x), \end{aligned}$$

showing that indeed multiplication is distributive, thus $\mathrm{End}(E)$ indeed forms a ring. □

Next, we record some properties of this newly defined endomorphism ring that are going to be needed for investigating what kind of rings $\mathrm{End}(E)$ could be isomorphic to.

Proposition 2.5.15. *The following are true for the ring $\mathrm{End}(E)$:*

(i) It has characteristic zero.

(ii) It has no zero divisors.

(iii) For $\alpha \in \mathrm{End}(E)$, the product $\alpha\hat{\alpha} = \mathrm{N}(\alpha)$ is a nonnegative integer, and $\mathrm{N}(\alpha) = 0$ if and only if $\alpha = [0]$.

(iv) It has rank at most 4 as a $\mathbb{Z}$-module.

Proof. (i) Suppose we have $\varphi \in \mathrm{End}(E)$ and $m \neq 0, m \in \mathbb{Z}$ such that they satisfy

$$m\varphi = [m] \circ \varphi = [0].$$

By Theorem 2.2.21(iii), taking degrees on both sides gives

$$(\deg[m])(\deg\varphi) = m^2 \deg\varphi = 0.$$

Since $m \neq 0$, we must have that $\deg\varphi = 0$, and therefore $\varphi = [0]$.

(ii) Let $\alpha, \beta \in \mathrm{End}(E)$ be two nonzero endomorphisms. Since both α and β are surjective, then clearly $\alpha\beta = \alpha \circ \beta$ is also surjective, and in particular a nonzero endomorphism. This shows that $\alpha\beta \neq [0]$, i.e. $\mathrm{End}(E)$ indeed has no zero divisors.

(iii) The first statement is just Definition 2.2.20. Assume $\mathrm{N}(\alpha) = 0$, then by Proposition 2.5.3, we have that $\deg\alpha = 0$, which implies $\alpha = [0]$. The converse is trivial, since any function post-composed with [0] will result in [0].

(iv) This is not too hard to show, but it uses Tate Modules, which are not introduced in this paper. The proof can be found in [69, Corollary III.7.5]. □

Now that we have gathered a decent understanding of the endomorphism ring $\mathrm{End}(E)$, we can begin proving more interesting theorems that are going to be relevant for our next chapter, the cryptographic algorithm of CSIDH. First, we begin by a simple lemma that justifies our decision to sometimes refer to the multiplication-by-m maps $[m]$ without the brackets.

Lemma 2.5.16. *Let $E(k)$ be an elliptic curve defined over the field k. Then the map $[\cdot] : \mathbb{Z} \hookrightarrow \mathrm{End}_k(E)$ is an injective ring homomorphism.*

Proof. It is clearly a ring homomorphism, since the additive and multiplicative identities are conserved $(0 \to [0], 1 \to [1])$, and by the definition of the multiplication-by-m map, we have that $[m+n] = [m]+[n]$ and $[mn] = [m][n]$. It remains to show injectivity.

Let $[m] = [m']$ be two multiplication maps. Then we have that $[m]-[m'] = [0] \implies [m-m'] = [0]$. This means that $[m-m']$ maps all points $P \in E(\overline{k})$ to the point at infinity $\infty \in E(\overline{k})$, which means that $[m-m']$ must be a constant morphism. Since the degree of a constant morphism is 0 and $\deg[m-m'] = (m-m')^2$, the only possibility is that $m - m' = 0 \implies m = m'$, showing injectivity. □

For most elliptic curves defined over k with $\mathrm{char}(k) = 0$, this inclusion defines the full endomorphism ring, i.e. $\mathrm{End}(E) \cong \mathbb{Z}$. In the zero-characteristic setting, it may happen that $E(k)$ has extra endomorphisms, in which case we say that E has *complex multiplication.* These curves are indeed interesting with special properties, but they are not used in CSIDH, so we refer the intrigued reader to a discussion in [69, C §11].

If k is a finite field, i.e. $\mathrm{char}(k) = p > 0$, then $\mathrm{End}(E)$ is always larger than $\mathbb{Z}$. We prove this fact in full generality, then give specifications pertaining to supersingular elliptic curves defined over a finite field, since this is the main setting of CSIDH and the subsequent improvements discussed in Chapter 5. For the general case, we start by giving a definition.

Definition 2.5.17. A **Quaternion algebra** is an algebra of the form

$$B = \mathbb{Q} + \mathbb{Q}\alpha + \mathbb{Q}\beta + \mathbb{Q}\alpha\beta,$$

where the multiplication of the non-integers is defined by

$$\alpha^2, \beta^2 \in \mathbb{Q}, \quad \alpha^2 < 0, \quad \beta^2 < 0, \quad \text{and} \quad \alpha\beta = -\beta\alpha.$$

Theorem 2.5.18. *Let $E(k)$ be an elliptic curve defined over the field k. The endomorphism ring* $\mathrm{End}(E)$ *is either*

(i) isomorphic to $\mathbb{Z}$,

(ii) an order in an imaginary quadratic field,

(iii) an order in a quaternion algebra.

Proof. [69, Theorem III.9.3.] Let $K = \mathrm{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ be $\mathrm{End}(E)$ with an extended scalar field $\mathbb{Q}$, so fractional multiplication maps are also defined. Since $\mathrm{End}(E)$ is a finitely generated $\mathbb{Z}$-module by Proposition 2.5.15(iv), it suffices to show that K is either $\mathbb{Q}$, an imaginary quadratic field or a quaternion algebra. Since we've extended the scalars to the rational numbers, we also have that the trace and norm of any element $\alpha \in K$ is also a rational number, i.e. $\mathrm{tr}(\alpha), \mathrm{N}(\alpha) \in \mathbb{Q}$. Note also that if $\alpha \in \mathbb{Q}$, then $\hat{\alpha} = \alpha \implies \mathrm{tr}(\alpha) = 2\alpha$. Furthermore, if $\alpha \in K$ satisfies $\mathrm{tr}(\alpha) = 0$, then we have that

$$0 = (\alpha - \alpha)(\alpha - \hat{\alpha}) = \alpha^2 - \mathrm{tr}(\alpha)\,\alpha + \mathrm{N}(\alpha) = \alpha^2 + \mathrm{N}(\alpha) \implies \alpha^2 = -\mathrm{N}(\alpha),$$

which in turn yields the implication

$$\alpha \neq 0 \quad \text{and} \quad \mathrm{tr}(\alpha) = 0 \implies \alpha^2 \in \mathbb{Q} \quad \text{and} \quad \alpha^2 < 0. \tag{2.4}$$

If $K = \mathbb{Q}$, there is nothing more to prove, since $\mathbb{Q} = \mathrm{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ can only be true when $\mathrm{End}(E) \cong \mathbb{Z}$ (recall that $\mathrm{End}(E)$ is a $\mathbb{Z}$-module of rank at most 4). Otherwise, we can find some $\alpha \in K$ with $\alpha \notin \mathbb{Q}$. Next, we can set $\alpha_0 = \alpha - \frac{1}{2}\mathrm{tr}(\alpha)$, then we have that

$$\begin{aligned}
\mathrm{tr}(\alpha_0) &= \mathrm{tr}\left(\alpha - \frac{1}{2}\mathrm{tr}(\alpha)\right) = \mathrm{tr}(\alpha) - \mathrm{tr}\left(\frac{1}{2}\mathrm{tr}(\alpha)\right) \\
&= \mathrm{tr}(\alpha) - 2 \cdot \frac{1}{2}\mathrm{tr}(\alpha) \\
&= 0.
\end{aligned}$$

By Equation 2.4, $\alpha_0^2 < 0 \in \mathbb{Q}$, so $\mathbb{Q}(\alpha_0)$ is a quadratic imaginary field. If $K = \mathbb{Q}(\alpha)$, we are again done.

Let $\alpha = \alpha_0$ and suppose now that $K \neq \mathbb{Q}(\alpha)$. Then we can choose some $\beta \in K$ such that $\beta \notin \mathbb{Q}(\alpha)$. Similarly as before, we can set

$$\beta_0 = \beta - \frac{1}{2}\mathrm{tr}(\beta) - \frac{\mathrm{tr}(\alpha\beta)}{2\alpha^2}\alpha.$$

Taking traces once more yields

$$\begin{aligned}
\mathrm{tr}(\beta_0) &= \mathrm{tr}(\beta) - \mathrm{tr}\left(\frac{1}{2}\mathrm{tr}(\beta)\right) - \mathrm{tr}\left(\frac{\mathrm{tr}(\alpha\beta)}{2\alpha^2}\alpha\right) \\
&= \mathrm{tr}(\beta) - \mathrm{tr}(\beta) - \frac{\mathrm{tr}(\alpha\beta)}{2\alpha^2}\mathrm{tr}(\alpha) \\
&= 0,
\end{aligned}$$

where the second equation holds because $\alpha^2, \mathrm{tr}(\alpha\beta) \in \mathbb{Q}$, and the third equation holds since $\mathrm{tr}(\alpha) = 0$ from before. By a similar calculation, we also see that

$$\begin{aligned}
\mathrm{tr}(\alpha\beta_0) &= \mathrm{tr}(\alpha\beta) - \mathrm{tr}\left(\frac{1}{2}\mathrm{tr}(\beta)\,\alpha\right) - \mathrm{tr}\left(\frac{\mathrm{tr}(\alpha\beta)}{2\alpha^2}\alpha^2\right) \\
&= \mathrm{tr}(\alpha\beta) - 0 - \mathrm{tr}(\alpha\beta) \\
&= 0.
\end{aligned}$$

We have shown that $\mathrm{tr}(\beta_0) = \mathrm{tr}(\alpha\beta_0) = 0$, and in particular from Equation 2.4, we have $\beta^2 < 0 \in \mathbb{Q}$. We have now shown three different traces being 0, and if we once again write just β instead of β_0, we get that

$$\begin{aligned}
\mathrm{tr}(\alpha) = 0 &\implies \alpha = -\hat{\alpha}, \\
\mathrm{tr}(\beta) = 0 &\implies \beta = -\hat{\beta}, \\
\mathrm{tr}(\alpha\beta) = 0 &\implies \alpha\beta = -\hat{\beta}\hat{\alpha}.
\end{aligned}$$

Substituting the first two rows into the third one gives the relation $\alpha\beta = -\beta\alpha$, so we have that

$$\mathbb{Q}[\alpha, \beta] = \mathbb{Q} + \mathbb{Q}\alpha + \mathbb{Q}\beta + \mathbb{Q}\alpha\beta$$

is a quaternion algebra by Definition 2.5.17.

It remains to prove that $\mathbb{Q}[\alpha, \beta] = K$. This can be done by showing that the there is a $\mathbb{Q}$-linearly independent basis $\{1, \alpha, \beta, \alpha\beta\}$, because then both $\mathbb{Q}[\alpha, \beta]$ and K have dimension 4 over $\mathbb{Q}$, and thus are isomorphic. To this end, suppose that

$$w + x\alpha + y\beta + z\alpha\beta = 0, \quad w, x, y, z \in \mathbb{Q}.$$

Taking traces both sides yields $2w = 0$, so $w = 0$. Next, we multiply by α and β on the left and right respectively, and get

$$\alpha x\alpha\beta + \alpha y\beta^2 + \alpha z\alpha\beta^2 = (x\alpha^2)\beta + (y\beta^2)\alpha + x\alpha^2\beta^2 = 0,$$

where the first equation holds since rational numbers commute with endomorphisms. We have by construction from before that 1, α and β are $\mathbb{Q}$-linearly independent, hence the above equation implies that

$$x\alpha^2 = y\beta^2 = z\alpha^2\beta^2 = 0 \implies x = y = z = 0.$$

This is exactly what we wanted to show, and this concludes the proof. □

Definition 2.5.19. The scalar extension of $\mathrm{End}(E)$ in the beginning of the previous proof, denoted by $\mathrm{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q} = \mathrm{End}^0(E)$ is called the **endomorphism algebra** of E. It has the standard tensor addition, but also since it's an algebra, it also has a component-wise multiplication defined by

$$(\alpha \otimes r)(\beta \otimes s) = (\alpha\beta) \otimes (rs),$$

where $\alpha, \beta \in \mathrm{End}(E)$, and $r, s \in \mathbb{Q}$.

An important theorem of Tate [73] (or rather a corollary of it) relates the endomorphism algebra of two elliptic curves in the following way.

Theorem 2.5.20. *[29] Let E, E' be two elliptic curves defined over a finite field k. Then E and E' are k-isogenous if and only if their endomorphism algebras are isomorphic, i.e.* $\mathrm{End}^0(E) \cong \mathrm{End}^0(E')$.

Proof. See [73] for his original proof with Tate modules and other tools. □

For our case of supersingular curves due to CSIDH, an important corollary of this theorem is that the number of points of two isogenous curves are the same. This will be important in the next section, and is proved in an alternative way in Theorem 2.6.15. Let us now focus on the possible endomorphism rings of supersingular curves.

Proposition 2.5.21. *[31, Proposition 2.4.] Let $p > 3$ and let $E(\overline{\mathbb{F}}_p)$ be a supersingular elliptic curve over $\overline{\mathbb{F}}_p$. Then*

$$E \text{ is defined over } \mathbb{F}_p \iff \mathbb{Z}[\sqrt{-p}] \subseteq \mathrm{End}_p(E).$$

Proof. ($\implies$) Assume we have the elliptic curve $E(\mathbb{F}_p)$. Then π is in $\mathrm{End}(E)$, and it satisfies the polynomial $\pi^2 + p = 0$ by Proposition 2.5.6. Thus we have that $\pi = \sqrt{-p} \notin \mathbb{Z}$, so $\mathbb{Z}[\pi] \cong \mathbb{Z}[\sqrt{-p}] \subseteq \mathrm{End}(E)$.

($\impliedby$) Conversely, assume that $\varphi \in \mathrm{End}(E)$ satisfies $\varphi^2 = [-p]$, then φ is an isogeny of degree p and $\hat{\varphi} \circ \varphi = [p]$. Since E is supersingular, it follows that φ has kernel $\{\infty\}$, so it is inseparable. By Theorem 2.2.14, we have that

$$\varphi = \varphi_{\mathrm{sep}} \circ \pi \implies \deg_s \varphi = 1,$$

so φ_{sep} is an isomorphism, hence the j-invariants are equal for $j(E) = j(E^{(p)}) = j(E)^p$. This in turn means that $j(E) \in \mathbb{F}_p$, so E is defined over $\mathbb{F}_p$ by Proposition 2.1.17, as required. □

Theorem 2.5.22. *[70, Theorem 13.6.] Let $E(\mathbb{F}_p)$ be a supersingular elliptic curve over $\mathbb{F}_p$ and π its Frobenius endomorphism. Then the endomorphism algebra is isomorphic to*

$$\mathrm{End}_p^0(E) = \mathbb{Q}(\pi) \cong \mathbb{Q}(\sqrt{-p}).$$

Proof. The Frobenius endomorphism π is a root of its characteristic polynomial $x^2 - \mathrm{tr}\,(\pi)\,x + \deg(\pi)$ by Proposition 2.5.6. Since E is supersingular, we have that $\pi^2 + p = 0 \implies \pi = \sqrt{-p}$, so $\mathbb{Q}(\pi) \cong \mathbb{Q}(\sqrt{-p})$.

By Theorem 2.5.18, the endomorphism algebra could be isomorphic to a quaternion algebra as well, so we need to check what happens to an arbitrary element $\alpha \in \mathrm{End}_p^0(E)$. As elements, we can write α and π as

$$\alpha = \varphi \otimes s, \quad \text{and} \quad \pi_E = \pi \otimes 1,$$

for some $\varphi \in \mathrm{End}_p(E)$ and $s \in \mathbb{Q}$. We can furthermore write $\varphi = (r_1(x), r_2(x)y)$ in standard form, then by the definition of multiplication in Definition 2.5.19, we have

$$\alpha\pi_E = (\varphi \otimes s)(\pi \otimes 1) = (\varphi\pi) \otimes s.$$

Focusing on the first tensor element, we have that

$$(\varphi\pi)(x, y) = (r_1(x^p), r_2(x^p)y^p) = (r_1(x)^p, r_2(x)^p y^p) = (\pi\varphi)(x, y),$$

therefore

$$\alpha\pi_E = (\varphi\pi) \otimes s = (\pi\varphi) \otimes s = \pi_E\alpha.$$

Since $\mathbb{Q}$ is in the center of $\mathrm{End}_p^0(E)$, we have that α commutes with both $\mathbb{Q}$ and π_E. Just like in the proof of Theorem 2.5.18(iii), we can do the same translations to α and π_E to get that $\alpha\pi_E = -\pi_E\alpha$, but since they commute as shown above, we have that

$$2\alpha\pi_E = (2\varphi\pi) \otimes s = 0 \implies \varphi\pi = 0 \implies \varphi = 0 \in \mathbb{Q}(\pi) \cong \mathbb{Q}(\sqrt{-p}),$$

where the first implication follows by the fact that a pure tensor is 0 if one of its parts is 0 ($s \neq 0$), and the second implication follows by the fact that $\mathrm{End}_p(E)$ has no zero divisors and $\pi = \sqrt{-p} \notin \mathbb{Q}$. We have thus shown that any arbitrary element of the endomorphism algebra is an element of $\mathbb{Q}(\sqrt{-p})$, as required. □

Theorem 2.5.23. *Let $E(\mathbb{F}_p)$ be a supersingular elliptic curve defined over $\mathbb{F}_p$. Then the endomorphism ring $\mathrm{End}_p(E)$ falls into two cases:*

- *If $p \equiv 1 \pmod 4$, then $\mathrm{End}_p(E) = \mathbb{Z}[\sqrt{-p}]$.*
- *If $p \equiv 3 \pmod 4$, then $\mathrm{End}_p(E)$ is either $\mathbb{Z}[\sqrt{-p}]$ or $\mathbb{Z}\left[\frac{1+\sqrt{-p}}{2}\right]$.*

Proof. From Theorem 2.5.18 and 2.5.22 we know that $\mathrm{End}_p(E)$ is an order $\mathcal{O}$ in $K = \mathbb{Q}(\sqrt{-p})$. From Proposition 2.5.21, we also have that since E is defined over $\mathbb{F}_p$, the possibilities for the endomorphism ring is sandwiched between

$$\mathbb{Z}[\sqrt{-p}] \subseteq \mathrm{End}_p(E) \subseteq \mathcal{O}_K,$$

where $\mathcal{O}_K$ is the ring of integers of K. By Theorem 2.4.16, if $p \equiv 1 \pmod 4$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{-p}]$, and when $p \equiv 3 \pmod 4$, then $\mathcal{O}_K = \mathbb{Z}\left[\frac{1+\sqrt{-p}}{2}\right]$. In the latter case, $\mathbb{Z}[\sqrt{-p}]$ has conductor $f = 2$, so these are the only two options. □

2.6 Isogeny Graphs

In this section, we explore the possibilities of having cryptographic protocols defined on graph, and why it is justified for CSIDH to be using isogenies between elliptic curves for its key exchange. The contents of this section mainly follows the structure of [29, Part III.] and [31]. This section will also require some knowledge of graph theory, so we will start by giving a refresher on the most relevant definitions. We start by the easy definitions.

Throughout this section, we denote any graph by $G = (V, E)$, its vertex set by V and edge set by E. Vertices are denoted by v and edges are denoted by $e = (v_1, v_2)$, where v_1 is the starting vertex, and v_2 is the ending vertex. A **path** is a series of edges that that share the same ending vertex with the starting vertex of the next edge. A graph is called **connected** if for any two vertices, there exists at least one path between them, and otherwise disconnected. If the graph is disconnected, then it decomposes into a disjoint union of **connected components**, which are the connected subgraphs G_i of the main graph G. A graph is call **directed** if the edges $(v_1, v_2) \neq (v_2, v_1)$, otherwise we call it **undirected** and the start and end vertices of edges are not distinguished.

The **degree of a vertex** $\deg(v)$ is the number of edges going in and out of the vertex v. A graph is called **k-regular** if all vertices of the graph have exactly $\deg(v) = k$. The **distance** $d(v_1, v_2)$ between two vertices is the length of the shortest path between them, if there is no path, then we set the distance to infinity. The **diameter** of a graph is the largest of all distance between any of its vertices.

Now we separately record some definitions that are more advanced, and are the most important.

Definition 2.6.1. Let $G = (V, E)$ be a graph with n vertices. The **adjacency matrix** of G is the $n \times n$ matrix, where the (i, j)'th column is 1 if there is an edge between v_i and v_j. If G is undirected, then this matrix is symmetric, and it has n real eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Definition 2.6.2. Let G be a k-regular graph and let $\varepsilon > 0$. G is called a (one-sided) **ε-expander** if

$$\lambda_2 \leq (1 - \varepsilon)k,$$

and a **two-sided ε-expander** if it also satisfies

$$\lambda_n \geq -(1 - \varepsilon)k.$$

Definition 2.6.3. Let G be a k-regular graph and λ_i its eigenvalues. G called a **Ramanujan graph** if for all λ_i except for λ_1 it holds that

$$|\lambda_i| \leq 2\sqrt{k-1}.$$

Expander graphs are graphs where small sets of vertices have many neighbors, just as the k-regular definition also suggests. This also means that the diameter of these graphs is small, therefore there exist short paths between any two vertices. Ramanujan graphs are in some sense the limit of expander graphs, because they maximize the difference between the biggest and second-biggest eigenvalue of its adjacency matrix, which results in even smaller sets of vertices having even more neighbors. In the language of functional analysis, Ramanujan graphs have a near maximal spectral gap, which makes them a great type of graph for cryptographic purposes. Let us first see an example to familiarize ourselves with the above definitions, and then explore the reasons for why they are great for cryptography.

Example 2.6.4. Let G be the undirected graph as in Figure 6. It has 10 vertices and it's 4-regular since all vertices have degree 4. The adjacency matrix is

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

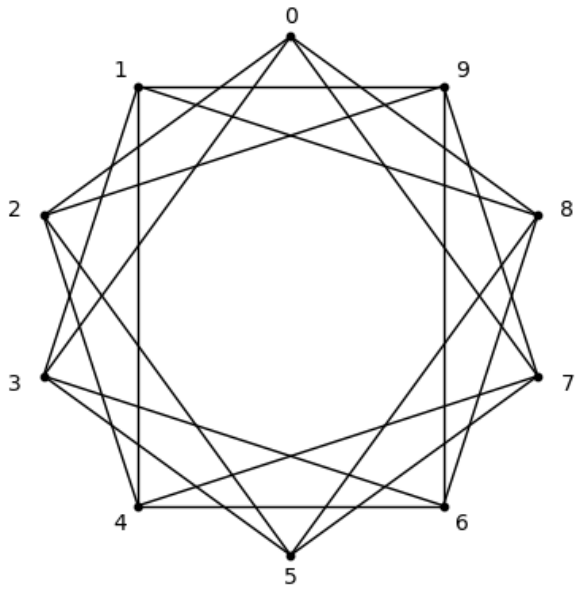


Figure 6: A 4-regular graph on 10 vertices. Figure created with SageMath [39].

with sorted list of eigenvalues

$$\lambda_1 = 4, 1.236, 1.236, 0, 0, 0, 0, 0, -3.236, -3.236 = \lambda_n.$$

From this, we can see that G is an ε_1-expander for $\varepsilon_1 = 0.69$ and a two-sided ε_2-expander for $\varepsilon_2 = 0.191$. We also have that $2\sqrt{k-1} = 2\sqrt{3} \sim 3.4641$, and $\max\{|\lambda_2|, |\lambda_{10}|\} = 3.236 \leq 3.4641$, we have that this graph is a Ramanujan graph.

Ramanujan graphs, as mentioned before, have a very attractive property of having relatively small (and thus more similar) eigenvalues in its adjacency matrix. To illustrate this [41], let A be the adjacency matrix, v_i be the vector where the i'th entry is 1 and the others are 0, then for any vertex j and integer d, the number of d-long paths from i to j is given by the j'th entry of $A^d v_i$, or simply $(i, j) \in A^d$. We can also write v_i in the eigenvector-basis of A

$$v_i = e_1 + e_2 + \cdots + e_n,$$

with their respective eigenvalues $\lambda_1, \ldots, \lambda_n$. Then we have that

$$A^d v_i = \lambda_1^d e_1 + \lambda_2^d e_2 + \cdots + \lambda_n^d e_n.$$

Clearly, if all the other eigenvalues are smaller than λ_1, then λ_1 will dominate. Then the smaller the other eigenvalues are, the more uniform are the number of d-long paths between any two vertices. This in turn means that if one starts taking a random walk from vertex i, then it can quickly reach any other vertex in roughly the same amount of steps. Cryptographically, this means that after a necessary amount of steps, there is an equal chance of ending up on any of the vertices of a graph. If we connect back to the discussion in the introductory section 1.2, we can immediately realize that this property is essentially like the"trap-door" function: easy to take random steps, hard to go back. This suggests that perhaps Ramanujan graphs (or in general expander graphs) could be a good foundation for cryptographic purposes, which is exactly what Charles, Goren and Lauter proposed in [19] specifically for hash functions, and what the paper [24] explores in general. Beyond these, expander families of graphs other applications as well, such as pseudo-random number generators or error-correcting codes.

Definition 2.6.5. A **random walk** of length i is a path $v_1 \to \cdots \to v_i$, where the next vertex is defined by a random process selecting v_i uniformly from the neighbors of v_{i-1}.

The next proposition is about how random walks behave in two-sided ε-expander graphs.

Proposition 2.6.6 (Mixing theorem)**.** *[29, Proposition 46.] Let $G = (V, E)$ be a k-regular two-sided ε-expander. Let $S \subset V$ be any subset of the vertices of G, and let v be any vertex in V. Then a random walk of length at least*

$$\frac{\ln(\sqrt{|S|}/2|V|)}{\ln(1-\varepsilon)}$$

starting from v will land in S with probability at least $\frac{1}{2}\frac{|S|}{|V|}$.

Proof. The proof uses representation theory, which was not introduced in this thesis, so for a full proof, see [44, Lemma 2.1.]. The proof then follows by setting $\Gamma = G$ and $c = (1-\varepsilon)k$ in the statement of the proposition. □

Definition 2.6.7. The expression in Proposition 2.6.6 is called the **mixing length** of the expander graph.

Another relevant consequence of the previous proposition is a theorem that relates to the class group of $\mathrm{End}(E)$, which is an order in an imaginary quadratic field $K = \mathbb{Q}(\sqrt{\Delta})$ when E is ordinary with complex multiplication, or when it is $\mathrm{End}_p(E)$. The proof of this theorem is using unintroduced tools, so we refer the interested reader again to [44]. Before this theorem, we will give some necessary definitions first.

Definition 2.6.8. Let A be a subset of X. The **symmetric subset** S of X is the multiset $A \cup A^{-1}$, where A^{-1} denotes the inverses of the elements of A.

Definition 2.6.9. Let G be a group and $S \subset G$ a symmetric subset. The **Cayley Graph** $\mathrm{Cay}(G, S)$ is a graph where the elements of G are the vertices, and there is an edge between g and h if $h = sg$ for some $s \in S$.

Example 2.6.10. Let G be the cyclic group $\mathbb{Z}/10\mathbb{Z}$ and g its generator. Let H be the graph of Figure 6, where the vertices are represented by the exponents of the generator g. Let $A = \{g^2, g^3\}$, so then $S = \{g^2, g^3, g^{-2}, g^{-3}\}$. Then the graph in Figure 6 is the Cayley graph $\mathrm{Cay}(G, S)$.

Theorem 2.6.11. *[22, Theorem 2.1.] Let $\mathcal{O}$ be an imaginary quadratic order of discriminant $\Delta < 0$ and conductor f. Set $G = \mathrm{cl}(\mathcal{O})$ and let B and x be real numbers satisfying $B > 2$ and $x \geq (\ln|\Delta|)^B$. Let S_x be the symmetric set $A \cup A^{-1}$ where*

$$A = \{[\mathfrak{p}] \in G \mid \gcd(f, \mathfrak{p}) = 1, \text{ and } \mathrm{N}(\mathfrak{p}) \leq x \text{ is a prime}\},$$

where $\mathrm{N}(\mathfrak{p})$ denotes the norm of $\mathfrak{p}$. Then, assuming the Generalized Riemann Hypothesis, there exists a positive absolute constant $C > 1$, depending only on B, such that for all Δ, a random walk of length

$$t \geq C\frac{\ln|G|}{\ln\ln|\Delta|}$$

in the Cayley graph $\mathrm{Cay}(G, S_x)$ from any starting vertex lands in any fixed subset $S \in G$ with probability at least $\frac{1}{2}\frac{|S|}{|G|}$.

Proof. Apply [44, Corollary 1.3] with the parameters

- $K =$ the field of fractions of $\mathcal{O}$,
- $G = \mathrm{cl}(\mathcal{O})$,
- $q = |\Delta|$. □

We will later see that even though this theorem is for the Cayley graph of the class group of $\mathcal{O}$, in the last section we will see that there is a generalization of Cayley graphs to group actions called Schreier graphs, where this theorem will be applicable to the CSIDH graph as well. Let us now state what this CSIDH graph is.

Definition 2.6.12. [17, Definition 2.] For a field k and a $L = \{\ell_1, \ell_2, \ldots, \ell_n\}$ list of primes $\ell_i \nmid \mathrm{char}(k)$, the **$k$-rational isogeny graph** $G_{k,L}$ is a graph, whose vertices are (representatives of) the isomorphism class of elliptic curves defined over k, and having a directed edge (E_1, E_2) if there is a k-rational ℓ_i-isogeny, $\ell_i \in L$, between E_1 and E_2. If there is only one prime ℓ, then we denote the isogeny graph by $G_{k,\ell}$.

Remark 2.6.13. [70, Remark 22.2.] By definition, the isogeny graph $G_{k,L}$ is a directed graph, but by Definition 2.2.20, every isogeny has a dual isogeny, so there should be equal amount of outgoing and incoming edges between two elliptic curves. Unfortunately, if the j-invariant is 0 or 1728, then those curves have extra automorphisms. To see why this is a problem, let E_1, E_2 be two elliptic curves and $\varphi : E_1 \to E_2$ an isogeny between them. Normally, the only automorphisms are 1 and -1, in which case the isogenies $\varphi \circ [-1] = [-1] \circ \varphi$ have the same kernel, so they're the same. The automorphism groups of $j = 0$ and $j = 1728$ are 6 and 4 respectively, and have automorphisms $\rho : (x, y) \mapsto (\rho x, -y)$ and $i : (x, y) \mapsto (-x, iy)$ respectively. (see Example 2.1.21) These automorphisms fix only two cyclic subgroups of $E[\ell]$, and thus in the case of $j(E_1) = 0, j(E_2) \neq 0$, the isogenies $\varphi, \varphi \circ \rho$ and $\varphi \circ \rho^2$ have different kernels, but the corresponding dual isogenies have the same kernels. Likewise for $j = 1728$, but with φ and $\varphi \circ i$ instead.

Remark 2.6.14. [17, Remark 3.] If the isogeny graph is defined over k, then the two j-invariants from Remark 2.6.13 have more outgoing edges if and only if $j(E) = 0$ and $\sqrt{-3} \in k$, or if $j(E) = 1728$ and $\sqrt{-1} \in k$. To avoid this case, we need to have that the Legendre symbols $(\frac{-3}{p})$ and $(\frac{-1}{p})$ both evaluate to -1 (i.e. they're not squares in k). This happens when $p \equiv 11 \pmod{12}$, so taking a prime satisfying this relation is enough to view $G_{k,L}$ as an undirected graph.

We will assume from now on that the conditions of Remark 2.6.14 are met, so we can assume that the isogeny graphs are undirected.

Theorem 2.6.15. *[70, Theorem 13.2.] Let E_1, E_2 be two elliptic curves with $\varphi : E_1 \to E_2$ an isogeny between them. Then E_1 is supersingular if and only if E_2 is supersingular.*

Proof. Let $p_1 \mathrm{End}_k(E_1)$ and $p_2 \in \mathrm{End}_k(E_2)$ denote the multiplication-by-p maps on E_1 and E_2 respectively. We thus have that

$$\begin{aligned} p_2 \circ \varphi &= \varphi \circ p_1 \\ \deg_s(p_2 \circ \varphi) &= \deg_s(\varphi \circ p_1) \\ \deg_s(p_2)\deg_s(\varphi) &= \deg_s(\varphi)\deg_s(p_1) \\ \deg_s(p_2) &= \deg_s(p_1). \end{aligned}$$

By Definition 2.2.24, E being supersingular is the same as having $E[p] = \{\infty\}$, so in particular $|\ker p| = 1$. By Corollary 2.2.15, $|\ker p| = \deg_s[p] = 1$, so if E_1 is supersingular, then on E_2, p_1 must have separable degree 1 as well, proving the theorem. □

Remark 2.6.16. The previous theorem follows immediately from Tate's Isogeny Theorem 2.5.20, but since proving Tate's theorem is definitely out of scope for this thesis, the alternative proof for Theorem 2.6.15 is given as a compensation.

These isogeny graphs are a disjoint union of connected components, where the components are determined by the number of k-rational points, as given by Tate's Isogeny Theorem 2.5.20. Since the number of points also dictates whether an elliptic curve is ordinary or supersingular (Definition 2.1.24), the connected components can be divided into ordinary and supersingular components. The ordinary components have been studied by Kohel [45] in his PhD thesis, and later on they have been named *isogeny volcanoes* by Fouquet and Morain [34], based on their appearance (cf. Figure 8).

We will not spend much time with the ordinary components, as CSIDH uses the supersingular component, with a *twist*. The supersingular component over the algebraic closure resembles more of an *isogeny yarnball* rather than a volcano, as can been seen in Figure 9. An example of an isogeny graph over $\mathbb{F}_p$ (instead of $\overline{\mathbb{F}}_p$) that mimics the prime field setup of CSIDH with $k = \mathbb{F}_{659}$ and $L = \{3, 5, 11\}$ is shown in Figure 7.

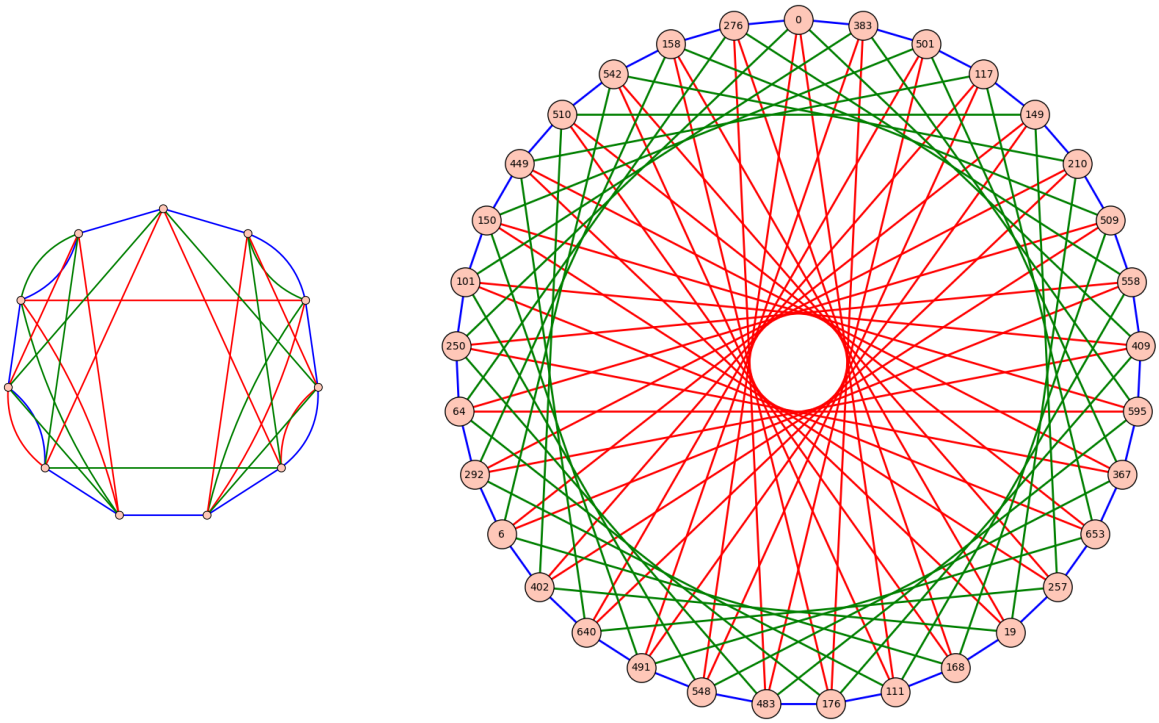


Figure 7: The supersingular components of the isogeny graph $G_{\mathbb{F}_p,L}$ with $p = 659$ and $L = \{3, 5, 11\}$. The left side corresponds to the ring of integers $\mathbb{Z}[\frac{1+\sqrt{-659}}{2}]$, and the right side corresponds to the elliptic curves with endomorphism ring $\mathbb{Z}[\sqrt{-659}]$, where the vertices are labeled by the curves' Montgomery coefficients. Figure created with SageMath [39].

The highly-connected structure of the full supersingular isogeny graph is due to the following proposition.

Proposition 2.6.17. *Let $G_{\overline{\mathbb{F}}_p,\ell}$ be the full supersingular component of the isogeny graph of some ℓ-isogeny $2 < \ell < p$. Then $G_{\overline{\mathbb{F}}_p,\ell}$ is an $\ell + 1$-regular graph.*

Let us now investigate the structure of these isogeny graphs $G_{k,\ell}$ for different fields k and primes ℓ in relation to the conductor of the endomorphism ring. We begin with a theorem and some terminology regarding the main parts of an isogeny graph.

Theorem 2.6.18. *[45, Proposition 21.] Let $\varphi : E(k) \to E'(k)$ be an ℓ-isogeny defined over k. Let $\mathcal{O}$ and $\mathcal{O}'$ be the corresponding orders of $E(k), E(k)'$ in the imaginary quadratic field $K = \mathrm{End}^0(E) \cong \mathbb{Q}(\sqrt{\Delta_K})$, then one of the following holds:*

(i) $\mathcal{O} = \mathcal{O}'$,

(ii) $[\mathcal{O} : \mathcal{O}'] = \ell$,

(iii) $[\mathcal{O}' : \mathcal{O}] = \ell$.

Proof. [70, Theorem 22.3.] By Tate's Isogeny Theorem 2.5.20, we have that since E, E' are isogenous, their endomorphism algebras are isomorphic as well, so it is justified to state the theorem as both orders $\mathcal{O}$ and $\mathcal{O}_K$ being in K. Let these orders be generated as $\mathcal{O} = [1, \tau]$ and $\mathcal{O}' = [1, \tau']$ respectively. Then we have that the composition $\varphi \circ \tau \circ \hat{\varphi}$ is an element of $\mathrm{End}_k(E') = \mathcal{O}'$, and in particular it has the same characteristic polynomial as $\ell\tau \in \mathcal{O}$, which implies that $\ell\tau \in \mathcal{O}'$. A similar argument applies in the other direction too, so we also have $\ell\tau' \in \mathcal{O}$. We thus have

$$[1, \ell\tau] \subseteq [1, \tau'] \quad \text{and} \quad [1, \ell\tau'] \subseteq [1, \tau] \implies [1, \ell^2\tau] \subseteq [1, \ell\tau'] \subseteq [1, \tau].$$

The index of $[1, \ell^2\tau]$ in $[1, \tau]$ is ℓ^2, so by the tower law, we must have that the index of $[1, \ell\tau']$ in $[1, \tau]$ is either 1, ℓ, or ℓ^2. If the index is 1, then $[1, \ell\tau'] = [1, \tau]$, so $[\mathcal{O}' : \mathcal{O}] = \frac{\#[1,\tau']}{\#[1,\tau]} = \frac{\#[1,\tau']}{\#[1,\ell\tau']} = \ell$, which corresponds to case (iii). The other two options are calculated in a similar way. □

Definition 2.6.19. Let $E(k), E'(k)$ be two elliptic curves with orders $\mathcal{O}, \mathcal{O}'$ respectively, and $\varphi : E(k) \to E'(k)$ be a k-rational ℓ-isogeny between them. Then we give names to the cases of Theorem 2.6.18 by saying that if

- $\mathcal{O} = \mathcal{O}'$, then φ is a **horizontal** isogeny,
- $[\mathcal{O}' : \mathcal{O}] = \ell$, then φ is a **ascending** isogeny,
- $[\mathcal{O} : \mathcal{O}'] = \ell$, then φ is a **descending** isogeny.

The cycles of horizontal isogenies as a subgraph on an isogeny volcano can be divided into **levels** $V_0, \ldots, V_d$ based on their endomorphism rings. The topmost level V_0, which has no ascending isogenies, is called the **surface** and the lowest level V_d, which has no descending isogenies, is called the **floor**.

Remark 2.6.20. Notice that when considering primes ℓ, ascending and descending isogenies can exist only for primes that divide the conductor of $\mathbb{Z}[\pi_k]$, since by (an extension of) Proposition 2.5.21, E is defined over k if and only if $\mathbb{Z}[\pi_k]$ is in $\mathrm{End}_k(E)$, which means that it lies on the floor of the isogeny graph. Therefore the horizontal case is the generic one, but this fact can also be used to define an algorithm to find the floor of any isogeny together with its depth. (see [45, pp. 46])

With this definition, we can investigate the isogeny volcano of Figure 8 again from a bird's eye view. The levels V_0, V_1, V_2 correspond to the concentric circles of vertices, starting from the inside. The innermost circle is the surface, with a simple 2-regular cycle of horizontal isogenies. The other two circles are then connected to each other *vertically* by ascending and descending isogenies with a ratio of $1 : 3$. Counting the degree of non-floor vertices, we see that they are all $4 = \ell + 1$, which is also true for the isogeny yarnball of Figure 9, even if it is hard to count. This is no coincidence, since this behavior is coming from an easy lemma for the isogeny yarnball, and from a general theorem about elliptic curves defined over $\mathbb{C}$, which is then reduced to finite fields by Deuring's Reduction Theorem [47, Theorem 13.4.12] for the isogeny volcano (ordinary) case.

Lemma 2.6.21. *Let $E(\overline{k})$ be an elliptic curve defined over an algebraically closed field $\overline{k}$. There are exactly $\ell + 1$ distinct ℓ-isogenies with domain E.*

Proof. From Theorem 2.2.23, we know that for primes $\ell \nmid \mathrm{char}(k)$, we have that the ℓ-torsion subgroup splits as

$$E[\ell] = \mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}.$$

From Theorem 2.3.1, we know that separable isogenies are uniquely defined by their kernel, so to count how many distinct ℓ-isogenies go out of E, we need to count the number of cyclic subgroups of $E[\ell]$ that have order ℓ. We have that $|E[\ell]| = \ell^2$, but then removing the point at infinity gives $\ell^2 - 1$ non-zero points. Every subgroup of order ℓ has $\ell - 1$ generators by Euler's totient function, so the total number of cyclic subgroups is $(\ell^2 - 1)/(\ell - 1) = \ell + 1$. □

Theorem 2.6.22. *[45, Proposition 23] Let $E(\mathbb{C})$ be an elliptic curve defined over $\mathbb{C}$, $\mathcal{O}$ its endomorphism ring with discriminant $\Delta = f^2\Delta_K$ and conductor f in an imaginary quadratic field $K = \mathbb{Q}(\sqrt{\Delta_K})$, and let ℓ be a prime. We have two possibilities depending on the prime factorization of f:*

(i) If $\ell \mid f$, then E has 0 horizontal, ℓ descending and 1 ascending ℓ-isogenies.

(ii) If $\ell \nmid f$, then E has $1 + \left(\frac{\Delta}{\ell}\right)$ horizontal, $\ell - \left(\frac{\Delta}{\ell}\right)$ descending and 0 ascending ℓ-isogenies.

Proof. Since we did not introduce elliptic curves over $\mathbb{C}$ in this thesis, the proof can be found in Andrew Sutherland's lecture notes [70, Theorem 22.5.]. □

The previous lemma describes the structure of the full supersingular isogeny graph, and in particular, it is an $\ell + 1$-regular graph. Pizer [57] has proved that these isogeny yarnballs are actually Ramanujan graphs, so they are a great basis for cryptographic schemes. This is exactly the reasoning behind SIDH,

CSIDH's distant cousin, for using these full supersingular graphs. For a difference between SIDH and CSIDH, see Section 1.3, and for a brief overview of SIDH, see Subsection 3.1.3 .

As mentioned before, the ordinary components of isogeny graphs is characterized even further in Kohel's PhD thesis [45]. We will not spend more time on ordinary components, as this thesis is based on understanding CSIDH. Since CSIDH uses a supersingular isogeny graph, but defined over $\mathbb{F}_p$ instead of $\overline{\mathbb{F}}_p$, so let us investigate the structure of the supersingular components of $G_{\mathbb{F}_p,\ell}$.

Let $E = E(\mathbb{F}_p)$ be a supersingular elliptic curve defined over $\mathbb{F}_p$ from now on until the end of this chapter. We have seen in Theorem 2.5.23 that the possible endomorphism rings of E is quite scarce: it is either $\mathbb{Z}[\sqrt{-p}]$ or $\mathbb{Z}[\frac{1+\sqrt{-p}}{2}]$ based on the value of $p \pmod 4$ and its conductor f. If we lift $E(\mathbb{F}_p)$ to a number field by the Deuring Lifting Theorem, then we can construct an infinite volcano of elliptic curves whose endomorphism rings are orders in $\mathbb{Q}(\sqrt{-p})$. In this setting, we can apply Theorem 2.6.22, and then since the conductor of $\mathcal{O}$ is 1 or 2, the only interesting prime is $\ell = 2$ when $p \equiv 3 \pmod 4$ and $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$. This shows that the levels of these volcanoes are only connected by 2-isogenies. We can then reduce the volcano modulo p again, and this action keeps the curves supersingular (by a part of Deuring Reduction Theorem [47, Theorem 13.4.12]) and the isogenies are also defined over $\mathbb{F}_p$ (by [31, Proposition 2.6]). We can completely categorize the supersingular isogeny graph $G_{\mathbb{F}_p,\ell}$ as follows.

Theorem 2.6.23. *[31, Theorem 2.7.] Let $p > 3$ be a prime and let $G_{\mathbb{F}_p,\ell}$ be the supersingular isogeny graph and $h(\Delta)$ the class number of the order $\mathcal{O}$ with discriminant Δ. Then we have the following cases:*

1. $p \equiv 1 \pmod 4$: *There are $h(-4p)$ $\mathbb{F}_p$-isomorphism classes of supersingular elliptic curves over $\mathbb{F}_p$, and they all have $\mathrm{End}_p(E) \cong \mathbb{Z}[\sqrt{-p}]$. From every vertex of the graph, there is one outgoing $\mathbb{F}_p$-rational horizontal 2-isogeny, and two horizontal ℓ-isogenies for every prime $\ell > 2$ with $\left(\frac{-p}{\ell}\right) = 1$.*
2. $p \equiv 3 \pmod 4$: *There are two levels; the surface and the floor. From each vertex, there are two horizontal ℓ-isogenies for every prime $\ell > 2$ with $\left(\frac{-p}{\ell}\right) = 1$.*
   (a) *If $p \equiv 7 \pmod 8$, then each level has $h(-p)$ vertices. The surface and floor are connected 1:1 with 2-isogenies, and on the surface we also have two horizontal 2-isogenies from each vertex.*
   (b) *If $p \equiv 3 \pmod 8$, then the surface has $h(-p)$ vertices, and the floor has $3h(-p)$ vertices. The surface and floor are connected 1:3 with 2-isogenies, and there are no horizontal 2-isogenies.*

Proof. From the discussion above, we know that Theorem 2.6.22 is applicable to our case. By Proposition 2.1.17, we have that $\mathbb{Z}[\pi] = \mathbb{Z}[\sqrt{-p}]$ will be the order which defines the floor of the supersingular isogeny graph, because when going through the reduction, the orders situated below $\mathbb{Z}[\sqrt{-p}]$ are of the form $\mathbb{Z}[\ell^r\sqrt{-p}]$ for some $r > 0$. Since these do not contain $\sqrt{-p}$, they will not show up as vertices in our graph. Furthermore, we also saw in Theorem 2.5.23 that for $p \equiv 1 \pmod 4$, the only endomorphism ring is $\mathbb{Z}[\sqrt{-p}]$, therefore for all ℓ, we're in case (ii) of Theorem 2.6.22. Since $\Delta = -4p$ in this case, and we are on the floor, from every vertex there are $1 + \left(\frac{-4p}{\ell}\right)$ horizontal isogenies, and no descending or ascending isogenies. The number of horizontal isogenies then follows from how the Legendre symbol works, concluding case (1).

For the second part, we proceed by analyzing the ℓ-torsion points of E. We have that the $\mathbb{F}_p$-rational ℓ-isogenies correspond to Galois-invariant cyclic subgroups of $E[\ell]$. We know from Theorem 2.2.23 that $E[\ell]$ is a 2-dimensional vector space $\mathbb{F}_\ell^2$, and π acts linearly on $E[\ell]$. If we fix a basis $\{P, Q\}$, then this action is represented by a 2×2 matrix, which also satisfies the Frobenius characteristic polynomial $\pi^2 + p = 0$ modulo ℓ. There are three possibilities for what happens:

1. it factors as $(\pi - a)^2$,
2. it factors as $(\pi - a)(\pi - b)$ with $a \not\equiv b \pmod \ell$,
3. it is irreducible.

Suppose there is a cyclic subgroup $G = \langle P \rangle$ of $E[\ell]$ with $\pi(G) = G$. Then it follows that $\pi(P) = [a]P$ in G for some integer a. We have that the linear map π has eigenspace G with eigenvalue a, and so the characteristic polynomial has a root $(\pi - a)$. Changing the basis now to use this point P and some other point Q means the 2×2 matrix is represented by either

$$\begin{pmatrix} a & 0 \\ b & a \end{pmatrix} \text{ or } \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

depending on whether the characteristic polynomial ramifies or splits respectively.

In the case when the polynomial $x^2 + p \pmod{\ell}$ splits, we clearly have 2 cyclic Galois-invariant subgroups, and 0 when the polynomial stays irreducible. Furthermore, we are only interested in ℓ such that $\left(\frac{-p}{\ell}\right) = 1$, since otherwise $\ell\mathcal{O}$ is inert in $\mathbb{Q}(\sqrt{-p})$ and there are no prime ideals of norm ℓ in $\mathbb{Z}[\sqrt{-p}]$, so there is no edge in that graph.

Lastly, the polynomial $x^2 + p \pmod{\ell}$ can only have a repeated root when $\ell = 2$, because then $x^2 + p \equiv (x + \sqrt{p})^2 \pmod{\ell}$, provided that p is a quadratic residue. Taking a look at of the matrix $\begin{pmatrix} a & 0 \\ b & a \end{pmatrix}$ again, we see that if $b \equiv 0 \pmod 2$, we get three subgroups of $E[2]$, but we also have that $b \equiv 0 \pmod 2$ if and only if $\pi(P) = P$ and $\pi(Q) = Q$. This means that the whole 2-torsion is included in $\ker(1 + \pi)$. Since the multiplication-by-2 map is separable (Example 2.2.8), there exists a unique isogeny $\varphi \in \mathrm{End}(E)$ such that $1 + \pi = 2\varphi$ by [69, Corollary III.4.11.]. We then have that φ is $\mathbb{F}_p$-rational since it is a quotient of $\mathbb{F}_p$-rational maps, so $\varphi \in \mathrm{End}_p(E)$. This is equivalent to saying that the order on the surface $\mathbb{Z}[\frac{1+\sqrt{-p}}{2}]$ has three outgoing 2-isogenies. The other case when $b \equiv 1 \pmod 2$ in the matrix has 1 Galois-invariant subgroup, and a similar argument show that it corresponds to $\mathbb{Z}[\sqrt{-p}]$ having one outgoing 2-isogeny.

The number of $\mathbb{F}_p$-isomorphism classes follows from the fact that the classes are in bijection with the ideal class group $\mathrm{cl}(\mathcal{O})$ of its order, so their sizes are equal. This fact will be proved in Theorem 2.7.13, and the ideal class group size differences are explained in Corollary 2.4.38 for part 2(c), otherwise they follow from Theorem 2.4.37. □

This theorem also sheds a new light onto the fact that in Example 2.4.23, we showed that $\mathfrak{l}_2 = (2, 1 + \sqrt{-p})$ was not an invertible fractional $\mathcal{O}$-ideal of $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$. Indeed, $\mathcal{O}$ is not the ring of integers in $\mathbb{Q}(\sqrt{-5})$, so it is the endomorphism ring of the floor in the isogeny graph $G_{\mathbb{F}_p,2}$, therefore there are no horizontal 2-isogenies. Even if E, whose endomorphism ring is $\mathcal{O}$, is an ordinary curve, Theorem 2.6.22 provides the same conclusion, since the conductor of $\mathcal{O}$ is 2, and $\mathrm{N}(\mathfrak{l}_2) = 2$ divides that. This is no coincidence, and is the topic of the next section.

2.7 Group Actions and Torsors

In the previous section we have developed a lot of mathematical machinery, such as giving a categorization of the possible endomorphism rings, and describing the structure of the supersingular isogeny graph over $\mathbb{F}_p$. We can feel proud of ourselves, and march proudly forward. In this section we connect everything together, and arrive at the finish line of understanding the core promise of CSIDH, which is the ideal class group action. This section is based mostly on [70, 38].

Let us start by giving some definitions.

Definition 2.7.1. Let G be a group with identity element e and X a set. The **(left) group action** of G on X is a function from the Cartesian product of G and X into X:

$$\begin{aligned} * : G \times X &\longrightarrow X \\ (g, x) &\longmapsto g * x, \end{aligned}$$

satisfying the following two axioms:

(i) $e * x = x$,
(ii) $(gh) * x = g * (h * x)$.

In this case, we say that G **acts on** X.

Definition 2.7.2. We say that a group action is **free** if there are no non-identity elements in G that fixes any element of X. Notationally, for all $g \in G, x \in X$, we have that

$$g * x = x \iff g = e.$$

Definition 2.7.3. We say that a group action is **transitive** if for any two elements $x, y \in X$, there exists an element $g \in G$ such that $g * x = y$.

Definition 2.7.4. If a group action is both free and transitive, then it is called **regular**. In this case, the set X is said to be a G-**Torsor** or a **principal homogeneous space** for G.

The astute reader might have noticed that these definitions mirror the notions of injectivity, surjectivity and bijectivity respectively. Indeed, there is an important connection to group actions that is quite important.

Theorem 2.7.5. *Let G be a group acting on X. If the action is free and transitive, then $G \cong X$ as sets, and in particular if G is finite, then $|G| = |X|$.*

Proof. Fix $x_0 \in X$, and let $f : G \to X$ be a map defined by $f(g) = g * x_0$. To prove f is a bijection, we show that it's both injective and surjective.

We start with injectivity. Assume $f(g_1) = f(g_2)$ for some $g_1, g_2 \in G$, we then have that

$$\begin{aligned}
f(g_1) = f(g_2) &\iff g_1 * x_0 = g_2 * x_0 \\
&\iff g_1^{-1} * (g_1 * x_0) = g_1^{-1} * (g_2 * x_0) \\
&\iff (g_1^{-1} g_1) * x_0 = (g_1^{-1} g_2) * x_0 && \textit{Def. 2.7.1(ii)} \\
&\iff e * x_0 = (g_1^{-1} g_2) * x_0 \\
&\iff e = (g_1^{-1} g_2) && \textit{Def. 2.7.2} \\
&\iff g_1 = g_2.
\end{aligned}$$

For surjectivity, we need to show that for every $y \in X$, there exists some $g \in G$ such that $f(g) = y$. This follows immediately from the assumption that the action is transitive, because there exists at least one element $g \in G$ for which $g * x_0 = y$, but this is the definition of $f(g) = g * x_0 = y$. □

Next, we will investigate the interplay of fractional $\mathcal{O}$-ideals and isogenies. Recall that for an elliptic curve $E(\mathbb{F}_p)$, we have that $\mathrm{End}_p(E) \cong \mathcal{O}$ is an order in an imaginary quadratic field. Also recall from Section 2.4 that $\mathfrak{a}$ denotes an $\mathcal{O}$-ideal whose norm is defined as $\mathrm{N}(\mathfrak{a}) = \gcd(\{\, \mathrm{N}(\alpha) \mid \alpha \in \mathfrak{a} \,\})$. Just like in Definition 2.2.22, for an elliptic curve E we are able to define the $\mathfrak{a}$-torsion point as well.

Definition 2.7.6. Let $E(\mathbb{F}_p)$ be an elliptic curve defined over $\mathbb{F}_p$, and let $\mathrm{End}(E) \cong \mathcal{O}$. Let $\mathfrak{a} \subset \mathcal{O}$ be an invertible ideal of norm coprime to p. We define the $\mathfrak{a}$-**torsion subgroup** of E as

$$E[\mathfrak{a}] = \{\, P \in E(\overline{\mathbb{F}}_p) \mid \alpha(P) = 0 \text{ for all } \alpha \in \mathfrak{a} \,\}.$$

Note that if $\mathfrak{a} \neq (0)$, then $E[\mathfrak{a}]$ is finite since it is the intersection of the kernel of all elements in $\mathfrak{a}$, and the kernels are clearly finite subgroups. As a result of this, by Theorem 2.3.1, there exists an elliptic curve $E/\mathfrak{a}$ and a separable isogeny $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ with $\ker \varphi_{\mathfrak{a}} = E[\mathfrak{a}]$. Also note that since $\mathfrak{a}$ is an $\mathcal{O}$-ideal, the elements of $\mathfrak{a}$ are endomorphisms.

We are now going to slowly work our way towards proving the main theorem of this section, Theorem 2.7.13, which is that the class group $\mathrm{cl}(\mathcal{O})$ acts freely and transitively on the set of isomorphism classes of supersingular elliptic curves over $\mathbb{F}_p$, which we will denote by $Ell_p(\mathcal{O}, \pi)$.

Lemma 2.7.7. *Let $\mathfrak{a} \subseteq \mathcal{O}$ be a proper $\mathcal{O}$-ideal and $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ be the separable isogeny with kernel $E[\mathfrak{a}]$. Then we have that $\deg(\varphi_{\mathfrak{a}}) = \mathrm{N}(\mathfrak{a})$.*

Proof. Since $\varphi_{\mathfrak{a}}$ is separable, we have that $\deg \varphi_{\mathfrak{a}} = |\ker \varphi_{\mathfrak{a}}| = |E[\mathfrak{a}]|$, so we need to only show that $|E[\mathfrak{a}]| = \mathrm{N}(\mathfrak{a})$. By Corollary 2.4.27, Theorem 2.4.29 and the fact that the degree map is multiplicative, it is enough to show this for principal and prime $\mathcal{O}$-ideals. To this end, let $\mathfrak{a} = (\alpha)$ be a principal ideal generated by $\alpha \in \mathcal{O}$. Then clearly $E[(\alpha)] = \ker(\alpha)$, but we also have that $\deg(\alpha) = N(\alpha)$ so we have

$$\deg \varphi_{(\alpha)} = |\ker \alpha| = \deg(\alpha) = \mathrm{N}(\alpha) = \mathrm{N}((\alpha)).$$

Now let $\mathfrak{p} \subseteq \mathcal{O}$ be a prime ideal. We can examine what happens to $p_0 = \mathfrak{p} \cap \mathbb{Z}$ in the order. There are three cases to consider, as usual:

Case 1.: If the element p_0 is inert, then it remains a prime in $\mathcal{O}$, so $\mathfrak{p} = (p_0)$ is a principal ideal, which is covered by the previous discussion.

Case 2.: If the element p_0 splits, then we have $(p_0) = \mathfrak{p}\overline{\mathfrak{p}}$ with $\mathfrak{p} \neq \overline{\mathfrak{p}}$ and $\mathrm{N}(\mathfrak{p}) = p_0$. Since the ideals are coprime, the p_0-torsion group decomposes as a direct sum:

$$E[p_0] = E[\mathfrak{p}] \oplus E[\overline{\mathfrak{p}}].$$

Taking the size of both sets gives by Theorem 2.2.23

$$p_0^2 = |E[\mathfrak{p}]| \cdot |E[\overline{\mathfrak{p}}]|.$$

Since $\mathfrak{p}$ and $\overline{\mathfrak{p}}$ are conjugate, their torsion sizes are equal, so indeed $|E[\mathfrak{p}]| = p_0 = \mathrm{N}(\mathfrak{p})$.

Case 3.: If the element p_0 ramifies, then $(p_0) = \mathfrak{p}^2$, and $\mathrm{N}(\mathfrak{p}) = p_0$. The inclusion of ideals $(p_0) = \mathfrak{p}^2 \subset \mathfrak{p} \subset \mathcal{O}$ induces a reversed inclusion

$$\{\infty\} \subset E[\mathfrak{p}] \subset E[\mathfrak{p}^2] = E[p_0].$$

Taking the sizes of these subgroups, Lagrange's theorem forces $|E[\mathfrak{p}]| = p_0$, showing that all three cases satisfy the statement of the theorem. □

Lemma 2.7.8. *[38, Lemma 3.3.] Let $E(\mathbb{F}_p)$ be an elliptic curve over $\mathbb{F}_p$, and $\mathfrak{p} \subseteq \mathcal{O} = \mathbb{Z}[\sqrt{-p}]$ be a proper $\mathcal{O}$-ideal. Let $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ be the associated isogeny, and π, π' the Frobenius endomorphism on E and E'. Then*

$$\varphi_{\mathfrak{a}} \circ \pi = \pi \circ \varphi_{\mathfrak{a}}.$$

In other words, the isogeny associated to $\mathfrak{a}$ and its image curve are both defined over $\mathbb{F}_p$.

Proof. We have that $E[\mathfrak{a}]$, by definition, is the intersection of the kernel of all elements of $\mathfrak{a}$, which can be expressed as $a + b\pi$ for some $a, b \in \mathbb{Z}$. Let $P \in \ker(a + b\pi)$, hen we have that

$$\infty = \pi(\infty) = \pi(aP + b\pi(P)) = a\pi(P) + b\pi(\pi(P)) = (a + b\pi)(\pi(P)).$$

This shows that $\pi(P)$ is also in the kernel of $a + b\pi$, from which it follows that $\pi(E[\mathfrak{a}]) = E[\mathfrak{a}]$. This then means that all points are invariant under the Frobenius endomorphism, so they are defined over $\mathbb{F}_p$, and by Vélu's formulae, the isogeny is a composition of function with $\mathbb{F}_p$-rational coefficients, therefore the isogeny is defined over $\mathbb{F}_p$ as well. □

If we fix E to be supersingular, then by the previous theorem and Theorem 2.6.15, we have that $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ is also supersingular, and defined over $\mathbb{F}_p$. This is great, because this suggests that there might be a group action from $\mathcal{O}$-ideals to supersingular curves over $\mathbb{F}_p$. To this end, let us define for an $\mathcal{O}$-ideal $\mathfrak{a}$ the group action $\mathfrak{a} * E = E/\mathfrak{a}$, where $E/\mathfrak{a}$ is the supersingular elliptic curve $E/E[\mathfrak{a}]$, as seen from before. It is not enough to just proclaim this, so we will prove this via the axioms.

Lemma 2.7.9. *A non-zero principal $\mathcal{O}$-ideal $\mathfrak{a} = (\alpha)$ is associated with an endomorphism.*

Proof. Every element of (α) is a multiple of α, so it has kernel containing $\ker(\alpha)$. Conversely, since $\alpha \in \mathfrak{a}$, we have $E[I] \subseteq \ker(\alpha)$ by the reverse inclusion, therefore $E[\mathfrak{a}] = \ker(\alpha)$. Let $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ be the uniquely defined isogeny with kernel $\ker(\alpha)$. Since α is an endomorphism by definition with the same kernel $\ker(\alpha)$, Theorem 2.3.1 says that we must have that $\varphi_{\mathfrak{a}} = \alpha$ up to isomorphism, so $\varphi_{\mathfrak{a}}$ is indeed an endomorphism. □

Lemma 2.7.10. *Let $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O} = \mathbb{Z}[\sqrt{-p}]$ be non-zero ideals. Then $\varphi_{\mathfrak{ab}} = \varphi_{\mathfrak{b}} \circ \varphi_{\mathfrak{a}}$.*

Proof. The strategy to prove this is to show that their kernels are the same. We will prove this by showing both inclusions hold. Assume first that $P \in \ker(\varphi_{\mathfrak{b}} \circ \varphi_{\mathfrak{a}})$, then clearly $\varphi_{\mathfrak{a}}(P) \in \ker \varphi_{\mathfrak{b}}$. For any element $\beta \in \mathfrak{b}$, we have that $\beta(\varphi_{\mathfrak{a}}(P)) = \infty$. By Lemma 2.7.8, we have that $\varphi_{\mathfrak{a}}$ commutes with all

elements of $\mathbb{Z}[\sqrt{-p}]$, so we can reverse the order $\varphi_{\mathfrak{a}}(\beta(P)) = \infty$, also have that for all $\alpha \in \mathfrak{a}$, it holds that $\alpha(\beta(P)) = \infty$. This shows that P is also in $\ker(\varphi_{\mathfrak{ab}})$.

Now assume that $P \in \ker \varphi_{\mathfrak{ab}}$. Then the same line of argument works; for all $\alpha \in \mathfrak{a}, \beta \in \mathfrak{b}$, we have that

$$\alpha(\beta(P)) = \infty \implies \beta(P) \in E[\mathfrak{a}] \implies \varphi_{\mathfrak{a}}(\beta(P)) = \infty.$$

Again, by Lemma 2.7.8 we can swap the order of β and $\varphi_{\mathfrak{a}}$, so then fo all β,

$$\beta(\varphi_{\mathfrak{a}}(P)) = \infty \implies \varphi_{\mathfrak{a}}(P) \in E[\mathfrak{b}] \implies P \in \ker(\varphi_{\mathfrak{b}} \circ \varphi_{\mathfrak{a}}),$$

so the reverse inclusion holds as well, therefore by the uniqueness of the kernel of an isogeny, the statement holds. □

These two lemmas were actually the two axioms from the group action Definition 2.7.1. Now we will prove two lemmas that concern the well-definedness of the group action, which is always necessary when one works with classes, and in this case, we have not only an ideal *class* group, but also isomorphism *classes* of supersingular elliptic curves.

Lemma 2.7.11. *Let $\mathfrak{a}, \mathfrak{b} \in \mathcal{O}$ be two non-zero proper $\mathcal{O}$-ideals. Suppose there exist $\alpha, \beta \in \mathcal{O}$ such that $(\alpha)\mathfrak{a} = (\beta)\mathfrak{b}$. Let $\varphi_{\mathfrak{a}} : E \to E/\mathfrak{a}$ and $\varphi_{\mathfrak{b}} : E \to E/\mathfrak{b}$ be their associated isogenies. Then $E/\mathfrak{a} \cong E/\mathfrak{b}$.*

Proof. If $(\alpha)\mathfrak{a} = (\beta)\mathfrak{b}$, then we have that their associated isogenies are also equal,

$$\varphi_{(\alpha)\mathfrak{a}} = \varphi_{(\beta)\mathfrak{b}},$$

and by Lemma 2.7.10, we can factor them as

$$\varphi_{\mathfrak{a}} \circ \varphi_{(\alpha)} = \varphi_{\mathfrak{a}} \circ \alpha = \alpha \circ \varphi_{\mathfrak{a}},$$

and similarly $\varphi_{(\beta)\mathfrak{b}} = \beta \circ \varphi_{\mathfrak{b}}$. Note that by Lemma 2.7.9, the principal ideals are endomorphisms. Evaluating these isogenies end up in $E/\mathfrak{a}$ and $E/\mathfrak{b}$ respectively, but by assumption

$$\varphi_{(\alpha)\mathfrak{a}} = \varphi_{(\beta)\mathfrak{b}},$$

which means that their target curve is also isomorphic, so $E/\mathfrak{a} \cong E/\mathfrak{b}$, as required. □

Lemma 2.7.12. *Let E_1, E_2 be supersingular elliptic curves over $\mathbb{F}_p$, such that $E_1 \cong E_2$ over $\mathbb{F}_p$. Let $\mathfrak{a}$ be an $\mathcal{O}$-ideal in $\mathbb{Z}[\sqrt{-p}]$. Then $\mathfrak{a} * E_1 \cong \mathfrak{a} * E_2$.*

Proof. Let $\alpha : E_1 \to E_2$ be an isomorphism defined over $\mathbb{F}_p$, and let $\varphi_{\mathfrak{a}} : E_2 \to E_3 = \mathfrak{a} * E_2$. We want to show that two curves are isomorphic, so we can proceed as in the previous lemmas, so we need to show that $\ker(\varphi_{\mathfrak{a}} \circ \alpha) = E_1[\mathfrak{a}]$.

Let $P \in E_1$. Then it is in the kernel

$$P \in \ker(\varphi_{\mathfrak{a}} \circ \alpha) \iff \varphi_{\mathfrak{a}}(\alpha(P)) = \infty \implies \alpha(P) \in E_2[\mathfrak{a}].$$

By definition, the condition $\alpha(P) \in E_2[\mathfrak{a}]$ means that for every element $a + b\pi \in \mathfrak{a}$, we have $(a + b\pi)(\alpha(P)) = \infty$, then since α is defined over $\mathbb{F}_p$, in particular it commutes with Frobenius by 2.7.8. Also since α is an isomorphism, in particular it is injective, so we get

$$\infty = (a + b\pi)(\alpha(P)) = \alpha(a(P) + b\pi(P)) \implies a(P) + b\pi(P) = \infty.$$

Since this holds for all $a + b\pi \in \mathfrak{a}$, we have shown that $P \in E_{\mathfrak{a}}$. The statement then follows from the uniqueness of the quotient up to isomorphism. □

Now we can give the full theorem in all of its glory.

Theorem 2.7.13. *Let $p \equiv 3 \pmod 4$ and $\mathcal{O}$ be an order in an imaginary quadratic field. Let $\pi \in \mathcal{O}$ and let $Ell_p(\mathcal{O}, \pi)$ denote the set of isomorphism classes of supersingular elliptic curves over $\mathbb{F}_p$ with*

endomorphism ring $\mathcal{O} \cong \mathbb{Z}[\pi]$. *Then the ideal class group* $\mathrm{cl}(\mathcal{O})$ *acts freely and transitively on the set* $Ell_p(\mathcal{O}, \pi)$ *via the map*

$$\begin{aligned} \mathrm{cl}(\mathcal{O}) \times Ell_p(\mathcal{O}, \pi) &\longrightarrow Ell_p(\mathcal{O}, \pi) \\ ([\mathfrak{a}], E) &\longmapsto E/\mathfrak{a}, \end{aligned}$$

where $\mathfrak{a}$ *is chosen as an integral representative.*

Proof. The group action satisfies the axioms of Definition 2.7.1 as shown in Lemma 2.7.9 and 2.7.10. We have also shown that this action is well-defined both for the isomorphism classes, and for the ideal classes in Lemma 2.7.11 and 2.7.12. Now there is only two claims left to show, that the group action is free and transitive.

Freeness follows from the fact that the ideal class group $\mathrm{cl}(\mathcal{O})$ is defined as the quotient of non-zero principal ideals, and we have seen that those are the exact ideals that correspond to endomorphisms. These principal ideals form the trivial class $[\mathcal{O}]$ in $\mathrm{cl}(\mathcal{O})$, and it acts like the neutral element. A group action is free if the only group element $[a] \in \mathrm{cl}(\mathcal{O})$ fixing any set element $E \in Ell_p(\mathcal{O}, \pi)$ is the neutral element.

Transitivity in this setting means that for any two isomorphism classes of curves $E_1, E_2 \in Ell_p(\mathcal{O}, \pi)$, there exists an ideal class $[a]$ such that $[a] * E_1 \cong E_2$. By Tate's isogeny theorem 2.5.20, if two curves have the same endomorphism algebra, then they are isogenous. In $Ell_p(\mathcal{O}, \pi)$, all curves are supersingular, and that means in turn that they have the same trace of Frobenius $t = 0$. This shows that between any two supersingular elliptic curve, there exists an $\mathbb{F}_p$-rational isogeny $\varphi : E_1 \to E_2$.

Define the subset of $\mathrm{End}_p(E_1)$ that factors through φ as

$$I_\varphi = \{\, \alpha \in \mathrm{End}_p(E_1) \mid \alpha(\ker \varphi) = 0 \,\}.$$

This subset forms a non-zero proper $\mathcal{O}$-ideal of $\mathcal{O}$. By construction, the kernel of this ideal $E_1[I_\varphi]$ is precisely the kernel of the isogeny φ. Evaluating the class action on E_1 via I_φ results in

$$[I_\varphi] * [E_1] = [E_1/E_1[I_\varphi]] \cong [E_1/\ker(\varphi)] \cong [E_2].$$

Since this ideal was defined for an arbitrary isogeny φ which is guaranteed by Tate's isogeny theorem, this shows that the action is transitive. For a nicer proof with kernel ideals and more algebraic geometry, see [76, Theorem 4.5] or [38, §3.3]. □

Corollary 2.7.14. *The number of isomorphism classes of supersingular elliptic curves over* $\mathbb{F}_p$ *with endomorphism ring* $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$ *satisfies*

$$|Ell_p(\mathcal{O}, \pi)| \approx \sqrt{p} \quad \text{as} \quad p \longrightarrow \infty.$$

Proof. Since the class group $\mathrm{cl}(\mathcal{O})$ acts freely and transitively on $Ell_p(\mathcal{O}, \pi)$, by Theorem 2.7.5 there is a bijection between these two sets (if they are finite, but the ideal class group is finite for all imaginary number fields [28, Theorem 5.22]). Having said that, by Theorem 2.4.39, we have that $h(\mathcal{O}_K) \approx \sqrt{|\Delta_K|}$ as the fundamental discriminant approaches negative infinity. For the order $\mathbb{Z}[\sqrt{-p}]$, the ring of integers is either $\mathbb{Z}[\sqrt{-p}]$ or $\mathbb{Z}[\frac{1+\sqrt{-p}}{2}]$ by Theorem 2.4.16, so Δ_K is either $-p$ or $-4p$. Then substituting in the values into the asymptotic formula gives the corollary statement. □

Let us go back to the isogeny graphs for a bit, because we have seen that we can also act on the vertices of isomorphism classes. There is an equivalent formulation to the isogeny graph that is for group actions.

Definition 2.7.15. Let G be a group acting freely on a set X, and let $S \subset G$ be a symmetric subset. Then the **Schreier graph** of (S, X) is the graph whose vertex set is the elements of X, and edges are drawn between vertices $x, y \in X$ if there is an element $s \in S$ such that $s * x = y$.

As per the discussion preceding this definition, we have already seen a lot of Schreier graphs in the form of isogeny graphs. We have also seen Figure 6, which was an example of a Cayley graph, but

multiplying group elements by other elements can be viewed as a group acting on itself by left (or right) multiplication. The relevant graph theory for Schreier graphs in our case is that under certain conditions, these graphs are expander graphs [44]. This was already mentioned in Theorem 2.6.11. Unfortunately, this theorem does not apply to CSIDH, because as stated in the theorem, one would need to have much more isogenies that have norm at least $\log(4q)^2$ to become an expander graph.

We finish this section, and the whole chapter, with the theorem of the CSIDH team that allowed the (re)birth of this protocol.

Theorem 2.7.16. *[17, Proposition 8] Let $p > 5$ be a prime such that $p \equiv 3 \pmod 8$, and let $E(\mathbb{F}_p)$ be a supersingular elliptic curve. Then $\mathrm{End}_p(E) = Z[\pi]$ if and only if there exists $A \in \mathbb{F}_p$ such that E is $\mathbb{F}_p$-isomorphic to the curve $E_A : y^2 = x^3 + Ax^2 + x$. Moreover, if such an A exists then it is unique.*

Proof. First assume that E is isomorphic over $\mathbb{F}_p$ to E_A for some $A \in \mathbb{F}_p$. We claim that E_A has no full 2-torsion part. To arrive at a contradiction, assume that E_A does have the full 2-torsion $E_A[2]$. Since it is in Montgomery form, the point $(0, 0)$ is always a 2-torsion point. Furthermore, every 2-torsion point has the form $(x, 0)$ for some $x \in \overline{\mathbb{F}}_p$. We can see that we have full 2-torsion if the polynomial $x^2 + Ax + 1$ splits in $\mathbb{F}_p$. This can be determined by the Legendre symbol $\left(\frac{\Delta}{p}\right)$, where $\Delta = A^2 - 4 = (A-2)(A+2)$ is the discriminant of the previous polynomial. This expression is a square if and only if either both or neither $(A-2)$ and $(A+2)$ are square. Table 2.1 shows a table of 4-torsion points of Montgomery curves divided into the exact condition. Rows 1 and 3 are the relevant here, and they show that either the curve

B	$A-2$ and $A+2$	$E_A(\mathbb{F}_q)[4]$	$E_A^t(\mathbb{F}_q)[4]$
square	both square	$\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$
square	one square	$\mathbb{Z}/4\mathbb{Z}$	$\mathbb{Z}/4\mathbb{Z}$
square	neither square	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$

Table 2.1: 4-torsion structures on a Montgomery curve E_A and its quadratic twist E_A^t. Excerpt [26, Table 1]

E_A or its twist contains $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Due to Lagrange's theorem, 8 must divide the order of the group $E(\mathbb{F}_p)$, but since we are in the case $p \equiv 3 \pmod 8$, both the curve and its twist has order $p + 1 \equiv 4 \pmod 8$, therefore it cannot have a full 2-torsion subgroup $E_A[2]$. Then by Theorem 2.6.23, since E_A does not have all the order-2 points, it must be on the floor of the isogeny graph. As a result, we have that $\mathrm{End}_p(E_A) \cong \mathrm{End}_p(E) = \mathbb{Z}[\pi]$.

Now assume that $\mathrm{End}_p(E) = \mathbb{Z}[\pi]$. By Theorem 2.7.13, the class group $\mathrm{cl}(\mathbb{Z}[\pi])$ acts transitively on $Ell_p(\mathbb{Z}[\pi], \pi)$, so there exists a proper $\mathcal{O}$-ideal $[\mathfrak{a}]$ such that $[\mathfrak{a}] * E_0 = E_A$. It is always possible to choose a representative of the ideal class to be coprime to $2p$, so without loss of generality assume so. We then get the separable $\mathbb{F}_p$ isogeny $\varphi_{\mathfrak{a}} : E_0 \to E$ which has odd degree. Then by Theorem 2.3.5, there exists a separable isogeny $\psi : E_0 \to E_A : y^2 = x^3 + Ax^2 + x$ defined over $\mathbb{F}_p$ such that $\ker \psi = \ker \varphi_{\mathfrak{a}}$. By Theorem 2.3.1, isogenies defined over $\mathbb{F}_p$ are unique up to $\mathbb{F}_p$-isomorphism with a given kernel, therefore E is indeed $\mathbb{F}_p$-isomorphic to E_A.

The last thing remaining is showing uniqueness. Let $B \in \mathbb{F}_p$ be another Montgomery coefficient such that $E_A \cong E_B : Y^2 = X^3 + BX^2 + X$. Then by [69, Proposition III.3.1(b)] there exist $u \in \mathbb{F}_p^\times$, and $r, s, t \in \mathbb{F}_p$ such that

$$x = u^2X + r, \quad \text{and} \quad y = u^3Y + su^2X + t.$$

If one substitutes these into the original equation for E_A and collects like terms, the following relations are obtained:

$$s = t = 0, \quad 3r^2 + 2Ar + 1 = u^4, \quad 3r + A = Bu^2 \quad \text{and} \quad r^3 + Ar^2 + r = 0.$$

The only $r \in \mathbb{F}_p$ for which $r^3 + Ar^2 + r = 0$ is $r = 0$, because as shown before, E_A has a single $\mathbb{F}_p$-rational 2-torsion point. In this case the only surviving relations are $u^4 = 1$ and $A = Bu^2$. The only possibilities for $u^4 = 1$ are ± 1, but it's squared in the other relation, so it does not make a difference, hence $A = B$ as claimed. □

The fact that every representative curve in the supersingular isomorphism classes with endomorphism rings $\mathbb{Z}[\pi]$ can be uniquely represented by $A \in \mathbb{F}_p$ is a very good thing in cryptography, because the public key that needs to be sent can be extremely tiny compared to other methods. In the next chapter, we explore the thought process that could have made you (or me!) discover this protocol.

Chapter 3

The Design and Details of CSIDH

Now that we have developed the necessarily mathematical tools, we are well-equipped to understand the ingenuity of this non-interactive key exchange protocol. Let us first discuss some relevant terminology within a cryptographic context before diving into the inner workings of CSIDH.

There are two major collections of algorithms that can be used for key exchanges.

Definition 3.0.1. [10, Definition 1] A **Non-Interactive Key Exchange (NIKE)** is a collection of two algorithms, `KeyGen` and `SharedKey`, where

- `KeyGen` is a probabilistic algorithm that on input 1^k, where k is a security parameter, outputs a keypair (sk, pk); and
- `SharedKey` is a deterministic algorithm that takes (sk, pk) as an input, and outputs a shared key K.

We say that a NIKE is **correct** if for any two generated pair (sk_1, pk_1) and (sk_2, pk_2), it holds that $\texttt{SharedKey}(sk_1, pk_2) = \texttt{SharedKey}(sk_2, pk_1)$.

Definition 3.0.2. [10, Definition 2] A **Key-Encapsulation Mechanism (KEM)** is a collection of three algorithms, `KeyGen`, `Encaps` and `Decaps`, where

- `KeyGen` is a probabilistic algorithm that on input 1^k, where k is a security parameter, outputs a keypair (sk, pk); and
- `Encaps` is a probabilistic algorithm that takes as input a public key pk, and outputs a pair of ciphertext and shared key (ct, K)
- `Decaps` is a deterministic algorithm that on input (ct, sk) outputs the shared key K.

We say that a KEM is **correct** if for any $(sk, pk) \leftarrow \texttt{KeyGen}(1^k)$ and $(ct, K) \leftarrow \texttt{Encaps}(pk)$ it holds that $\texttt{Decaps}(ct, sk) = K$.

These are the abstract definitions, that serve the purpose of categorizing different cryptographic primitives. Of course, for a NIKE-style protocol, to generate the shared key K mentioned in the definition, the protocol must ensure that the two communicating parties have a way of sharing each other's public keys.

To connect back to the earlier Definition 1.2.1, KEMs are clearly asymmetric, but NIKE's case could be a bit more confusing. The most known NIKE is ECDH, which is the elliptic curve version of the Diffie-Hellman protocol (see Section 3.1.1), which *produces* symmetric keys that can be used later by other algorithms, but it itself is an asymmetric protocol, since both parties have their own private keys. The non-interactive nature of NIKEs makes them more flexible than a KEM, but it could depend on the scenario as well. In some cases, both KEMs or NIKEs can be used, such as in the ephemeral key exchange of TLS 1.3. That protocol has been using mostly NIKE, but because of the threats of quantum attacks discussed in Section 4.2, they were considering migrating over to post-quantum KEMs, since there were no comparable post-quantum NIKE options. CSIDH is the potential answer to this hole.

Another important aspect of comparing KEMs to NIKEs is their capability for Static-Static Key Exchange (SSKE). Static-static key exchange is a non-interactive key derivation protocol, where the two parties can compute a shared secret key using only their pre-established, long-term private and public keys, hence the term "static". There are no ephemeral (short-lived) key material during the SSKE protocol. The obvious advantage of such a protocol is the elimination of online key-exchange payloads entirely. This is crucial for bandwidth-constrained networks, like IoT environments (for example IETF

LAKE / EDHOC). They also support the possibility of Zero Round-Trip Time (0-RTT), which essentially means that there is no requirement for an initial interactive handshake (like in TLS) or a response payload to secure transmission.

In a NIKE scheme, two parties holding static key pairs can compute a common shared secret independently and without interacting or exchanging ephemeral ciphertexts. However, Standard KEMs cannot natively support static-static Diffie-Hellman key exchange due to their one-way encapsulation design. When KEMs are adapted for mutual authentication, the extra encapsulation cipher texts bloat the message bandwidth consuption. The conflicting problem is that there were no viable post-quantum NIKE options, which is what CSIDH is trying to fill in.

Even though CSIDH was late for the NIST standardization (see Section 4.2.1), the IETF LAKE Working Group has formally requested that the Crypto Forum Research Group evaluates candidates for future lightweight PQC NIKE standardization. The CSIDH framework natively supports NIKE and SSIKE, so it would be positioned as a prominent candidate (provided the quantum security is not degraded further, see Subsection 4.2.5)

Now we that we have given a bit of context, let us explore how this protocol came to be.

3.1 CSIDH – Letter By Letter

This section is going to give a less straining introduction to how CSIDH came to be, and what the history is broadly. Before CSIDH, there have not been many options for an efficient isogeny-based cryptographic protocol, as previous iterations, such as the Couveignes-Rostovtsev-Stolbunov [27, 61] protocol, was too slow for any real life practical purposes. We explore the intuition behind key exchange protocols, and give a detailed review of how the failures of the past shaped the future.

3.1.1 DH: Diffie-Hellman

CSIDH is a quantum-secure iteration of an idea for a non-interactive key exchange protocol, initially developed in 1976 by Whitfield Diffie and Martin Hellman[32]. As its first rendition, it was defined as a protocol to be able to securely share a secret key over an insecure channel that can be used later in a symmetric key cipher. This protocol used the commutativity of exponents of group elements, together with the fact that it's a very hard problem to reverse-engineer the exponent used (this is also referred to as the **discrete logarithm problem**, abbreviated as DLP). The key exchange goes like this:

We have Alice and Bob, who would like to agree on a common shared key to use to encrypt messages later. They have previously publicly chosen a cyclic group with a large prime p order, and a generator g. Then Alice chooses a secret exponent a and computes $A = g^a \bmod p$. Similarly, Bob chooses a secret exponent b and computes $B = g^b \bmod p$. Next, both Alice and Bob exchange their computed values via their chosen channel, which is being eavesdropped by Eve. Eve therefore gains knowledge of A and B, but has no knowledge of the secret exponents a and b, since they are protected by DLP. In order to arrive at the shared secret, Alice raises Bob's value to her own secret exponent again, getting $S = B^a$, and Bob computes likewise A^b. Note that this is the same secret value, since group exponentiation is commutative:

$$A^b \equiv (g^a)^b \bmod p \equiv g^{ab} \bmod p \equiv g^{ba} \bmod p \equiv (g^b)^a \bmod p \equiv B^a = S.$$

Eve, on the other hand, has no way of calculating $S = g^{ab} \bmod p$ by only knowing $A = g^a \bmod p$ and $B = g^b \bmod p$.

In the context of elliptic curves, the equivalent protocol is called ECDH, which stands for Elliptic Curve Diffie-Hellman. As discussed in Section 2.1, The set of points of an elliptic curves is an abelian group, therefore the same idea can be applied to elliptic curves, in which case, if we keep the standard notation of a and b for Alice and Bob respectively, the secret key becomes

$$S = (ab)G = (ba)G,$$

where $G = (x_0, y_0)$ is a publicly agreed generator with prime order $d \sim |E(\mathbb{F}_p)|$. ECDH is currently used

widely in cryptographic environments today, most notably in TLS, Signal and Bitcoin. As marvelous an invention it was, unfortunately it will have to be superseded by a quantum-secure version, since a curve with a 256-bit modulus can be broken with only 2330 qubits, and 126 billion ($\sim 2^{37}$) Toffoli gates [60].

This is where CSIDH comes into the picture. The perspective is shifted from doing computations within a single elliptic curve, where the order of operation does not matter, to having a fixed set of elliptic curves and jumping between them with the help of isogenies. This change of perspective also has the consequence that the same quantum algorithm that breaks ECDH does not apply to CSIDH. The reason for this is discussed in Section 4.2.3, and the more relevant quantum algorithm for attacking CSIDH is discussed in Section 4.2.5.

3.1.2 I: Isogeny-based

As mentioned before, isogenies are the connections between two elliptic curves. If we choose a point in one elliptic curve, apply an isogeny to it, then we end up on another curve, and thanks to the existence of dual isogenies, we are able to travel back to the starting curve. Isogenies themselves are defined by rational maps, that is certain substitutions of the points' coordinates from $P = (x, y)$ to $P' = (x', y')$. These maps, however, can become very complicated very fast:

Example 3.1.1. A degree 5 isogeny from the curve $E_1 : y^2 = x^3 + x$ to $E_2 : y^2 = x^3 + 111x^2 + x$, defined over the finite field $\mathbb{F}_{659}$ is given by

$$(x, y) \mapsto \left(\frac{295x^5 + 57x^4 + 73x^3 + 322x^2 + x}{x^4 + 322x^3 + 73x^2 + 57x + 295}, y \cdot \frac{265x^6 + 149x^5 + 97x^4 + 298x^3 + 86x^2 + 170x - 295}{x^6 - 176x^5 + 110x^4 + 5x^3 - 177x^2 + 141x - 265} \right).$$

As you can see, we are merely dealing with a degree 5 isogeny, whereas in the CSIDH protocol, we get up to degree 587![1] But thankfully, we don't actually need to know the exact mapping function, since we only care about the elliptic curve that we end up on, whose coefficients can also be calculated with the help of Vélu's formula (or by another formula in the case of CSIDH, cf. Section 2.3).

Unfortunately, if we take any two random elliptic curves without any restrictions, they very rarely have isogenies connecting them. In the rare case of when they do, the Isogeny Graphs are one way of visualizing these isogeny connections between a set of elliptic curves. See Definition **??** for a rigorous definition, but in essence, the isogeny graph is a graph where the nodes are isomorphism classes of elliptic curves (or represented by their j-invariants), and the edges between them are different isogenies. A priori these graphs are directed, but by Definition 2.2.20, there exists an "inverse" isogeny, so we can view these graphs as undirected (modulo some minor lie, cf. Remark 2.6.13). They are also disconnected, having connected components that range from singular vertices to structures that look like volcanoes. See Figure 8 for an example. These graphs are still considered somewhat manageable, since there is a clear structure to the edges. On the other hand, in the case of the full supersingular isogeny graph over the algebraic closure $\overline{\mathbb{F}}_p$, the graph resembles more of a ball of yarn than a volcano. (Figure 9). Isogenies and the isogeny graph as a basis for cryptographic purposes was already proposed in the mid 2000s[27, 61, 19]. Even though they were well-founded as mathematical theories, there were many obstacles in implementing them as an efficient algorithm. However, with the correct manipulations of choosing the primes, elliptic curves, endomorphisms and which field the graph is defined over, they can be tamed, which is exactly what the authors of CSIDH have managed. The mathematics of such graphs and how they can be manipulated is explored extensively in Section 2.6.

These graphs provide a natural way of visualizing the Diffie-Hellman protocol applied to elliptic curves and isogenies, which parallels the way of visualizing the Diffie-Hellman protocol on groups and exponents. To illustrate how CSIDH uses its isogeny graph, let us take a toy model with a small prime and three available isogenies. If we are able to find a graph that is connected, cyclic and commutative (more about this later in Subsection 3.1.4!). Once we have such a graph, then the Diffie-Hellman protocol on an isogeny graph is illustrated in Figure 10, and the figure is explained in the next paragraph.

The nodes of the graph represent the Montgomery coefficient A of the curve $E_A : y^2 = x^3 + Ax^2 + x$. The common starting point is E_0, which is the top vertex in the graph. We now let Alice and Bob choose

[1] This is not a factorial!

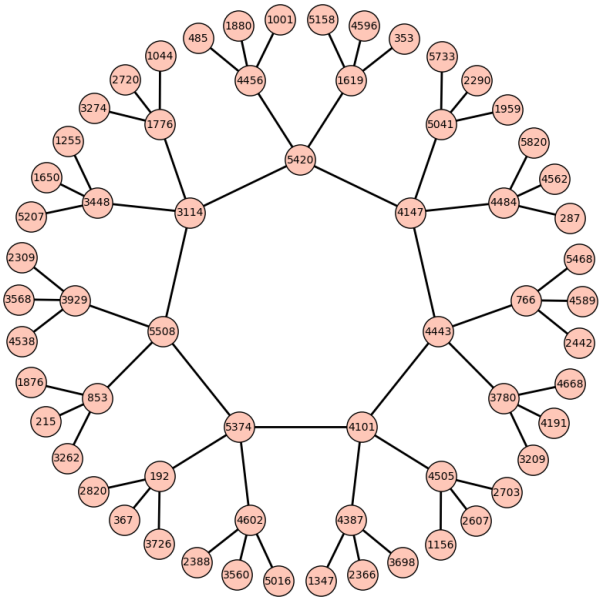


Figure 8: The connected component of $G_{\overline{\mathbb{F}}_p,\ell}$ for $p = 5851, \ell = 3$. The nodes are represented by the j-invariants of the isomorphism classes, and the edges are the 3-isogenies between them. Figure created with SageMath [39].

a set of secret isogenies, which correspond to the edges in the isogeny graph. Their respective secret steps are shown color-coded at the bottom, where the plus sign means taking a step counter-clockwise. They both take their respective steps in the graph by following the edges of their secret isogenies, both starting from curve E_0. These paths are denoted by the dotted lines, and their target curves are colored yellow, and labelled E_A for Alice and E_B for Bob. Next, they exchange their respective nodes that they ended up on, and apply the same secret isogenies starting from each other's curves. These are the solid lines done by Bob on the left graph and by Alice on the right graph. By our assumptions, they arrive at the same vertex of the graph, which is now their shared secret, since $E_{AB} = E_{BA}$.

The classical and quantum security of this approach relies on the presupposed hardness of the following problem.

Problem 3.1.2 (Isogeny Problem)**.** Given two isogenous elliptic curves E_1, E_2, find an isogeny $\varphi : E_1 \to E_2$ between them.

Just like with everything in post-quantum cryptography, the hardness of this problem is also based on the fact that nobody has been able to find a way to solve this in polynomial time by a classical algorithm. As it will be discussed later in Section 4.2, quantum computers do pose a bigger risk to this problem, as there exists an algorithm that finds an isogeny in subexponential time.

Beyond the hardness of this problem, we also need to consider several other aspects regarding the isogeny graph. In a previous paragraph, I mentioned that we need to tailor the primes and fields in a way such that the isogeny graph itself is cyclic. This is a must if we ever wish to arrive at the same shared secret elliptic curve, because if we were to take walks in the isogeny graphs of Figure 8 or Figure 9, not only do we have more than 2 choices for choosing the next vertex, but taking steps on the floor and then on the surface (if they even exist) would give a different vertex than walking on the surface first, then on the floor.

Furthermore, for this approach to have the same level of security as the original version of Diffie-Hellman, we also require the property that random walks on the graph mix rapidly, and that the possible public keys (Alice's and Bob's target curve) are uniformly distributed. Rapid mixing refers to the fact that the number of random steps within the graph needed to end up at any vertex with equal probability is relatively much smaller than the full size of the graph. This is essential in order to be able to implement any algorithm, since the target vertex of the encryption needs to be "hard to guess", but if it takes too long to get to that state, the algorithm has no practical use. Thankfully, this property is satisfied for the

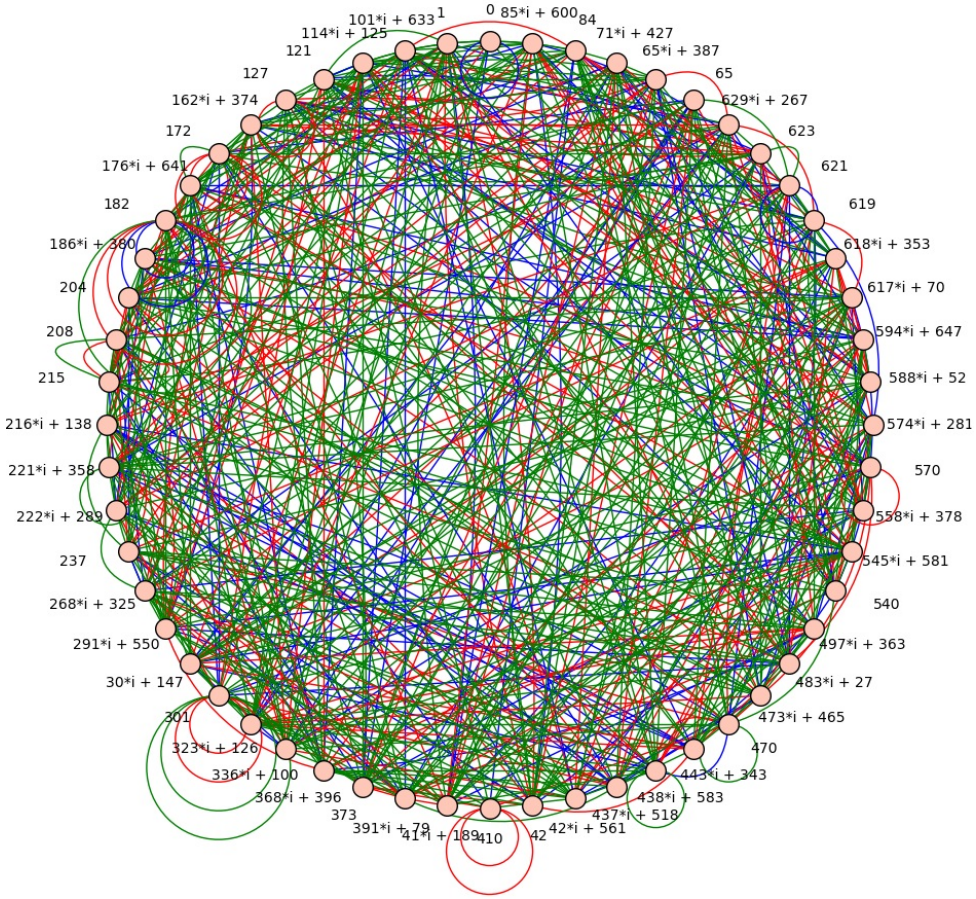


Figure 9: The full supersingular isogeny graph $G_{\overline{\mathbb{F}}_p,L}$ for $p = 659, L = \{3, 5, 11\}$. The vertices are labeled by the j-invariants, and i is the generator of $\mathbb{F}_{p^2}$. Figure created with SageMath [39].

above mentioned isogeny graphs, which are also discussed in Section 2.6.

3.1.3 S: Supersingular

Elliptic curves are by definition nonsingular, so this term is a bit unfortunate, but it originates from a time when people were exploring elliptic curves over the complex number $\mathbb{C}$. Mathematicians noticed that some curves have extra endomorphisms, and they described them as "rare", or as they said it back then, "singular". Later on, when the focus shifted towards more restricted domains, such as finite fields, a new type of strange elliptic curve popped up, hence the name "supersingular". There are quite a lot of reasons for though why one would consider supersingular graphs when looking to implement some kind of cryptographic primitive. A potentially attractive feature of supersingular curves when it comes to calculating multiples of a point is that sometimes it can be done faster than expected [75, pp. 132]. If we want to compute kP, where $P = (x, y) \in E(\mathbb{F}_{p^n})$ for some $n > 0$, then we usually do that by successive doubling, because doubling a point is more efficient for large k than adding it k times. In practice, this means that the integer k is expanded in base 2 as

$$k = k_0 + 2k_1 + 2^2k_2 + \cdots + 2^r k_r.$$

This allows us to do $r = \log k$ doublings, and at most r additions instead of k additions. In the supersingular case, it is possible to do even better. Recall that if E is supersingular then the trace of Frobenius is 0 (Theorem 2.5.9), and so the Frobenius endomorphism π satisfies the equation $\pi^2 + p = 0$ by Proposition 2.5.6. Therefore the multiplication-by-p map can be expressed as

$$[p]P = p(x, y) = -\pi^2(x, y) = \left(x^{p^2}, -y^{p^2}\right).$$

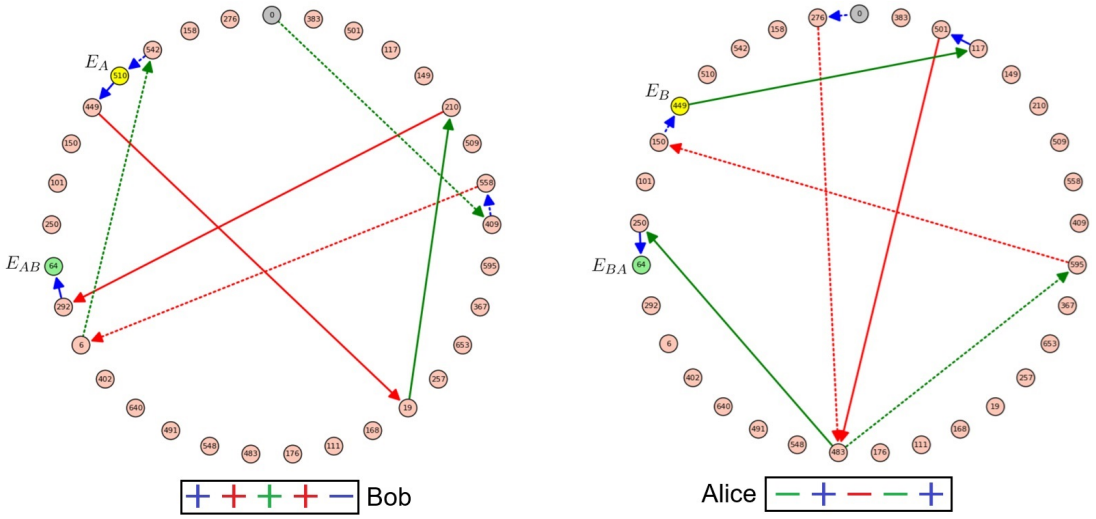


Figure 10: Isogeny graph key exchange protocol. Bob's and Alice's secret path are denoted below the graphs, where the colors corresponding to the color of the edge, and the plus-sign means taking a step counter-clockwise. The dotted lines show the first paths by the other person, with their target vertex denoted by yellow g_A and g_B for Alice and Bob respectively. The solid lines are the paths followed by the person written below the graph. The green vertex is the shared secret at the end of the protocol.

This is faster to do because we are trading arithmetic on the elliptic curve for arithmetic in a finite field. This can be even further sped up if the coordinates are expressed in terms of a normal basis (see [75, Appendix C]). So the faster procedure can be defined as

1. Expand k in base p:
$$k = k_0 + pk_1 + p^2k_2 + \cdots + p^rk_r, \quad 0 \leq k_i < p.$$
2. Compute k_iP for each i with the double-and-add method described above.
3. Compute p^ik_iP as $\left(x_i^{p^{2i}}, -y_i^{p^{2i}}\right)$, using the faster finite field arithmetic.
4. Sum the points p^ik_iP for $0 \leq i \leq r$.

The downside of this is that it requires the finite fields' characteristic to not be too big, otherwise the calculations will not be in the zone of efficiency. When it comes to cryptographic algorithms, performance through optimized arithmetic is a common practice, but there are other good reasons to consider.

When using isogeny graphs as basis for key exchange protocol, ideally one would like to have graphs that are k-regular, that is every vertex has exactly k neighbors, and it is also a desirable trait to have fast mixing times, as mentioned in the previous section. There is a theorem which states that if we look at the full supersingular graph (which is the isogeny yarnball that can be seen in Figure 9), then it has the Ramanujan property. This property means that one can take random walks in the graph that mix almost as fast as mathematically possible. For a cryptographic scheme, it is a crucial part of its design to be able to hide any information, which in this case would be the distribution of the public keys generated. Furthermore, the supersingular graphs are easier to control than the ordinary components, for the simple reason that when they are defined over the finite field $\mathbb{F}_p$, they *all* have exactly $p + 1$ points. Due to Hasse's theorem 2.5.7, the possible number of rational points on a curve is bound by $p + 1 \pm 2\sqrt{p}$, but there is only one number for which the curve is supersingular. This gives a lot better predictability when looking to use the intrinsic mathematical nature of the graphs.

Another consequence of the fact that the curve is supersingular is that it has the largest possible discriminant. The discriminant is connected to the size of the isogeny graph through Siegel's approximation 2.4.39. It states that the number of isomorphism classes grow with $\sqrt{p}$, when they are defined over $\mathbb{F}_p$, so clearly supersingular curves are the best choice in that regard. When one considers the full supersingular graph, there are about $p - 1/12$, which scales even better. Another benefit of supersingular curves is that

they are all defined over $\mathbb{F}_{p^2}$, so there is no need to hunt for huge field extensions in the hopes of finding a rational curve.

Theorem 2.5.18 gives a description of all possible endomorphism rings for any curve, ordinary or supersingular, but it is only in the supersingular case when the endomorphism ring can be non-commutative. when that property trickles over to isogeny graphs, then we can view it in two opposing way: (1) It's a bad thing that we lost commutativity; or (2) It's a *good* thing that we lost commutativity! This is what Luca De Feo, David Jao and Jérôme Plût took advantage of in 2011 [30] to create an efficient and interesting key exchange protocol, which they called **SIDH**, which stands for "Supersingular Isogeny Diffie-Hellman". It is just shy of one more letter...

As mentioned briefly in the introductory section 1.3, the mechanics and assumptions of SIDH are quite different from that of CSIDH, but what they have in common is obviously the supersingularity. That's about it, because the actual protocol can be described as follows:

The SIDH elliptic curves are defined over $\mathbb{F}_{p^2}$, where the prime is manufactured to be of the form $p = l_A^{e_A} l_B^{e_B} \cdot f \pm 1$. This demonstrates the deliberate manipulation of the prime field to suit one's needs, which is a courtesy of supersingular curves. In the original SIDH paper [30], they instantiate the smallest primes $(l_A, l_B) = (2, 3)$, which helps with efficient isogeny calculations. The other remaining numbers are then adjusted to have roughly the same security level for both parties. From Theorem 2.2.23, we know that for primes not dividing the characteristic of the field, the ℓ-torsion group splits as $E[\ell] = \mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}$. A similar decomposition happens when one looks at the whole group as well.

Since the endomorphism algebra is noncommutative, doing a Diffie-Hellman style key exchange is a fruitless endeavor. If Alice's public key is A, Bob's is B, then in the full supersingular case, we have that $AB \neq BA$, so there is no shared secret like this. The solution to this was to have two isogeny graphs for the two chosen primes $(2, 3)$. The drawback (and the eventual downfall) with this approach is that after evaluating the isogeny steps, both parties need to include their kernel images in the public key. The reason for that is if they are taking isogeny steps on two different isogeny trees, their kernels will be disjoint. When that happens, combining them into one big kernel has the latent effect of being commutative.

SIDH had a good run, but many researchers were skeptical of publishing kernels of the secret isogenies, as it seemed like a great way to leak information. SIDH joined the call for post-quantum cryptographic standardization that was initiated by the national Institute of Standards and Technology (NIST for short). They changed name to SIKE, which perfectly reflects what happened in 2022 – Castryck and Decru [15] published a paper, where they broke the SIDH protocol in mere hours on a classical computer. Further papers [48, 58] have cemented the fate of SIDH in the Post-Quantum race, but it clearly had upsides as well in the form of new mathematical research, and better protocols.

The classical break of SIDH indeed used the information gathered from the published kernels, but it was also a different type of approach. It was based on the fact that finding the endomorphism ring of a given supersingular curve defined over $\mathbb{F}_{p^2}$ is a very hard problem. The ingenuity of the attack was to transfer the problem to higher dimension. There are simply more possibilities for endomorphisms when one considers the product of elliptic curves $E \times E$ instead of just one. Kani's theorem ensures that in 2 dimensions, any-degree isogenies are constructible in a simple way. So perhaps commutative is the way to go...

3.1.4 C: Commutative

We have seen an example of a key exchange protocol that was based on supersingular curves, but was also non-commutative, so torsion points had to be published in the public keys, which led to a fatal attack. This started in 2011, but already in 1996, Couveignes [27] has thought about using group actions and graphs together for encryption. He submitted his paper to a the Crypto 1997 conference, where his paper introducing hard homogeneous spaces and isogeny-based cryptography got rejected. He forgot about it, until in 2006, Rostovtsev and Stolbunov [61] independently rediscovered what Couveignes had try to forget.

They were the first to suggest basing key exchange on hard homogeneous spaces, which are Torsors over a group action, where certain problems are easy, and others are hard. They perfectly map onto

Schreier graphs (Definition 2.7.15), which are graphs that draw edges between vertices of a set, if the two elements are connected by some group action. We have already seen an example of the proposed mechanics of the CRS key exchange protocol in Figure 10. The steps of the key exchange are explained in the caption of the figure, but we can see a quick explanation of why it works.

Alice chooses a random walk in the Schreier graph, then so does Bob. They share with each other the vertex where they finished their walk, that is their public key. After they both verify that the public key is correct, they perform their steps again, and end up at the same place. This is commutativity at work, which follows from the fact that the $\mathrm{cl}(\mathcal{O})$ is itself commutative, and this flows over with the group action.

This method is in some sense the opposite of SIDH, since these isogeny steps are commutative, and they use ordinary curves over $\mathbb{F}_q$. The interesting thing is that they never explicitly compute $\mathcal{O}$, the endomorphism ring, instead they determine parameters in the following order [29].

1. A large enough finite field $\mathbb{F}_q$,
2. A curve E defined over $\mathbb{F}_q$,
3. With point counting, determine the Frobenius discriminant $D_\pi = t_\pi^2 - 4q$, and then verify that it has a large enough prime factor,
4. Determine a set $L = \{ \ell_1, \ldots, \ell_m \}$ of primes that split in $\mathbb{Z}[\pi]$,
5. For each prime ℓ_i, factorize the Frobenius polynomial

$$\pi^2 - t_\pi \pi + q = (\pi - \lambda_i)(\pi - \mu_i) \pmod{\ell_i}.$$

Looking back at this setup from 2026, certainly feels like a lot of things are unoptimized, but of course one needs to keep in mind that at that time, this was the pioneering concept. One of the biggest bottlenecks when trying to implement this protocol was the immense amount of computational resources needed. As mentioned before, supersingular curves are all defined over at most $\mathbb{F}_{p^2}$, but there isn't such a cap for ordinary curves. That is why it could happen that if the $E[\ell_i]$ torsion points are not defined over $\mathbb{F}_q$, then the computation must be extended over a potentially high-degree field $\mathbb{F}_{q^k}$. The splitting behavior of primes is also hard to predict in ordinary curves, which also added a lot of computational overhead for step 4. The algorithm essentially finds $\mathbb{F}_q$ first, then looks for elliptic curves that split for as many primes as possible. In the end, it was only practical to enforce it for 7 primes.

Commutativity is both a blessing and curse, since quantum attacks have a higher chance of being applicable if the setting is commutative in some way. This is discussed later in 4.2. On the other hand, there are many good things to take away from both of these protocols, SIDH and CRS. The supersingular case has the advantage of being able to better control the sizes of subgroups, and having the upper limit of the prime field $\mathbb{F}_{p^2}$ as worst case scenario. The ordinary case could have very fast isogeny steps in a more connected graph, then the overall cost of evaluating the isogenies could be reduced as well.

3.1.5 Which prime to choose?

Let us summarize what we have discovered so far, and what we would like. We want to have a key exchange protocol, that is similar to how Diffie-Hellman works. CRS has showed us that Schreier graphs can be a good primitive to work with. SIDH's full supersingular graph had very good mixing properties, and it had good potential. The problem is with its non-commutativity, which was solved by publishing torsion points. What we have not tried yet is trying something *commutative*, like the action of the ideal class group, in a *supersingular* setting, because then perhaps we can control better what we want.

First of all, when considering what prime our field should be, if we look at the trouble $\mathbb{F}_{p^2}$ made with the isogeny graph (Figure 9) then it is perhaps worth trying $\mathbb{F}_p$. From the CRS protocol, we know that the best would be to have as many small primes splitting in our endomorphism ring $\mathcal{O}$, because than we have a lot more choices to take steps in. It would also be beneficial to have these primes be as small as possible, since the isogeny evaluation scales as $O(\ell)$ (or $O(\sqrt{\ell})$ [3]) with the size of the ℓ-isogeny.

Let us also recall some important theorems from the previous chapter. As per Theorem 2.6.23, there are only 2 levels to the isogeny graph when the curve is supersingular, and there are 2 horizontal isogenies when $\left(\frac{-p}{\ell}\right) = 1$. This is an advantage compared to the ordinary isogeny graph, which has one crater, and

then a tree-like structure (see Figure 8). Theorem 2.6.15 proves that if the starting curve is supersingular, then the isogenous curve is supersingular as well. Theorem 2.5.13 shows that the curve $E : y^2 = x^3 + x$ is supersingular if $p \equiv 3 \pmod 4$. From Theorem 2.7.16, we see that if $p \equiv 3 \pmod 8$, then there is a unique and efficient way of representing the supersingular curves, which would put CSIDH at the forefront of tiny key exchanges. Remark 2.6.14 highlights that for the isogeny graph to be undirected, we need to have $p \equiv 11 \pmod{12}$. And of course, we would like to have $\ell_i \mid \#E(\mathbb{F}_p)$ for as many primes ℓ_i as possible.

The CSIDH team has chosen the prime to be in the form of

$$p = 4 \cdot \ell_1 \cdot \ell_2 \cdots \ell_n - 1,$$

where the ℓ_i denote the list of primes starting from 3. Let us see if this form achieves everything that we wanted. Since we decided to work in the supersingular setting, we know that $\#E(\mathbb{F}_p) = p + 1$. Therefore it is clear that all the small primes ℓ_i divide the group order, therefore they exist as orders of elements. This allows us to sample points of order ℓ_i for ℓ_i-isogenies. Let us investigate how the isogeny graph looks like with this prime p. We indeed have $p \equiv 3 \pmod 4$, so with this prime, we have that $E_0 : y^2 = x^3 + x$ is in our isogeny graph, and we know that we are on the floor as well. We also have now that the endomorphism ring of a supersingular curve in this field is $\mathbb{Z}[\pi] = \mathbb{Z}[\sqrt{-p}]$, we thus can check what happens to our primes ℓ_i in this order. Since we are in the supersingular setting, the trace of Frobenius is 0 (Theorem 2.5.9), so the minimal polynomial for π is $x^2 + p$. Taking module ℓ_i gives

$$x^2 + p \equiv x^2 - 1 = (x + 1)(x - 1) \pmod{\ell_i}$$

for all ℓ_i. This means that $\ell_i\mathcal{O}$ splits as the ideals $[\mathfrak{l}_i][\overline{\mathfrak{l}_i}]$ with $\mathfrak{l}_i = (\ell_i, \pi - 1)$. Since all the ℓ_i small primes divide the group order, and also split into two prime ideals, we have a $2n$-regular isogeny graph with a well-defined orientation. Furthermore, acting by the ideal class $[\mathfrak{l}_i]$ is by Definition 2.7.6

$$\begin{aligned} E[\mathfrak{l}_i] &= \{\, P \in E(\overline{\mathbb{F}}_p) \mid [\ell_i]P = \infty, (\pi - 1)P = \infty \,\} \\ &= \{\, P \in E(\overline{\mathbb{F}}_p) \mid [\ell_i]P = \infty, \pi(P) = P \,\} \\ &= \{\, P \in E(\mathbb{F}_p) \mid [\ell_i]P = \infty \,\} \\ &= E(\mathbb{F}_p)[\ell_i], \end{aligned}$$

thus we see that the ideal class group action evaluation would be equivalent to sampling a point with order ℓ_i, and using that point to generate a kernel for the isogeny.

Since we are on a Schreier graph, in order to achieve the heuristic bound for graph to be an expander, we would need to have all the primes with smaller norm than $\log(4p)^2$ to be in the symmetric set. At cryptographically relevant sizes, that is not feasible, so there should be another heuristic for deciding how many small primes we would need to have in the CSIDH prime p. One such heuristic could be the bounding box argument.

If we set a bound for the number of $[\mathfrak{l}_i]$ actions to be performed between $\{\, -m, \ldots, m \,\}$, then get that the potentially reachable size of the key space is $(2m + 1)^n$. Based on existing attacks, the CSIDH team estimated that the CSIDH prime should be about 2^{512}, with which we have an estimate on the number of $\mathbb{F}_p$-rational isomorphism classes as $\sim 2^{256}$ by Theorem 2.4.39. With this in mind, the parameters chosen for n, m turned out to be $n = 74$ and $m = 5$, which together gives $11^{74} \approx 2^{256}$ possible secret keys. Let us see now how the actual protocol works.

3.2 Non-Interactive Key Exchange Protocol

As per Definition 3.0.1, a NIKE protocol needs to have (at least) two components: `KeyGen` and `SharedKey`. Other algorithms include key verifications to protect against adversaries. In this section, we present the full key exchange protocol for **CSIDH-512**, which is chosen to fulfill the requirements of NIST Security Level 1 (see Section 4.2.1 for the definition). Every mathematical concept is discussed in the relevant sections of Chapter 2. The lower-level algorithms for calculations are discussed in the remaining

subsections, and they are described in general, i.e. the algorithm are not specific for the CSIDH-512 instantiation. This section follows more or less the original CSIDH paper [17, Section 5, 6, 8].

3.2.1 Key Exchange Protocol

The **global parameters** are the prime p and the starting curve E_0. The prime is in the form

$$p = 4 \cdot \prod_{i=1}^{74} \ell_i - 1 = 4 \cdot \ell_1 \cdot \ell_2 \cdots \ell_{74} - 1,$$

where the $\ell_1, \ldots, \ell_{73}$ denote the smallest 73 *odd* primes, and $\ell_{74} = 587$. The global starting curve is defined over the finite field $\mathbb{F}_p$, and it has the endomorphism ring $\mathcal{O} = \mathbb{Z}[\sqrt{-p}]$. The curve is given in Montgomery form as

$$E_0 : y^2 = x^3 + x.$$

`KeyGen`– The **key generation** is as follows: Both parties 74 numbers uniformly at random sampled from the range $\{-5, \ldots, 5\}$. The sampled 74-tuple $(e_1, \ldots, e_{74})$ is the **private key**. These integers represent the exponents of the ideal class representatives

$$[\mathfrak{a}] = \left[\mathfrak{l}_1^{e_1} \mathfrak{l}_2^{e_2} \cdots \mathfrak{l}_{74}^{e_{74}}\right] \in \mathrm{cl}(\mathcal{O}),$$

where the $\mathfrak{l}_i$ denote the fractional $\mathcal{O}$-ideals $(\ell_i, \sqrt{-p} - 1)$. Algorithm 1 is used to compute the group action, which returns the Montgomery coefficient $A \in \mathbb{F}_p$ of the elliptic curve $[\mathfrak{a}]E_0 : y^2 = x^3 + Ax^2 + x$. This coefficient A is the **public key**.

`SharedKey` – The **key exchange** is as follows: Suppose Alice has the private-public key pair $([\mathfrak{a}], A)$, and Bob has the private-public key pair $([\mathfrak{b}], B)$. When either of the parties receive each other's keys, they verify the public key with Algorithm 2 (potentially including Algorithm 3). If the verification is successful, then Alice applies her secret key $[\mathfrak{a}]$ to the curve $E_B : y^2 = x^3 + Bx^2 + x$ with Algorithm 1, and gets the curve $[\mathfrak{a}]E_B = [\mathfrak{a}][\mathfrak{b}]E_0$. Similarly, Bob applies his secret key $[\mathfrak{b}]$ to Alice's public key $E_A : y^2 + x^3 + Ax^2 + x$ with Algorithm 1, and gets the curve $[\mathfrak{b}]E_A = [\mathfrak{b}][\mathfrak{a}]E_0$. The **shared secret key** is the Montgomery coefficient S of the curve $[\mathfrak{a}][\mathfrak{b}]E_0 = [\mathfrak{b}][\mathfrak{a}]E_0$, written in the form $E_S : y^2 = x^3 + Sx^2 + x$. This coefficient is the same for both parties, because of the commutativity of $\mathrm{cl}(\mathcal{O})$ and Theorem 2.7.16, so by Definition 3.0.1, this NIKE is correct.

3.2.2 Isogeny calculation

CSIDH does not use the Vélu formula for its calculation of the isogenies, rather they use the projectivized variant of Renes' formula [17, pp. 26] discussed in Section 2.3. For completeness sake, we record it here as well, since Renes' formula is given in its affine form in Section 2.3.

For a fixed prime $\ell \geq 3$, a point P of order ℓ, the Montgomery coefficient A and an integer $k \in \{1, \ldots, \ell - 1\}$, let $(X_k : Z_k)$ be the projectivized x-coordinate of $[k]P$. Then if we define $c_i \in \mathbb{F}_p$, such that the following are equal as polynomials in w,

$$\prod_{i=1}^{\ell-1} (Z_i w + X_i) = \sum_{i=0}^{\ell-1} c_i w^i,$$

then we have that

$$(\tau(A - 3\sigma) : 1) = (Ac_0c_{\ell-1} - 3(c_0c_{\ell-2} - c_1c_{\ell-1}) : c_{\ell-1}^2),$$

where

$$\tau = \prod_{i=1}^{\ell-1} \frac{X_i}{Z_i}, \quad \text{and} \quad \sigma = \sum_{i=1}^{\ell-1} \left(\frac{X_i}{Z_i} - \frac{Z_i}{X_i} \right).$$

As a standard feature of elliptic curves, we have that $x([k]P) = x([\ell - k]P)$, so in particular we can half the required computations to get the Montgomery curve of the codomain, which is $B = \tau(A - 3\sigma)$ for the isogeny with kernel $\langle P \rangle$.

3.2.3 Evaluation of the ideal class group action

The group action is defined by evaluating ℓ_i-isogenies and calculating their Montgomery coefficient A. For the ℓ_i-isogenies, we need points P_i on the curve with order ℓ_i, so that it can generate the kernel for the isogeny. A random $x \in \mathbb{F}_p$ is sampled, and then a quadratic residue check is performed. If the point is a square, we are on E, but if not, then we can go over to the quadratic twist E^t, and perform the steps there. These steps are repeated until all of the private key exponents e_i have been calculated, since the e_i is decreased (or increased) only when a successful isogeny calculation has been computed. Note also that because of the commutativity of the class group $\mathrm{cl}(\mathcal{O})$, the order of isogeny evaluations does not matter. The algorithm is defined as follows.

Algorithm 1 Evaluating the ideal class group action

Input: $A \in \mathbb{F}_p$ and a list of integers $(e_1, \ldots, e_n)$.
Output: $B \in \mathbb{F}_p$, such that $[\mathfrak{l}_1^{e_1} \cdots \mathfrak{l}_n^{e_n}]E_A = E_B$, where $E_B : y^2 = x^3 + Bx^2 + x$).

1: **while** some $e_i \neq 0$ **do**
2: Sample a random $x \in \mathbb{F}_p$.
3: **if** $x^3 + Ax^2 + x$ is a square in $\mathbb{F}_p$ **then**
4: $s \leftarrow +1$ ▷ Step right
5: $P \leftarrow (x, \sqrt{x^3 + Ax^2 + x})$
6: **else**
7: $s \leftarrow -1$ ▷ Step left
8: $P \leftarrow (-x, \sqrt{-x^3 - Ax^2 - x})$ ▷ Quadratic twist $E_A^t \cong E_{-A}$
9: **end if**
10: $S = \{\, i \mid e_i \neq 0,\ \mathrm{sign}(e_i) = s \,\}$ ▷ Collect all primes with same direction.
11: **if** S is empty **then** start over with a new x.
12: **else**
13: $k \leftarrow \prod_{i \in S} \ell_i$
14: $Q \leftarrow [(p+1)/k]P$ ▷ order of Q divides $k = \prod \ell_i$
15: **for** $i \in S$ **do**
16: $R \leftarrow [k/\ell_i]Q$
17: **if** $R = \infty$ **then** skip this i ▷ if $R = \infty$, then initial point P had no ℓ_i-torsion
18: **else**
19: Compute an isogeny $\varphi : E_A \to E_B : y^2 = x^3 + Bx^2 + x$ with $\ker \varphi = R$.
20: $Q \leftarrow \varphi(Q)$ ▷ Push Q through the isogeny to avoid scalar multiplication.
21: $A \leftarrow B, k \leftarrow k/\ell_i, e_i \leftarrow e_i - s$
22: **end if**
23: **end for**
24: **end if**
25: **end while**
26: **return** A

Remark 3.2.1. The original [17, Algorithm 2] does not state how P is assigned, so that is fixed in line 5 and 8.

The crucial performance optimization is the fact that the generated point Q is pushed through the isogeny, so it can be used again until all the ℓ_i primes are calculated that $\mathrm{ord}(Q)$ contained. This saves a lot of time, because we do not have to sample a random point, perform a square-root check, and calculate a big scalar multiplication every time. It also helps to not discard a sampled x-coordinate, but instead go over to the quadratic twist, which has Montgomery coefficient $-A$.

3.2.4 Public Key Validation

In this subsection, we are going to discuss and give the algorithm used for verifying a public key in the CSIDH protocol. Since we are working over supersingular curves, by Theorem 2.5.9, we have that $\#E(\mathbb{F}_p) = p + 1$. Therefore, we need to count the number of points on the public curve E_A, and verify that it is $p + 1$. By a standard theorem of number theory (Lagrange's theorem), any order of a point

$P \in E_A$ must divide the size of the full group. Hasse's theorem 2.5.7 gives us a range of possible number of points as

$$[p + 1 - 2\sqrt{p}, \quad p + 1 + 2\sqrt{p}].$$

If we find a point whose order d is greater than $4\sqrt{p}$, and divides $p+1$, then that means there exists only one multiple of d in the Hasse interval, so we must have that $\#E(\mathbb{F}_p) = p + 1$, so it is a valid key. If the

Algorithm 2 Public Key Verification

Input: Elliptic curve $E_A(\mathbb{F}_p) : y^2 = x^3 + Ax^2 + x$ with $p = 4 \cdot \ell_1 \cdots \ell_n - 1$.
Output: TRUE or FALSE

1: Pick a point $P \in E(\mathbb{F}_p)$ uniformly at random.
2: $d \leftarrow 1$
3: **for** each ℓ_i **do**
4: $\quad Q_i \leftarrow [(p+1)/\ell_i]P$
5: $\quad$ **if** $[\ell_i]\mathbb{Q}_i \neq \infty$ **then**
6: $\quad\quad$ **return** FALSE ▷ $\#E(\mathbb{F}_p)$ does not divide $p + 1$.
7: $\quad$ **end if**
8: $\quad$ **if** $Q_i \neq \infty$ **then**
9: $\quad\quad d \leftarrow \ell_i \cdot d$ ▷ In this case ℓ_i divides the order of P.
10: $\quad$ **end if**
11: $\quad$ **if** $d > 4\sqrt{p}$ **then**
12: $\quad\quad$ **return** TRUE
13: $\quad$ **end if**
14: **end for**

algorithm didn't return TRUE or FALSE, then the sampled P had too small order. This is very unlikely, however, because by construction, the structure of

$$E(\mathbb{F}_p) = \mathbb{Z}/4\mathbb{Z} \times \prod_{i=1}^{n} \mathbb{Z}/\ell_i\mathbb{Z},$$

and since the point P is chosen randomly, the chance that ℓ_i divides the order of P is $1 - 1/\ell_i$. Ignoring the $\mathbb{Z}/4\mathbb{Z}$ part, the random variable d''s contribution to the order is the expectation

$$d' = \prod_{i=1}^{n} X_i, \quad \text{where } X_i = \begin{cases} \ell_i & \text{with probability } (\ell_i - 1)/\ell_i \\ 1 & \text{with probability } 1/\ell_i \end{cases}.$$

This shows that the total expectation is then

$$\mathbb{E}(d') = \prod_{i=1}^{n} \mathbb{E}[X_i] = \prod_{i=1}^{n} \left(\ell_i \cdot \frac{\ell_i - 1}{\ell_i} + \frac{1}{\ell_i} \right) = \prod_{i=1}^{n} \left(\ell_i - 1 + \frac{1}{\ell_i} \right).$$

This product is very close to the product of all the ℓ_i, which is approximately $(p + 1)/4 \gg 4\sqrt{p}$. This shows that the chance of failure is almost 0.

Actually, we do not have to sample a full point, since the x-coordinate of any point works. This is true because the quadratic twist E^t of E is also supersingular by Remark 2.5.12, so the specific curve does not matter. Another performance speed up is possible if we trade the running time to memory cost. Algorithm 3 can be used to precompute a lot of points, that can be then used already for the verification step, this way we are saving on a lot of serial scalar multiplication steps. The recursive algorithm is the following.

This algorithm can be used by picking a random point $P \in E(\mathbb{F}_p)$ and run the algorithm on input $[4]P$ and $(\ell_1, \ldots, \ell_n)$, which will compute the points $Q_i = [(p + 1)\ell_i]P$. We can then go back to Algorithm 2 to verify that E is supersingular using these precomputed points. Crucially, this algorithm works on global parameters, so in particular it does *not* need to be secured against timing/side-channel attacks.

Algorithm 3 Batch cofactor multiplication

Input: An elliptic curve point P and positive integers $(k_1, \ldots, k_n)$

Output: The points $(Q_1, \ldots, Q_n)$, where $Q_i = \left[\prod_{j \neq i} k_j\right] P$.

1: **if** $n = 1$ **then** ▷ base case
2: **return** (P)
3: **end if**
4: $m \leftarrow \lceil n/2 \rceil$
5: $u \leftarrow \prod_{i=i}^{m} k_i$
6: $v \leftarrow \prod_{i=m+1}^{n} k_i$
7: $L \leftarrow [v]P$
8: $R \leftarrow [u]P$
9: **Recurse** with input $L, (k_1, \ldots, k_m)$ giving $(Q_1, \ldots, Q_m)$ ▷ left half
10: **Recurse** with input $R, (k_{m+1}, \ldots, k_n)$ giving $(Q_{m+1}, \ldots, Q_n)$ ▷ right half
11: $(Q_1, \ldots, Q_n)$

Chapter 4

Security of CSIDH

In this chapter, we are going to give the security considerations that were used to determine the specific instantiation of the prime p used in the protocol. It is not enough to just focus on the quantum attacks, since everybody still uses classical computers, so we will start our discussion with such classical attacks. In the section after that, we are going to give a short background on quantum information theory, which should give enough knowledge to understand the mechanisms of the most known quantum attacks – Shor's algorithm and Grover's algorithm. In subsection 4.2.5, we will present the current best quantum attack against CSIDH, which has been a substantial blow to the presumed security.

First, before going into the attacks and potential fall of CSIDH, let us start by the proven breakthroughs and advantages of CSIDH in the context of isogeny-based cryptography. As discussed in Section 3.1.4, the precursor of CSIDH was the key exchange protocol developed by Couveignes [27], and then independently by Rostovtsev-Stolbunov [61]. Their scheme was the first NIKE scheme built on the action of an ideal class group, which worked on *ordinary* elliptic curves with a common endomorphism ring $\mathcal{O}$. One problem was that it was hard to find primes ℓ_i that split in $\mathbb{Z}[\sqrt{-p}]$ so that the isogeny graph had horizontal steps. Another, bigger issue was that computing ℓ-isogenies with ℓ-torsion points over ordinary curves had to be over large extension fields $\mathbb{F}_{q^k}$, which resulted in running times of minutes to hours even. CSIDH's ingenious idea was to do the same protocol over a supersingular curve, where point counting is much easier, and controlling the splitting primes is also much easier.

SIDH is based on supersingular curves as well, but it uses the full supersingular graph, which is acted on noncommutatively, so they also had to include torsion points in their public keys during the key exchange. CSIDH does not use any torsion points, so the devastating attacks of SIDH [15, 48, 58] had no impact on CSIDH, surviving the fallout. Furthermore, CSIDH boasts the smallest public keys out among all the post-quantum cryptographic schemes. Since CSIDH only needs the Montgomery coefficient A to uniquely identify the vertex of the isogeny graph, it only needs 64 bytes. Compare this to ML-KEM's [54] 800 bytes and HQC's [36] 2241 bytes, the two standardized PQC schemes by NIST.

CSIDH also has a very fast public key verification, since it is very easy to verify whether a public key A is the Montgomery coefficient of a valid supersingular elliptic curve. This gives a natural protection against the **Chosen-Ciphertext Attack (CCA)**, which means that an attacker has access to the decapsulation algorithm and can manipulate the ciphertexts, from which they can deduce information about the secret key. In the form of a NIKE, there is no ciphertext to talk about, but CSIDH can form the basis for a KEM as well, in which case because of the fast verification, and determinism of the group action, there is no efficient way to leak information about the secret key. Other KEMs need to use the Fujisaki-Okamoto Transform [35] to achieve CCA-security, which adds some overhead to the calculations.

4.1 Classical Security

In this section, we are going to discuss the extent to which classical attacks apply to break CSIDH. To start with, we state the underlying hard problem that is the basis for CSIDH's cryptographic relevance.

Problem 4.1.1 (Key recovery)**.** [17, Problem 10] Given two supersingular elliptic curves $E(\mathbb{F}_p), E'(\mathbb{F}_p)$ such that $\mathrm{End}_p(E) \cong \mathrm{End}_p(E') = \mathcal{O}$, find an ideal $\mathfrak{a}$ of $\mathcal{O}$ such that $[\mathfrak{a}]E = E'$. It is also required that the ideal $\mathfrak{a}$ must be represented in a way that the action of $\mathfrak{a}$ on the curve can be evaluated efficiently.

This problem is a specification of the more general problem, which is called the **Group Action Inverse Problem (GAIP)**.

Problem 4.1.2 (GAIP)**.** Let G be a group acting faithfully and transitively on a set X. Given two elements $x, y \in X$, find the group element $g \in G$ such that $g * x = y$.

4.1.1 Brute force attack

Brute force attack is the most obvious way to attack any cryptographic algorithm, and unsurprisingly it is also the least effective. It is based on the try-and-fail approach, where an attacker would try all the possible private keys, and see if it matches. As explained before in Subsection 3.2.1, the private key in CSIDH consists of the 74 exponents of the ideals $\mathfrak{l}_1^{e_1}, \ldots, \mathfrak{l}_{74}^{e_{74}}$. All of the e_i's are in the range $\{-5, \ldots, 5\}$, so the total key space has size $11^{74} \approx 2^{256}$ possible key vectors. This is no coincidence, because as shown in Remark **??**, the class group $\mathrm{cl}(\mathcal{O})$ is in bijection with $Ell_p(\mathcal{O}, \pi)$, which constitutes the vertices of the isogeny graph. Then by Theorem 2.4.39, we have that $|\mathrm{cl}(\mathcal{O}_K)| \approx \sqrt{|\Delta_K|}$, and since $\Delta_K = -p$, for our case to heuristically make sure that every vertex (elliptic curve) is reachable from the $\sqrt{p} = 2^{256}$ possibilities, we need to tune the size of the secret key vector bounding size. The full argument of why this sampling method is justified, involving estimating with min-entropy, is further detailed in [17, §7.1]. The reason for why this argument had to be made is that there can be a lot of different configurations of the secret keys $(e_1, \ldots, e_n)$ that once are evaluated, result in the same elliptic curve as another, potentially smaller representation. At the time of publishing, they did not know the exact size and structure of the class group used, but since then [5] has calculated that.

The CSI-FiSh team has proved that $\mathrm{cl}(\mathbb{Z}[\sqrt{-p}])$ is cyclic, is generated by the first ideal $\mathfrak{l}_1$ and has size $\sim 2^{257}$, with the knowledge of the structure of the class group, one could get true random sampling for the secret key, but analysis showed that it did not change much in performance. Knowing the relations matrix therefore does not aid or harm CSIDH, or an attacker looking to break CSIDH.

4.1.2 Meet-in-the-middle attack

A more sophisticated approach is the style of the **Meet-in-the-Middle (MITM)** attack. This is more of an umbrella term for all kinds of algorithms that divide the search space in some way, and then through some searching algorithm finds a collision. The classic example of such an attack is the **Baby-step giant-step** method. The baby-steps giant-steps method was first introduced by Daniel Shanks in [65] and subsequently generalized by many authors. It solves the discrete logarithm problem in a cyclic group in the following way. This algorithm is guaranteed to find a solution, since the indexing i, j cover the full

Algorithm 4 Pseudo-code for Baby-step Giant-step

Input: A cyclic group G of order n, a generator g, and an element h
Output: x such that $g^x = h$

$m \leftarrow \lceil \sqrt{n} \rceil$
$L \leftarrow []$ ▷ Empty list
for $0 \leq j < m$ **do**
    Compute g^j
    Append (j, g^j) to L ▷ List of baby steps
end for
$A \leftarrow g^{-m}$
$H \leftarrow h$
for $0 \leq i < m$ **do**
    if H is in L **then**
        return $im + j$ ▷ Get j from first component of the tuple
    else
        $H \leftarrow H \cdot A$ ▷ Take a giant step
    end if
end for

size of the group G. It returns the discrete logarithm, because when a match is found, then we have that

$$g^j = h \cdot (g^{-m})^i \implies g^{j+mi} = h.$$

This algorithm has $O(\sqrt{n})$ space and time complexity, which is exponentially better than the naïve brute force attack.

In our case, since the private key space naturally decomposes into two smaller sets, the baby-step giant-step algorithm is applicable to CSIDH by splitting $[\mathfrak{l}_1^{e_1}\mathfrak{l}_2^{e_2}\cdots\mathfrak{l}_n^{e_n}]$ into $[\mathfrak{l}_1^{e_1}\cdots\mathfrak{l}_k^{e_k}]\cdot[\mathfrak{l}_{k+1}^{e_{k+1}}\cdots\mathfrak{l}_n^{e_n}]$ for some $k \in \{1,\ldots,n\}$. Since the size of the underlying set is $\sqrt{p}$, the result of this algorithm is that it runs in $O(\sqrt[4]{p})$ steps. We know from before that CSIDH-512 has a 512-bit prime, so the classical security of CSIDH is estimated to be $\lambda = \log(2^{512/4}) = 128$.

Another approach that uses the idea of dividing up a search concerns the isogeny graph itself. If an attacker finds a way to solve the Isogeny Finding Problem (see Problem 3.1.2), then they have broken CSIDH, since then they can apply the broken isogeny on the other person's public key, and arrive at the shared key. This problem can be solved by starting a breadth-first search from the starting node E_0 and one of the public keys E_A, then if a collision is found, then the path to the point can be composed into an efficient isogeny (provided the search isogeny steps were also easily computable). One would expect that this method also has $O(\sqrt{n})$ time complexity, but Benjamin Wesolowski [77] published a new method that lowers this time complexity to $O(p^{1/3+o(1)})$, which is a breakthrough from the previous $O(p^{1/2}\cdot(\log p)^{O(1)})$. This does not affect CSIDH, however, because that problem is for the full supersingular graph in $\mathbb{F}_{p^2}$, but CSIDH lives in $\mathbb{F}_p$. Finding an isogeny over $\mathbb{F}_{p^2}$ has a very small chance of being defined over $\mathbb{F}_p$, but even if it applied to CSIDH, it has a much bigger quantum problem.

4.1.3 Pohlig-Hellman-style attacks

The Pohlig-Hellman attack is an efficient method for computing discrete logarithms when the order of the group G is smooth, meaning that it decomposes into small prime powers. The algorithm relies on this decomposition, solving the DLP in the small prime-power subgroups, then uses the Chinese Remainder Theorem to build the global discrete logarithm back up. A lot of small primes might sound exactly like CSIDH, but the CSIDH prime is irrelevant in this case. The reason is because the set $Ell_p(\mathcal{O},\pi)$ has no group structure on its own, and the class group $\mathrm{cl}(\mathcal{O})$ is large, and crucially by [5], we also know that it is very much non-smooth.

4.2 Quantum Security

In the previous chapter, we have looked at the possible classical attacks against the CSIDH protocol, and what it means for the size of the prime p to be large enough for classical safety. In this chapter, we give a short explanation of the quantum security levels proposed by the National Institute of Standards and Technology (NIST), which provide the basis of comparing different post-quantum options. In section 4.2.2, we give a basic introduction to quantum information theory, which covers the theoretical background necessary to understand the coming sections. In the last three section, we explore the most known quantum attacks, explain their applicability to CSIDH, and review the biggest quantum threat to CSIDH as of now, which is Kuperberg's algorithm. These sections will roughly follow the Lecture notes of Changpeng Shao [66] and Nielsen-Chuang's seminal book on quantum information theory [23].

4.2.1 NIST Security Levels

The National Institute of Standards and Technology (NIST) has called for a Post-Quantum Cryptography Standardization competition for finding new protocols that could withstand the threat of quantum computers. In their Call for Proposals [53], they specify the minimum acceptability requirements and evaluation criteria among many other (perhaps rather bureaucratic) things. They specify [53, §4.A.5] the following security levels for the contestants, listed in order of increasing strength:

Level 1 Any attack must require comparable or more resources than breaking **AES-128** by *bruteforce*.
Level 2 Any attack must require comparable or more resources than breaking **SHA-256** by *collision*.
Level 3 Any attack must require comparable or more resources than breaking **AES-192** by *bruteforce*.
Level 4 Any attack must require comparable or more resources than breaking **SHA-384** by *collision*.
Level 5 Any attack must require comparable or more resources than breaking **AES-256** by *bruteforce*.

As per NIST recommendations, they suggested restricting the quantum attacks to a fixed running time / circuit depth. They called this parameter **MAXDEPTH**, and is motivated by the difficulty of running too long serial computations on quantum computers. Their proposed [53, §4.A.5] values are ranging between

- 2^{40} logical gates, corresponding to the estimated number of gates for a *quantum* computing architecture to serially perform *in a year*;
- 2^{64} logical gates, corresponding to the estimated number of gates for a *classical* computing architecture to serially perform *in a decade*;
- 2^{96} logical gates, corresponding to the estimated number of gates that atomic scale qubits with speed of light propagation times could perform *in a millennium*.

These complexities can be measured in terms of T-gates (see Section 4.2). NIST gives the following estimates for the 5 security levels:

1. **AES-128:** 2^{170}/MAXDEPTH quantum gates or 2^{143} classical gates.
2. **SHA-256:** 2^{146} classical gates.
3. **AES-192:** 2^{233}/MAXDEPTH quantum gates or 2^{207} classical gates.
4. **SHA-384:** 2^{210} classical gates.
5. **AES-256:** 2^{298}/MAXDEPTH quantum gates or 2^{272} classical gates.

These definitions and requirements will be discussed at length in Section 4.2.

We will not mention digital signature schemes in this thesis, but they are equally as important to secure against quantum attacks as well. Alas, there are many different types of KEMs and public-key encryption schemes that have competed in NIST's competition. In the following paragraphs, we will briefly overview the mathematical foundations of their hardness assumptions, and present the winners of the NIST competition.

The first and most well-studied algorithmic family is **lattice-based** cryptography. One of its earliest versions is called NTRU, and it is based on hardness of geometric problems over high-dimensional lattices. A toy example would look like choosing a point in the Cartesian plane, then it is easy to describe as coefficients of the usual basis. However, if there is a horrible basis chosen (for example the basis $e_1 = (1, 1000000)^T$ and $e_2 = (1, 1349738)^T$); this task can become very difficult. This can also be combined with something called **Learning With Errors**, which can be described so. Let $R_q = \mathbb{F}_q[X]/(X^n + 1)$ be a polynomial ring, then given a public matrix $A \in R_q^{k \times k}$, a secret vector $\mathbf{s}$ and error $\mathbf{e}$, then we get a cipher text

$$\mathbf{b} = A\mathbf{s} + \mathbf{e} \pmod{q}.$$

Recovering $\mathbf{s}$ from this linear equation requires solving the Shortest Vector Problem. This problem is thought to be very hard for even quantum algorithms. This is the basis that the first winner of NIST also used, which was called CRYSTALS-KYBER in the contest, but now has been standardized with the new name **ML-KEM** [54].

Code-based cryptography is also a popular and well-founded post-quantum basis. They rely on the NP-hardness of decoding a general random linear error-correcting code. The second winner that has been chosen for standardization uses this coding theory basis, and it's called HQC [36] for "Hamming Quasi-Cyclic" as of writing this thesis. The intrigued reader may find a lot of coding theory basics in the HQC paper.

The last (but perhaps least?) major algorithmic family is the topic of this thesis, which is called **isogeny-based** cryptography. The idea of these schemes are based on the maps between elliptic curves, and their hard properties to find these maps given two elliptic curves. This area is the youngest, as it only has been really studied for the past 25 years, and as a result, it is the most volatile. Unfortunately, CSIDH could not compete in the NIST competition, since its development dates 2 years after the start. There was, however, an isogeny-based candidate, which shares a very similar name.

4.2.2 Quantum Information Theory Basics

Classical computers operate on the basis of having two states that can be manipulated; the bits 0 and 1. These bits are then manipulated by logic gates, which can take some number of inputs. A 1-input gate

example is `NOT`, which flips the bit to the other one, and 2-input gates are the well-known `AND`, `OR`, `XOR` etc. Crucially, these gates "destroy information", since they take 2 bits, and output 1 bit only. This will be important later.

In quantum computing, we do not only work with 2 states. We have access to infinite states since these quantum systems can be in superposition of many different states at the same time. This also affects the physical computations, because keeping these states in superposition / entangled requires high-end, state-of-the-art engineering, but let us ignore these materialistic challenges – we are mathematicians after all!

Definition 4.2.1. The **Dirac notation** is the system to denote vectors and states in a Hilbert space, and thus in quantum algorithms. The symbol $|\psi\rangle$ is called **ket**, and denotes the column vector in the Hilbert space. The symbol $\langle\psi|$ is called **bra**, and denotes the conjugate (Hermitian) transpose of $|\psi\rangle$. This operation is also denoted by the dagger notation $\langle\psi| = |\psi\rangle^\dagger$.

Definition 4.2.2. A **qubit** is unit vector in a two-dimensional complex Hilbert space $\mathbb{C}^2$. A collection of qubits is called a **(quantum) register**. The qubits are represented in Dirac notation as a linear combination of the orthogonal basis $|0\rangle$ and $|1\rangle$. We call this basis the **computational basis**, and we define

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix} \quad \text{and} \quad |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}.$$

A general qubit $|\psi\rangle$ is then written as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \quad \alpha,\beta \in \mathbb{C} \quad \text{and} \quad |\alpha|^2 + |\beta|^2 = 1.$$

The last part is there to ensure that the state is normalized. By taking the tensor products of qubits, we get a basis of the form $\{|x\rangle : x \in \{0,1\}^n\}$. For ease of notation, we will denote this basis in various ways, based on whichever aids the understanding more, so we have

$$|x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle = |x_1\rangle\,|x_2\rangle \ldots |x_n\rangle = |x_1, x_2, \ldots, x_n\rangle = |x\rangle\,.$$

Since these are pure tensors, any quantum state is a unit linear combination of $|x\rangle\,, x \in \{0,1\}^n$. As a result, quantum states are unit, so we can use unitary operators to them while keeping the norm.

Definition 4.2.3. An operator $U \in \mathbb{C}^n$ is called **unitary** if $UU^\dagger = U^\dagger U = I$. A **quantum gate** is some unitary operator $U \in \mathbb{C}^2$.

Some examples of quantum gates that are especially useful are the **Hadamard gate**

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix},$$

the **Pauli matrices**

$$X = \sigma_x = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}, \quad Y = \sigma_y = \begin{pmatrix}0 & i\\-i & 0\end{pmatrix} \quad \text{and} \quad Z = \sigma_z = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix},$$

the **T-gate**

$$\begin{pmatrix}1 & 0\\0 & e^{i\pi/4}\end{pmatrix},$$

and the **Controlled-G gates**. For any gate G, Controlled-G gates use an extra qubit that controls if the gate is applied or not. So we have that $CG\,|0\rangle\,|x\rangle = |0\rangle\,|x\rangle$, and $CG\,|1\rangle\,|x\rangle = |0\rangle\,G\,|x\rangle$. The T-gates are referred to most in quantum query complexity analysis (for example the NIST requirements in Section 4.2.1, because they are the most expensive compared to other gates. [4, Appendix A.4]

Definition 4.2.4. A **measurement** of a single qubit $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ collapses the state into $|0\rangle$ or $|1\rangle$ with probability $|\alpha|^2$ and $|\beta|^2$ respectively. Measuring a single qubit i in a superposition of n qubits

is called **partial measurement**, and the probability of measuring $x \in \{0,1\}$ is

$$p(x) = \sum_{j=1}^{n} |\alpha_x|^2,$$

where α_x are the coefficients of the basis states that have value x in the i'th qubit. When x is seen, the collapsed state is the sum of the states with x in position i, renormalized again by $\sqrt{p(x)}$ to 1.

Example 4.2.5. Let $|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle)$ be a superposition of two qubits. Measuring the first qubit will give 0 with probability $|\alpha|^2 + |\beta|^2$, and result in the new state

$$|\psi_{new}\rangle = \frac{\alpha|00\rangle + \beta|01\rangle}{\sqrt{|\alpha|^2 + |\beta|^2}}.$$

Since the computational basis is orthonormal, measuring any qubit is equivalent to projecting the state onto the subspace spanned by that qubit. For this reason, the measurement will sometimes be denoted by a projection $P_x = |x\rangle\langle x|$, and in this case, the probability of seeing $x \in \{0,1\}^n$ in $|\psi\rangle = \sum \alpha_x |x\rangle$ is expressed as

$$p(x) = \langle\psi| P_x |\psi\rangle = |\alpha_x|^2,$$

and the state of the system becomes

$$|\psi_{new}\rangle = \frac{P_x|\psi\rangle}{\sqrt{p(x)}}.$$

As mentioned before, classical algorithms are information-destroying, or in other words, non-reversible. The problem in quantum computing is that operations have to be reversible (since the operators are unitary matrices), so one might think that classical algorithms cannot be imported into the quantum setting. This is actually false, because indeed reversible variants of classic algorithm can be implemented in the following way.

Lemma 4.2.6. *Any classical circuit can be implemented as a quantum circuit (modulo physical challenges).*

Proof. If we have a classical function $f : \{0,1\}^n \to \{0,1\}^m$ that takes a string of n bits to m bits, then we can attach an **ancilla register** of m bits to the right set to 0, and modify f to give a new operation

$$f' : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m \times \{0,1\}^m,$$

defined by

$$f'(x,y) = (x, y \oplus f(x)),$$

where $\oplus$ is the bitwise `XOR` operation (also known as addition modulo 2). clearly we have then

$$f'(f'(x,y)) = f'(x, y \oplus f(x)) = (x, y \oplus f(x) \oplus f(x)) = (x,y),$$

so $f'^2 = 1$, hence reversible. Any classical circuit is made up of these bitstring functions in the form of classical logic gates, so the claim is shown. □

Example 4.2.7. We can implement the classical `AND` gate $f(x_1,x_2) = x_1 \wedge x_2 = x_1x_2$ on a quantum computer as

$$f'(x_1,x_2,y) = (x_1,x_2,y \oplus x_1x_2).$$

In matrix form, it looks like

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and is known as the **Toffoli gate** or **CCNOT gate**.

The Toffoli gate is very special when paired with the Hadamard gate, since it is known to be *sufficient* for universal quantum computation, which means that like the `NAND` gate in classical algorithms, the Toffoli and Hadamard gates are sufficient to represent any quantum computation or circuit.

Time complexity in a quantum setting refers to the number of elementary gates used in a universal set, but this is hard to determine due to technical details obscuring the key features of an algorithm. To circumvent this, we usually talk about *query complexity* instead, in which we assume that we have access to an **oracle** (or a "black box"), which returns answers to queries. Thus the goal of query complexity is to minimize the number of queries to achieve something. In light of the previous construction, let us define two more important operators.

Definition 4.2.8. Let $f : \{0,1\}^n \to \{0,1\}^m$ be a function. The **Evaluation oracle** is the operator defined by

$$O_f : |x, y\rangle \longmapsto |x, y \oplus f(x)\rangle .$$

When $m = 1$, then the **Phase oracle** is an operator defined by flipping the phase of an input state

$$U_f : |x\rangle \longmapsto (-1)^{f(x)} |x\rangle .$$

4.2.3 Shor's Algorithm

Now that we have introduced the basics of quantum programming, we are soon able to describe Shor's Algorithm [67], and why it completely breaks cryptographic primitives that rely on factoring integers. First, we state the core mechanism inside Shor's algorithm.

Definition 4.2.9. The **Quantum Fourier Transform (QFT)** (over $\mathbb{Z}/N\mathbb{Z}$) is the unitary operator

$$Q_N : |x\rangle \longmapsto \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} \omega^{xy} |y\rangle ,$$

where $\omega = e^{2\pi i/N}$.

This is the same transformation as the Discrete Fourier Transform that the reader might be familiar with already. QFT is important to determine the period of a periodic function that is also one-to-one on each period (it is like a class function for sets). Assuming these conditions, our task is to determine the period r for $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/M\mathbb{Z}$, where we have that $f(x) = f(x + r)$. This can be done by the following algorithm.

Theorem 4.2.10. *Algorithm 5 correctly returns the period r.*

Proof. The first 3 lines of Algorithm 5 result in the state

$$|0\rangle |0\rangle \longmapsto \frac{1}{\sqrt{N}} \sum_{x=1}^{N-1} |x\rangle |0\rangle \longmapsto \frac{1}{\sqrt{N}} \sum_{x=1}^{N-1} |x\rangle |f(x)\rangle .$$

When the second register is measured, then we receive some value z. Since the new measured state is the sum of the states that has z in the second register, by the assumed periodicity and one-to-one properties

Algorithm 5 Pseudo-algorithm to determine the periodicity r of the function f

1: Initialize $|0\rangle |0\rangle$, where the first register has dimension N and the second has dimension M.
2: Apply Q_N to the first register.
3: Apply O_f to both registers.
4: Measure the second register.
5: Apply Q_N to the first register.
6: Measure the first register, save the answer k.
7: Simplify the fraction k/N as much as possible, and return the denominator.

of f, all input values $x \in \mathbb{Z}/N\mathbb{Z}$ for which $f(x) = z$ is true, are of the form $x_0 + jr$ for some integer x_0, j. Thus the first qubit's state collapses to

$$\sqrt{\frac{r}{N}} \sum_{j=0}^{N/r-1} |x_0 + jr\rangle ,$$

where the extra $\sqrt{r}$ term comes from the normalization step. After this, we apply Q_N again, resulting in the state

$$\sqrt{\frac{r}{N^2}} \sum_{y=0}^{N-1} \omega^{x_o y} \left(\sum_{j=0}^{N/r-1} \omega^{jry} \right) |y\rangle .$$

Note that if $\omega^{ry} \neq 1$, then

$$\sum_{j=0}^{N/r-1} \omega^{jry} = \frac{\omega^{ry(N/r)} - 1}{\omega^{jry} - 1} = 0.$$

When $\omega^{ry} = 1$, which is the same as N dividing ry, then the summation equals N/r, so the only surviving states combine to

$$\frac{1}{\sqrt{r}} \sum_{l=0}^{r-1} \omega^{x_o N l/r} |Nl/r\rangle .$$

When we perform line 6, that is measuring the first register again, we receive some outcome $k = l_0 N/r$ at random with uniform probability $0 \leq l_0 \leq r-1$. We now know both k and N, so we also know that

$$\frac{k}{N} = \frac{l_0}{r}.$$

If we are lucky and $\gcd(l_0, r) = 1$ then reducing the fraction k/N is indeed going to give l_0/r, and the denominator is going to be the period r as required. If it's not (which can be checked classically by computing $f(r)$ and comparing it to $f(0)$), then we run the circuit again. □

Now that we know how the QFT works, we are ready to give the (pseudo-)algorithm developed by Peter Shor [67] to factor large semiprimes of the form $N = p \cdot q$.

Algorithm 6 Pseudo-code for Shor's algorithm

Input: N
Output: p, q
1: Choose $1 < a < N$ uniformly at random.
2: Compute $d = \gcd(a, N)$. If $d > 1$, output d and stop.
3: Compute $r = \mathrm{ord}(a)$. If r is odd, go to step 1.
4: Compute $s = \gcd(a^{r/2} - 1, N)$. If $s = 1$, go to step 1.
5: **return** $s, N/s$.

This works because if a is coprime to N and r is even, then we have that

$$a^r - 1 = (a^{r/2} - 1)(a^{r/2} + 1) \equiv 0 \pmod{N},$$

so N divides the product $(a^{r/2}-1)(a^{r/2}+1)$. If N divides neither fully, then it must divide partly into each of them, so computing the greatest common divisor will obtain a factor of N. Step 3 in the algorithm (apply QFT to find the order) works because the function $f : \mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ defined by $f(x) = a^x \pmod N$ is both periodic with period r, and one-to-one on each period. Technically speaking $\mathbb{Z}$ is infinite, and we don't have infinite resources, so there has to be some cutoff to truncate it to a finite size. We will not go deep into this step, but it can be solved.

How does this affect CSIDH? Thankfully, not in any meaningful way. Shor's algorithm works, because there is an efficient way to represent / evaluate the function call. The scenarios, to which Shor's algorithm applies, have in common that they have a cyclic group G acting internally on itself (either by group multiplication for elliptic curves, or exponentiation for RSA). This allows the quantum computer to prepare the uniform superposition of the group elements as

$$\frac{1}{\sqrt{|G|}} \sum_{g \in G} |g\rangle \,,$$

and evaluate the one-to-one and periodic homomorphism $f : G \to X$, whose kernel (or period) corresponds to a hidden subgroup.

Problem 4.2.11 (Hidden Subgroup Problem)**.** Let G be a group, $H \leq G$ a subgroup, X a set, and $f : G \to X$ a function such that $f(g_1) = f(g_2)$ if and only if $g_1 H = g_2 H$. Find the hidden subgroup H.

If an algorithm's security can be reduced to this problem, then Shor's algorithm is applicable. In contrast, CSIDH operates via a free and transitive group action $\mathrm{cl}(\mathcal{O}) \times \mathit{Ell}_p(\mathcal{O}, \pi) \to \mathit{Ell}_p(\mathcal{O}, \pi)$ (see Section 2.7). Since the target is just a set without any group structure, so the computational domain of the group action is disconnected from the register inputs. The isogeny computations still need to be evaluated in order to go from one curve to the other, so there is no exponential speedup available. The security of CSIDH can be, however, reduced to a related problem, which is discussed later in section 4.2.5.

4.2.4 Grover's Algorithm

Arguably the second most influential quantum algorithm is Lov Grover's algorithm [40]. It concerns the following computational problem.

Problem 4.2.12 (Unstructured search problem)**.** Let $f : \mathbb{Z}/N\mathbb{Z} \to \{\,0, 1\,\}$ be a function with the promise that $f(x_0) = 1$ for a unique "marked" element x_0. Find x_0.

Classically, this problem in hopeless, the only thing we can do is start searching one by one, so the worst possible running time is N. However, in the quantum setting, it is possible to find this element in just $O(\sqrt{N})$ queries, by computing the following algorithm.

Algorithm 7 Pseudo-code for Grover's algorithm, where $N = 2^n$

1: Prepare $|0^n\rangle$.
2: Apply $H^{\otimes n}$.
3: **for** i in $1 \leq i \leq O(\sqrt{N})$ **do**
4: Apply U_f.
5: Apply $D := -H^{\otimes n} U_0 H^{\otimes n}$, where U_0 maps $|0^n\rangle$ to $-|0^n\rangle$, leaving others invariant.
6: **end for**
7: Measure all qubits and output the result.

Theorem 4.2.13. *Algorithm 7 correctly returns the marked element x_0.*

Proof. The first two lines create the state

$$|\psi\rangle = H^{\otimes n} |0^n\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \,.$$

We can rewrite this in another orthonormal basis, namely the state vector spanned by $|x_0\rangle$ and its orthogonal complement $|x_0^\perp\rangle = \frac{1}{\sqrt{N-1}} \sum_{x=x_0} |x\rangle$. We have then

$$|\psi\rangle = \frac{1}{\sqrt{N}} |x_0\rangle + \frac{\sqrt{N-1}}{\sqrt{N}} |x_0^\perp\rangle .$$

Denote $\sin\theta = 1/\sqrt{N}$ and $\cos\theta = \sqrt{N-1}/\sqrt{N}$. In step 3, we can denote $U_0 = I - 2|0^n\rangle\langle 0^n|$, so that $D = -(I - 2|\psi\rangle\langle\psi|)$. Then in step 3, the first U_f maps $|\psi\rangle$ to

$$-\sin\theta |x_0\rangle + \cos\theta |x_0^\perp\rangle .$$

Applying D to this state vector yields

$$\begin{aligned} DU_f |\psi\rangle &= \sin\theta(I - 2|\psi\rangle\langle\psi|)|x_0\rangle - \cos\theta(I - 2|\psi\rangle\langle\psi|)|x_0^\perp\rangle \\ &= \sin\theta \left(|x_0\rangle - 2\sin\theta(\sin\theta|x_0\rangle + \cos\theta|x_0^\perp\rangle)\right) - \cos\theta \left(|x_0^\perp\rangle - 2\cos\theta(\sin\theta|x_0\rangle + \cos\theta|x_0^\perp\rangle)\right) \\ &= \sin(3\theta)|x_0\rangle + \cos(3\theta)|x_0^\perp\rangle . \end{aligned}$$

By the definition of U_f, it is a reflection, and so is D since

$$\begin{aligned} D^2 &= (I - 2|\psi\rangle\langle\psi|)^2 \\ &= I - 4|\psi\rangle\langle\psi| + 4(|\psi\rangle\langle\psi|)(|\psi\rangle\langle\psi|) \\ &= I - 4|\psi\rangle\langle\psi| + 4|\psi\rangle \cdot 1 \cdot \langle\psi| \\ &= I. \end{aligned}$$

Then DU_f is the product of two reflections in the plane spanned by x_0 and its complement, so it is a rotation of angle 2θ. This was one step of step 3. Performing T more times results in

$$\sin((2T+1)\theta)|x_0\rangle + \cos((2T+1)\theta)|x_0^\perp\rangle .$$

Since we defined $\sin\theta = 1/\sqrt{N}$, we have the classic physics-style approximation of small angles, so $\theta \approx 1/\sqrt{N}$. We are in search for x_0, so we would like to maximize the probability of measuring x_0. This means that we want to make $\sin((2T+1)\theta)$ close to 1 as much as possible. A simple algebraic manipulation yields the estimate

$$T \approx \frac{\pi\sqrt{N}}{4} - \frac{1}{2},$$

which is indeed $O(\sqrt{N})$, as required. □

This surprising fact is the core mechanism the *claw-finding algorithm*. The problem related to it is the following.

Problem 4.2.14 (Claw finding)**.** [71] Given two functions $f : X \to Z$ and $g : Y \to Z$, find at least one tuple $(x, y) \in X \times Y$ such that $f(x) = g(y)$.

The intrigued reader can find the actual algorithm is detailed in Tani's article [71]. The query complexity of this quantum version of a Meet-in-the-Middle attack is $O(\sqrt[6]{p})$ for CSIDH, because one can define the search space into a classical $O(\sqrt[6]{p})$ target part and a $O(\sqrt[3]{p})$ search part, on which the quantum collision search can be run. This might seem alarming, but in light of the next section, the prime instantiation already handles this. As per the NIST definitions, the number of quantum queries to an AES-128 oracle needed to break it with Grover's algorithm needs about 2^{64} quantum queries. In light of CSIDH, this would require to set $p > 2^{6 \cdot 64} = 2^{384}$, but the smallest prime that CSIDH uses has 512-bits. The next section details why this is necessary.

4.2.5 Kuperberg's algorithm and Peikert's attack

Even though Shor's algorithm is not applicable, since the CSIDH protocol cannot be reduced to the Hidden Subgroup Problem 4.2.11, the commutative nature of the underlying ideal-class group $\mathrm{cl}(\mathcal{O})$ can definitely be exploited in a quantum setting. Kuperberg [46] has come up with an algorithm to solve the related Dihedral Hidden Subgroup Problem (DHSP) with time, query and space complexity $2^{O(\sqrt{\log N})}$.

Problem 4.2.15 (Dihedral Hidden Subgroup Problem)**.** Suppose $D_{2n} = \langle r, s \mid r^n = s^2 = 1, srs = r^{-1} \rangle$ is the dihedral group, S is a finite set, and $f : D_{2n} \to S$ is some function. Let $H \subseteq D_{2n}$ be a subgroup, such that $f(a) = f(b)$ if and only if a and b are in the same (right) coset of H. Finding the hidden subgroup H is the *dihedral hidden subgroup problem.*

A more direct and applicable problem for CSIDH, however, is called the Hidden Shift Problem.

Problem 4.2.16 ((Abelian) Hidden Shift Problem)**.** [46, pp. 179] Suppose A is an abelian group, S is a finite set and $f, g : A \to S$ are two injective functions that differ by a shift:

$$f(a) = g(a + s).$$

The task of finding s from f and g is the *abelian hidden shift problem.*

Translating this problem statement into our CSIDH language is simple [17, pp. 21]. We have that the functions f, g are defined by

$$\begin{aligned} f, g : \mathrm{cl}(\mathcal{O}) &\longrightarrow Ell_p(\mathcal{O}, \pi) \\ f : [x] &\longmapsto [x]E_0, \\ g : [x] &\longmapsto [x]E_{Alice} = [x][\mathfrak{a}]E_0. \end{aligned}$$

Clearly, we have that

$$g([\mathfrak{b}]) = [\mathfrak{b}][\mathfrak{a}]E_0 = f([\mathfrak{b}][\mathfrak{a}]),$$

where $[\mathfrak{a}]$ is the hidden shift that we are trying to find. The problem statement is using additive notation, but for our group $\mathrm{cl}(\mathcal{O})$, the group action is denoted multiplicatively.

In his paper, Kuperberg also shows that the hidden shift problem is equivalent to solving DHSP. First, we state a lemma that will simplify the previous statement.

Lemma 4.2.17 (Hidden Reflection Problem)**.** *Finding an arbitrary hidden subgroup H of D_{2n} reduces to finding the slope k of a hidden reflection sr^k.*

Proof. If H is not a reflection, then either it is the trivial group or it has a non-trivial intersection with the cyclic subgroup $\mathbb{Z}_n$ generated by $\langle r \rangle$. Let $H' = \mathbb{Z}_n \cap H$ be the intersection of the hidden subgroup and the cyclic rotational subgroup. Finding the hidden subgroup H' in $\mathbb{Z}_n$ is easy if we know the factors of N, and we can factor N using Shor's algorithm (c.f. Section 4.2.3). Then the quotient group H/H' is either trivial, or a reflection in the quotient group G/H'. □

For the statement of the proposition, we will need the following definition.

Definition 4.2.18. Let $\exp(A)$ denote the multiplicative form of the abelian group A and C_2 be the multiplicative cyclic group of order 2. The *generalized dihedral group* is the semidirect product

$$D_A \cong C_2 \ltimes \exp(A),$$

with the conjugation relation

$$x^{-1} = yxy$$

for all $x \in \exp(A)$ and for the nontrivial $y \in C_2$.

Proposition 4.2.19. *If A is an abelian group, the hidden shift problem in A is equivalent to the hidden reflection problem in D_A.*

Proof. If $a \in A$, let x^a denote the corresponding element in $\exp(A)$. Given $f, g : A \to S$, define

$$f(a) = h(x^a), \quad g(a) = h(yx^a).$$

Then we clearly have that

$$f(a) = g(a+s) \iff h(x^a) = h(yx^{a+s}).$$ □

As a result, Kuperberg's algorithm, which is given for the DHSP problem, can be directly applied to CSIDH, as it can be translated into a hidden shift problem. This is exactly what Childs-Jao-Soukharev [22] did in the ordinary elliptic curve case, and what Biasse-Jao-Sankar [6] did in the supersingular case.

Since then, Peikert [56] has improved on Kuperberg's algorithm in 2020, which has resulted in a substantial quantum blow to CSIDH. The inner workings of Peikert's quantum algorithm is outside of the scope of this paper, but it is based on Kuperberg's algorithm, which Galbraith [37] summarizes at a high-level in the following way. Kuperberg's algorithm contains 3 major parts:

1. A quantum circuit to compute the class group action $[\mathfrak{a}] * E$ for any ideal class, which is then used to create certain "labeled" quantum states (consisting of an integer, called the "label", stored classically as bits, and a quantum state).

2. A sieve algorithm to produce "good" labeled states from "bad" labeled states.

3. A measurement process that allows to compute some bits of the secret ideal class from a nice labeled state.

Immediately, the real-world challenges lie in minimizing the required quantum circuit to do step 1, and minimizing the size of the sieve to increase efficiency. Peikert's improvements target the second and third parts of Kuperberg's algorithm. As written by Peikert himself [56, §3], the main differences between Kuperberg's and his algorithm can be summarized as the following.

1. "The sieve creates phase vectors with multipliers in progressively smaller intervals of the integers, by collimating on the "most-significant bits" of the multipliers. (By contrast, Kuperberg makes the multipliers divisible by progressively larger powers of two, by collimating on the least-significant bits.)"

2. "After sieving down to an interval of size S, where S can be roughly as large as the amount of quantumly accessible classical memory (QRACM), the algorithm applies a quantum Fourier transform of dimension S and measures, to reveal about $\log(S)$ of the "most-significant bits" of the secret with good probability. (Kuperberg instead applies a two-dimensional Fourier transform and measures to recover the single least-significant bit of the secret, with certainty.)"

3. "Alternatively, instead of recovering just $\log(S)$ bits of the secret, the algorithm can perform additional independent sieves down to various "scaled" intervals. By combining the resulting phase vectors, the algorithm can recover about $\log(S)$ different secret bits per sieve, and in particular, it can recover the entire secret using about $\log(N)/\log(S) = \log_S(N)$ sieves. (Kuperberg's algorithm, after recovering the least-significant bit of the secret, effectively halves the secret and repeats to recover the remaining bits, using $\log(N)$ total sieves.)"

Bonnetain and Schrottenloher [8] have also made estimates of the quantum safety of CSIDH. In their paper, they estimated breaking CSIDH-512 takes about $2^{32.5}$ queries, together with 2^{31} quantum bits of memory. However, Peikert's new analysis greatly improves on this estimate by estimating that 2^{16} oracle queries with 2^{40} bits of QRACM (not quantum memory!) is enough. Selected results are presented in Table 4.1.

Peikert based his T-gate complexity estimates on the CSIDH authors second paper [4], where they estimated that the CSIDH-512 oracle can be implemented in about 2^{40}-2^{44} T-gates, which is then multiplied by the estimated number of queries from the relevant row of Table 4.1. Based on the MAXDEPTH restriction of the NIST competition 4.2.1, Level 1 security corresponds to the difficulty of breaking AES-128. NIST's estimate for that is 2^{170}/MAXDEPTH quantum gates, where MAXDEPTH is between 2^{40}

CSIDH-log p	log N	log QRACM	depth	log $\tilde{Q}_{\text{total}}$	log $T \leq$
512	257.1	32	11	18.7	63
		36	9	17.0	61
		40	8	15.7	60
		44	7	14.9	59
		48	6	14.1	58
1024	512	36	19	27.9	76
		40	16	25.5	74
		44	14	23.5	72
		48	13	22.1	70
		52	12	20.8	69
1792	896	40	29	39.2	90
		44	25	35.8	87
		48	23	33.2	84
		52	21	30.9	82
		56	19	29.2	80

Table 4.1: Peikert's complexity estimates [56, Figure 1]. Empty cells are equal to the one above them. The columns from left to right are: CSIDH implementation; estimated size of the class group $\mathrm{cl}(\mathcal{O})$; number of bits of QRACM; depth of the sieve's recursion tree; total number of queries to the quantum oracle to recover all but 56 bits of the secret; total estimated T-gates.

and 2^{96}. Even for the weakest assumption, CSIDH-512 is significantly below the required 2^{74} gates to qualify for Level 1. Furthemore, even CSIDH-1024 is below (or barely on the edge) of Level 1, resulting in a substantial blow to the presumed quantum security of CSIDH (and CSURF as well).

The authors of CSIDH's response to the Bonnetain-Schrottenloher [4] recount a couple of refutations for the estimates [17, pp. 23]. They argue that that the Bonnetain-Schrottenloher paper ignores most of the costs of evaluating isogenies, for example Algorithm **??** uses a lot of input-dependent branches, which is impossible in superposition, and that they skip finding the kernel generator points for the ℓ_i-isogeny. The CSIDH team also states [4, §1.3] that they have analyzed the cost of the isogeny-evaluation quantum oracle, and conclude that this paper does not fall below NIST level 1. Another criticism, this time from Steven Galbraith [37], regarding this discussion is that it is quite unclear whether a quantum computer with 2^{60} gates is a realistic object at all. It is worth noting that as of 2026, the processor with the most MOSFETs is the AI processor Cerebras WSE-2 [42] with 2.5 trillion transistors ($\sim 2^{31}$).

Chapter 5

Recent improvements of CSIDH

Now that we understand the inner workings of CSIDH, and what mathematical background goes into both its decisions for the programmatic implementation, and the ways to break it, we can begin pondering about potential ways to improve on CSIDH. The dear reader might have already gotten some ideas regarding potential improvements within the setup of the algorithm or to speed up the computations, for example: can we do better than bounding the number of isogeny steps between $[-5, \ldots, 5]$? In general, we can categorize the types of improvements into two big categories:

Easier: These improvements appear as immediate questions even to the beginner reader. These improvements can be implemented and tested relatively fast and easily. Some examples are the number of isogeny steps and its distribution, achieving constant-time implementation using dummy variables and swapping out functions for calculations ($\sqrt{\text{élu}}$ instead of Renes).

Harder: These improvements are non-trivial, they require developing new techniques, and potentially extensive cryptanalysis. Some examples are mitigating Peikert's quantum attack (Section 5.1), developing a dummy-free constant-time implementations (Section 5.2) or trying other curves (Section 5.3).

In the following sections, we explore instances of the harder improvements, and detailing the ideas for the improvements, and the empirical evidences backing up these claims. The general flow of these section is stating the problem in CSIDH, then giving some detail about how the respective implementation addresses this issue, then their contributions are shown by their proposed instantiations, together with performance test results to give more credence to their claims. Of course, these hard improvements also feature a lot of the easier improvements, but the history of those is rather summarized in the next paragraph, as their ideas are incorporated into the following sections' discussion.

Definition 5.0.1. [1, Section 2.4] Let A be a randomized algorithm that takes an input $i \in I$. Since A is randomized, the running time of A varies according to some distribution. If this running time distribution is the same for any two inputs $i, i' \in I$, then we say that A is **constant-time**.

After the publication of CSIDH, some of the authors came together to analyze the quantum evaluation of isogenies and their cost, which also happens to be the first constant-time implementation of CSIDH [4]. They used dummy isogenies to secure constant-time implementation. The dangers of that is discussed later in this section. Later on, Meyer, Campos and Reith [49] have improved on the efficiency of the previous implementation, with the additional introduction of some tricks for algorithms. One example is called the SIMBA technique (short for Splitting Isogenies into Multiple Batches), which is similar to the batching process of CTIDH that is shown in Section 5.2. Another change of their setup is that the secret key $e = (e_1, \ldots, e_n)$ is taken from $e_i \in [0, \ldots, 5]$ instead of the full range. In essence, their setup only takes steps in one direction.

The next major step towards better performance was proposed by Onuki, Aikawa, Yamazaki and Takagi [55], where they utilize two points P, Q for the group action evaluation instead of one, where $P \in E(\mathbb{F}_p)$ and $Q \in E(\mathbb{F}_{p^2})$ with $x(Q) \in \mathbb{F}_p$. Their 2-point algorithm runs with m_i iterations, and each iteration is only 1/3 as expensive and the previous one.

There have been many other papers with substantial theoretical and low-level improvements. One example is a more robust framework against physical attacks [18], where the isogenies are constructed e_i times, and the rest of the $m_i - |e_i|$ isogenies are alternatingly evaluated. Another example being able to adapt SIKE's optimal strategies to CSIDH [21, 43].

There are some immediate problems that are evident from the discussions in the Quantum Security Section 4.2. The fact that CSIDH is susceptible to the sub-exponential quantum attack is already a significant drawback when comparing CSIDH to other PQC protocols. Even though the CSIDH authors debate the practical relevance of Peikert's or Bonnetain and Schrottenloher's findings, clearly it is still the most important argument against CSIDH as it stands. This is going to be addressed in Section 5.1, where the authors of SQALE [**sqale**] increase the size of the prime p significantly, while finding many improvements in other areas that end up resulting in a dummy-free constant-time implementation, resulting in a very solid improvement with (proposed) security corresponding to NIST Level 1, 2 and 3.

The second most important shortcoming of the CSIDH paper (which is also mentioned by them in their paper [17, pp. 29]) is the fact that its implementation is *not side-channel resistant*, and in particular not constant-time. As per the discussion in Section 2.3, the cost of computing an ℓ-isogeny is linear in ℓ, meaning that the bigger the prime ℓ is, the more computation needs to be performed. Furthermore, the actual implementation of finding the kernel for the Vélu isogeny-calculation is done by randomly sampling a point on the curve, which is then multiplied enough times to make the new point have order ℓ. However, this algorithm can return the point at infinity, which causes the algorithm to start over. The probability of this happening is $1/\ell$, since out of the ℓ points, only one of them is ∞. These two facts can then be used to gain information about the secret key, because (1) measuring the time of taking one isogeny step is longer if that step is for a bigger isogeny, and (2) if the isogeny calculation fails at some rate $1/r$, then an attacker can assume that the specific isogeny step is most likely an r-isogeny. These attacks are some examples of non-software attacks, collectively called as **side-channel attacks.**

Both of these cases highlight the importance of constant-time algorithms, but the method of which one addresses these also matters. One of these approaches is done by using *dummy isogenies*. This means that there is a fix amount of isogenies that are computed in every group action evaluation step, but the ones that are not actually needed are discarded. This solves the problem of determining how many isogenies are used in the private key, but if this number is bounded by $|e_i| \leq m$ (in CSIDH, this is $m = 5$), then we have that $m - |e_i|$ of these are calculation done "in vain", heavily impacting the performance of the overall protocol. A more efficient constant-time implementation using as little dummy isogenies as possible by introducing a new way of defining the key space is tackled in Section 5.2.

Another drawback of dummy operations is due to an attack called *safe-error attack*, which is a type of fault-injection. Fault injections are basically physical hardware manipulations of the hardware that is running an algorithm. If such an injection faults an operation, but then the output remains unchanged, we can conclude that this operation was indeed a dummy operation. Fault-injection attacks against CSIDH are discussed in [12]. The SQALE team addressed this issue natively in their setup as discussed in Section 5.1, and the in 2025, a dummy-free version of CTIDH has been developed in [2] to efficiently mitigate these risks.

Lastly, as discussed in Section 3.1.5, we are on the floor in the isogeny graph of CSIDH. This in itself is not a security vulnerability, but it is natural to ask whether going to the surface is worth doing. One reason for considering this is the fact that because of how the mathematics work on the floor (see Section 2.6), we do not have the possibility of using 2-isogenies, which is possible on the surface. The actual mathematics and algorithms are discussed in Section 5.3.

5.1 SQALE

As mentioned before many times, Peikert's attack greatly lowers the assumed quantum security of CSIDH as it stands in its original form. The quantum security of the Abelian Hidden Shift Problem (see Problem 4.2.16 against Kuperberg's algorithm depends on the size of p only, so for CSIDH-512 to be viable at NIST Level 1, the size of the prime needs to increase to between 2260 and 5280 bits [8], at which point the algorithm becomes excruciatingly slow. This is the problem that the authors of SQALE; Chávez-Saab, Chi-Domínguez, Jaques and Rodríguez-Henríquez set out to solve.

SQALE [20] stands for "**S**ublinear Vélu **Q**uantum-resistant isogeny **A**ction with **L**ow **E**xponents". We will not go through this protocol letter by letter as before, but as the name suggests, it is using $\sqrt{\text{élu}}$ (see Section 2.3) for its isogeny step calculations. In particular, the number of steps taken is around $\tilde{O}(\sqrt{\ell})$ for an ℓ-isogeny, so using more larger primes than in the instantiation of CSIDH is not punished, and

might even end up being more efficient. The last part of the name is justified by the fact that in the SQALE instantiation, instead of the originally suggested bound of $e_i \in [-5, \ldots, 5]$, these exponents, i.e. the number of ℓ-isogenies are drawn only from $\{-1, 1\}$. The authors arrive at this minimal exponents by running experiments checking all bounds from $m = 5$ to $m = 1$, to check how many primes are needed to reach classical security of 128-bits and 192-bits, based on their own assessment of quantum costs. The reported results are in Table 5.1.

Bound m	**128-bits**		**192-bits**	
	CSIDH	SQALE	CSIDH	SQALE
5	64	86	89	119
4	70	95	97	132
3	79	111	109	153
2	95	139	132	193
1	139	221	193	306

Table 5.1: Number of small odd primes ℓ_i required for ensuring 128- and 192-bits of classical security given hardware bounds. [20, Table 3]

The prime p in SQALE is still in the same form as in CSIDH, namely

$$p = 4 \cdot \prod_{i=1}^{n} \ell_i - 1,$$

but if the number n of the first small primes ℓ_i is not raised, then this would only result in a massively reduced key space of 2^{74}, so naturally one might think that they require much more primes, which is partly true, but the surprising result is that the SQALE key generation uses only the $k \leq n$ smallest prime factors (since 0 steps are allowed). CSIDH used a heuristic bounding box argument for their choice of number of primes to be equal to the required number of vertices (the size of the class group) 2^{256}. On the other hand, the SQALE team chose to argue against the need to fill up the whole key space. This is similar to the idea of *lossy keys*, and they base their justification on the following.

Classically, attacks against large-prime CSIDH rely on Meet-in-the-Middle attack (Section 4.1.2) or on the vOW attack. The cost of these attacks is governed by the key space size N, not the class group size, so setting $N \sim 2^{2\lambda}$ guarantees λ-bits of classical security (or does it? [77]). For the quantum security, as stated in [20, §4], they believe that the small keyspace does not help a quantum adversary, because group elements still need to be represented as integers modulo $N = \mathrm{cl}(\mathcal{O})$, but low-exponent keys do not map to small integer multiples of a generator, so phase states still require $\sim \log(p)$ qubits. There are other, more technical reasons for this assumption, which can be read in [20, §4], but the the conclusion is that attackers still need to search the entire class group, and that the quantum oracle used in the attack still needs to evaluate arbitrary group actions, not just small exponents.

Based on these justifications, after running their own simulations/experiments, they suggest using CSIDH-4096 for NIST Level 1, CSIDH-6144 for NIST Level 2 and CSIDH-8192 for NIST Level 3, where CSIDH-n denotes the bit length of the prime p. They have applied the above mentioned optimizations to the three implementations of CSIDH mentioned in the previous section: the MCR-style [49], OAYT-style [55] and the Dummy-free style [18]. The corresponding performance in gigacycles is presented in Table 5.2, and the full list of suggested small primes for a given $\log(p)$ size is in Table 5.3.

To sum up, what the authors of SQALE have done is adopting the CSIDH framework to resist Peikert's refined Kuperberg-style attack, since previous CSIDH-4096 was taking literal minutes run. They achieved this by applying the improved $\sqrt{\text{élu}}$ formula to speed up the larger ℓ-isogeny calculations. They have also given a solid theoretical background to justify using very small exponents (and thus smaller private keys), since any quantum attacker would still need to attack the entire class group, regardless of the size of the exponents of the isogenies. One important side note is that it is still important to ensure that the selected small primes ℓ_i actually generate the class group, and it does not form its own subgroup.

Instantiation	**Implementation Style (gigacycles)**		
CSIDH-$\log(p)$	OAYT	MCR	Dummy-free
CSIDH-4096	23.21	28.50	39.35
CSIDH-6144	74.88	87.09	117.57
CSIDH-8192	199.15	236.13	322.57

Table 5.2: Clock cycles given in gigacycles of the suggested SQALE instantiations, broken down by style implemented. The measured clock cycles are the average of 1024 random instances without key validation. [20, Table 6]

$\log(p)$	n	Excluded	Included
1024	130	739	983
1792	207	149	1289
2048	231	5	3413
3072	326	37, 2053	2203, 2007
4096	417	1151	2897
5120	504	5	4133
6144	590	71, 4289	4337, 4339
8192	757	4937, 5749	5783, 5791
9216	838	263, 6373	6473, 6481

Table 5.3: Shape of the primes in the instantiation, where n denotes the number of the first odd prime numbers, excluding and including the primes present in the respective columns. [20, Table 5]

5.2 CTIDH

As mentioned before, the original CSIDH is not constant-time, so it is susceptible to side-channel attacks. Even though there existed constant-time implementations mentioned at the beginning of this chapter, they were not particular fast / near the speed of CSIDH, understandably so. The team of CTIDH [1]; Banegas, Bernstein, Campos, Chou, Lange, Meyer, Smith and Sotáková, have come up with a new way of solving the problem of efficient constant-time implementation. They introduce a new key space together with a new algorithm for evaluating the group action, which produces speed records for any previous constant-time CSIDH implementations.

CTIDH also utilizes the $\sqrt{\text{élu}}$ formula, but adaptively. The estimated ℓ for which the $\sqrt{\text{élu}}$ formula takes over the normal Vélu is around 100 [3, Appendix A], and the CTIDH paper states that it is around 89. This is useful for the speed-up of course, but the real issue was trying to find a way to mask the time it takes to compute ℓ-isogenies. One aspect of optimizing this is targeting the fact that there is no efficient way to deterministically generate points of order ℓ in $E(\mathbb{F}_p)$, so the only way is how CSIDH does it. Recall that in CSIDH, a random point T is sampled from $E(\mathbb{F}_p)$, and then $P = [(p+1)/\ell]T$ will have order ℓ with probability $1 - 1/\ell$ (since it could also be the point at infinity). This innate randomness can be taken advantage of, so there are some strategies to mitigate this. CSIDH uses a multiplicative strategy, where the isogenies are setup in a row, and the points are pushed through the isogeny one-by-one (cf. Section 3.2.3), so it minimizes the possibility of failure, and also avoids the huge cost of scalar multiplication by a near-512 bit number. Another aspect is not giving away the private key when an attacker is counting the time it takes to apply the ℓ-isogeny. CSIDH itself had no safeguards against this, but subsequent constant-time implementations were attempted by either using dummy isogenies, or pushing more points through the isogeny, as discussed before.

The approach CTIDH is taking to mask the time of isogeny evaluations is quite different, which is done by *batching* both the primes and the isogenies, which is basically the same thing as partitioning, and setting an L^1-norm bound,

Definition 5.2.1 (Prime batching)**.** [1, §3] Let $\ell_1, \ldots, \ell_n$ be a sequence of primes, $0 < B \leq n$ the number of batches and $N = (N_1, \ldots, N_B) \in \mathbb{Z}^B_{>0}$ be a vector with $\sum_{i=1}^{B} N_i = n$. The **prime batches** are the

subsequences of the consecutive primes, labeled as

$$\begin{aligned}(\ell_{1,1},\ldots,\ell_{1,N_1}) &:= (\ell_1,\ldots,\ell_{N_1}),\\ (\ell_{2,1},\ldots,\ell_{2,N_2}) &:= (\ell_{N_1+1},\ldots,\ell_{N_1+N_2}),\\ &\vdots\\ (\ell_{B,1},\ldots,\ell_{B,N_B}) &:= (\ell_{n-N_B+1},\ldots,\ell_n).\end{aligned}$$

Definition 5.2.2 (Isogeny batching)**.** [1, §3] Let $B_i = (\ell_{i,1},\ldots,\ell_{i,N_i})$ be the i'th batch of primes. **Isogeny batching** is sampling a prime $\ell_{i,j}$ from the batch B_i at random m_i times. In other words, there is a fixed bound on the whole batch given by $m_i \geq \sum_{j=1}^{N_i} |e_{i,j}|$, not just for singular $\ell_{i,j}$-isogenies.

Definition 5.2.3. [1, §3] Let $N \in \mathbb{Z}_{>0}^B$ and $m \in \mathbb{Z}_{\geq 0}^B$. Then we define the **key space** $\mathcal{K}_{N,m}$ as

$$\mathcal{K}_{N,m} := \{(e_1,\ldots,e_n) \in \mathbb{Z}^n \mid \sum_{j=1}^{N_i} |e_{i,j}| \leq m_i \text{ for } 1 \leq i \leq B\}.$$

Lemma 5.2.4. *[1, Lemma 1.] The size of $\mathcal{K}_{N,m}$ can be determined by*

$$\#\mathcal{K}_{N,m} = \prod_{i=1}^{B} \Phi(N_i, m_i), \text{ where } \Phi(x,y) = \sum_{k=0} \binom{x}{k} 2^k \binom{y}{k}$$

is a function that counts the vectors in $\mathbb{Z}^x$ with 1-norm at most y.

Proof. This follows from an easy counting argument. The number of nonzero entries in the x positions is k, so there are $\binom{x}{k}$ ways of permuting them without order. Each nonzero entry is either positive or negative for the two directions on the isogeny graph, hence the 2^k part. Lastly, the vector of partial sums over these k nonzero entries has k different integers in the range $[1, y]$, and these vectors uniquely match one assignment of partial sums. There are $\binom{y}{k}$ ways to pick k different integers in the range $[1, y]$. The size of the full key space is the product of the size of the batches, proving the lemma. □

This batching process gives a larger key space per isogeny computed, because combinatorially, there are more possibilities to choose isogenies from the batches. It's important to note that in a constant-time algorithm, when a batch is chosen, all the isogenies are calculated inside the batch, not just one of them picked randomly, therefore a bigger key space might not necessarily convert to better performance. Let us demonstrate this and the above definition with some examples, and we will also compute the isogenies per key bit (i.e. the base-2 logarithm of the key space size) to illustrate the impact of running every isogeny inside the batch.

Example 5.2.5. Let $\ell_1,\ldots,\ell_n$ be a sequence of primes. Let us see what happens when we choose different B's for batches.

B = n : If $B = n$, then we have only one choice for dividing the primes up, which is $N = (1,\ldots,1)$.

B = 1 : In this case, our choices are not really choices, and we have $N = (n)$. This essentially amounts to choosing $m_i = m$ isogenies at most randomly from the full list of small primes $\ell_1,\ldots,\ell_n$, which is the approach that [52] takes.

Example 5.2.6. Let $\ell_1,\ldots,\ell_{10}$ be a sequence of 10 primes, and let us allow a maximum of 10 isogenies, so the bounding vector for the exponents satisfies $\sum m_i = 10$. Let us count the size of the key space for different batches of primes and isogenies.

B = 10 : Because we are batching into 10 batches, and we're restricted to maximum 10 isogenies, this is equivalent to the key space of SQALE with allowing 0 as an exponent, since $e_i \in \{-1, 0, 1\}$. The size of the key space is thus 3^{10}, with key bit ~ 15.85.

B = 1 : We have one big batch, so we must have $N = 10$, and $m = 10$. Plugging these into the formula of Lemma 5.2.4 gives 8,097,453 keys, or ~ 22.95 key bits.

B = 3 : We have three batches, so let us choose a partition of $N = (3, 3, 4)$ and a similar partition of bounds $m = (3, 3, 4)$. Counting based on the formula yields 1274049 keys, or ~ 20.28 key bits.

In the first case, since every batch contains 1 isogeny, running 10 batches gives us 10 isogeny calculations. In the second case, we run the giant batch 10 times, but since it has 10 isogenies inside, this results in 100 isogeny calculations. The third case uses $9 + 9 + 16 = 34$ isogeny calculations. If we calculate the isogenies per key bit, we get 0.63, 4.36 and 1.68 respectively.

As it can be seen from the previous example, batching can have a big improvement on the number of isogenies evaluated compared to the key space size in bits, which has been optimized by the CTIDH authors to strike a balance between ℓ-isogeny calculations and generating points of order ℓ.

The next hurdle in balancing the performance-security tradeoff is finding a way to evaluation an $\ell_{i,j}$-isogeny in batch B_i in constant time, i.e. without revealing which isogeny was chosen. To this end, the CTIDH paper defines *restricted square-free atomic blocks*, which is an abstract way of thinking about the evaluation process, and making it easier to prove whether an algorithm is constant-time. It essentially processes one secretly selected prime per active batch, reaching the goal of masking the choice of isogeny.

The whole algorithm is quite complicated (see [1, Algorithm 4.], but the most important parts that constitute solutions to the challenges mentioned before are worth highlighting. One subroutine in the algorithm is called `PointAccept`, which is there to mask the different failure probabilities of finding points of different orders ℓ. This subroutine uses the following construction.

Lemma 5.2.7. *Let $\ell_{i,j}$ be the j'th prime in batch i. The probability of failure for generating a kernel generator P of order $\ell_{i,j}$ can be masked by accepting a coin toss with success probability*

$$\gamma = \frac{\ell_{i,j} \cdot (\ell_{i,1} - 1)}{\ell_{i,1} \cdot (\ell_{i,j} - 1)}.$$

Proof. The probability of failure for P is $1/\ell_{i,j}$, which depends on j only within the batch. We can accept the point P as a valid kernel generator if both $P \neq \infty$ and the coin toss is successful. The probability of acceptance is then

$$\begin{aligned}
\mathbb{P}(\text{Accept}) &= (1 - 1/\ell_{i,j}) \cdot \frac{\ell_{i,j} \cdot (\ell_{i,1} - 1)}{\ell_{i,1} \cdot (\ell_{i,j} - 1)} \\
&= \frac{\ell_{i,j} - 1}{\ell_{i,j}} \cdot \frac{\ell_{i,j} \cdot (\ell_{i,1} - 1)}{\ell_{i,1} \cdot (\ell_{i,j} - 1)} \\
&= (1 - 1/\ell_{i,1}).
\end{aligned}$$

Therefore the probability of acceptance is based on the smallest prime in the batch, independent of the chosen j'th prime. □

In practice, this γ-weighted cointoss is implemented by generating a uniform random 256-bit integer, which is reduced modulo $\ell_{i,1}(\ell_{i,j} - 1)$, then a simple check is done whether the reduction is smaller than $\ell_{i,j}(\ell_{i,1} - 1)$, all in constant time.

The other important subroutine is called `MatryoshkaIsogeny`. This subroutine masks the computation of the ℓ-isogeny by transforming it into a computation independent of j inside the batch. When the algorithm uses Vélu's formula, the first step is calculating the product

$$\prod_{s \in S} (X - x([s]P))$$

for some index set S. In our case, this set is $\{1, \ldots, (\ell-1)/2\}$, but then this set contains the corresponding integers of all lower degree isogenies $\ell_i \leq \ell$. Thus we can always compute the largest primes' Vélu kernel, and actually use the kernels of the smaller primes for their calculations. This property is like a Matryoshka doll, hence the name. When the algorithm uses $\sqrt{\text{élu}}$, then the sets U and V are defined as the optimal choices for the smallest degree in the batch $\ell_{i,1}$, and the leftover set W is chosen to make the formulae work even for the largest prime in the batch ℓ_{i,N_i}. The standard baby-step giant-step algorithm for $\sqrt{\text{élu}}$ stays the same, but during the iteration over W, the intermediate results are saved which correspond to the degrees $\ell_{i,j}$. This results in a constant time of evaluation, after which we can select the result for

the desired j index. When the primes are close together in the batches, this modified formula is not too much worse than the original formulae.

Now that the theory is solid, the next step is to verify this practically. In order to find the optimal batching vector together with the best bounds for them, the CTIDH team had to run an extensive optimization. Since the key space is much more complicated compared to the standard CSIDH key space, optimizing has a lot more variables. At the top, the number of batches B is chosen, and that controls the search space for the partition of n into B parts. The different batch sizes N_i influence the success rate of accepting an atomic block calculation, because `PointAccept` depends on the smallest prime in the batch, as seen in Lemma 5.2.7. The other part is minimizing the cost from the L^1-norm bounds $m = (m_1, \ldots, m_B)$, and by Lemma 5.2.4, the size of the key space depends on both N and m, which is the third part that needs minimizing, since the key space also needs to be around $2^{2\lambda}$ for security level λ. Their cost function is detailed in [1, §6], and the results of the searches are summarized in [1, Table 1, pp. 17]. The best performing batch size turned out to be $B = 14$, so this was chosen as their basis for the instantiation for CTIDH-512. The empirically discovered sizes of batches N_i, together with the small primes in the batches $\ell_{i,j}$ and their L^1-norm bounds m_i are recorded in Table 5.4. The approximate key space size for this instantiation is approximately $2^{256.009}$, which is very close to the actual class group, which has ~ 257.137-bits [5].

Batch N_i	**List of primes in the batch**	**Bound** m_i
N_1	3, 5	10
N_2	7, 11, 13	14
N_3	17, 19, 23, 29	16
N_4	31, 37, 41, 43	17
N_5	47, 53, 59, 61, 67	17
N_6	71, 73, 79, 83, 89	17
N_7	97, 101, 103, 107, 109, 113	18
N_8	127, 131, 137, 139, 149, 151, 157	18
N_9	163, 167, 173, 179, 181, 191, 193	18
N_{10}	197, 199, 211, 223, 227, 229, 233, 239	18
N_{11}	241, 251, 257, 263, 269, 271, 277, 281	18
N_{12}	283, 293, 307, 311, 313, 317	13
N_{13}	331, 337, 347, 349, 353, 359, 367, 373	13
N_{14}	587	1

Table 5.4: Instantiation of CTIDH-512. List of small primes $\ell_1, \ldots, \ell_{74}$ are from CSIDH-512, separated into the specified number of batches and respective bounds. [1, Table 1]

Unsurprisingly, the CTIDH authors ran performance checks as well, and it turns out that their implementation is by far the fastest from available choices. The tests were performed on a 3GHz Intel Xeon E3-1220 v5 (Skylake) CPU, so that the speeds can be comparable to other CSIDH constant-time implementations using the Skylake architecture as well. Some selected results comparing the performance of CTIDH-512 vs. the performance of the constant-time instantiations of CSIDH-512 mentioned in the beginning of this chapter are shown in Table 5.5. As it can be seen, the speedup is roughly $\sim 40\%$ in number of million Skylake cycles, and CTIDH uses roughly 50% less multiplications than the best alternative. For reference, the performance of CSIDH for the full group action and public key validation is around 111 Mcycles, so CTIDH is almost as close as the original instantiation, while being a lot more side-channel resistant. Furthermore, the software implementation of CTIDH has been formally audited using Valgrind, which confirmed zero secret-data flow to conditional branches or memory array indices. This cements CTIDH as the state-of-the-art best option for a constant-time implementation of CSIDH-512.

For higher security primes of CTIDH, there have been more improvements as recent as 2025. The team of [10] have implemented the higher-security primes with a couple modifications, such as a modified

Instantiation	Mcyc	M	S	a	Total
CTIDH [1]	129.64	330966	121787	497476	452752
CR [21]	218.42	665876	189377	691231	855253
HLKA [43]	238.51	632444	209310	704576	841754
CCCDRS [18]	239.00	657000	210000	691000	867000
OAYT [55]	—	732966	243838	680801	976804
MCR [49]	395.00	1054000	410000	1053000	1464000

Table 5.5: Comparison of speeds of constant-time CSIDH-512 implementations with 256-bit private key. The columns 2-6 denote the following in order: Millions of Skylake cycles; Number of multiplications (excluding squaring); Number of squarings; Number of additions; Total cost calculated as $\mathbf{M} + \mathbf{S}$. [1, Table 2]

prime structure to help with verification, and low-level arithmetic optimization. This resulted in a 3-times speedup compared to their deterministic implementation (which improved on SQALE by a factor 2.5). See **??** for the results of their real-life testing. In mid-2025, Campos, Hellenbrand, Meyer and Reijnders [11] have combined the deterministic variant of CSIDH with the batching strategy of CTIDH from [10], and improved their efficiency by roughly 17%. They called their this new version dCTIDH-2048, where d stands for "deterministic". Then in January 2026, almost immediately after this improvement, Hellenbrand teamed up with Banegas and Saldanha [2] to improve dCTIDH from [11]. The resulting protocol, dubbed "Hardened CTIDH", did not surpass dCTIDH, but it still performed 4 times faster then the deterministic version of [10]. Hardened CSIDH is the first *efficient* implementation of CSIDH-like protocols that are (1) deterministic, (2) constant-time and (3) fully dummy-free at the same time.

5.3 CSURF

As mentioned in Subsection 3.2.2, the CSIDH protocol specifies the finite field prime p as $p = 4 \cdot 3 \cdot \ldots \ell_{74} - 1$, so we are in case 2(b) of Theorem 2.6.23, furthermore, since the starting curve has order $\mathbb{Z}[\sqrt{-p}]$ then CSIDH operates on the floor of the isogeny graph $G_{\mathbb{F}_p, L}$. We have discussed before a couple times that the Vélu-style formulae scale with the size of ℓ-isogeny, so it could potentially be a performance speedup if the degree-2 isogenies are included in the protocol. However, we see that from Theorem 2.6.23 again, that on the floor of the CSIDH prime, there are no $\mathbb{F}_p$-rational horizontal 2-isogenies. The only 2-isogenies existing in the structure of the supersingular isogeny graphs are the "sides" of the volcano, connecting the floor to the surface in a ratio of 3:1. Furthermore, because of this 3:1 ratio, it turns out even though the CSIDH graph's floor has 3 times more vertices, there is a 3-to-1 homomorphism that reduces the hardness of the parallelization problem (see Problem **??**) to the smaller class group on the surface, so this extra multiple of 3 does not increase offer additional security.

This gave birth to the idea of CSURF [14], developed by Wouter Castryck and Thomas Decru (the same cryptographers behind the fall of SIDH [15]). CSURF stands for "CSIDH on the Surface", so as the name suggests, they proposed to move the calculation to the surface, where 2-isogenies could be traversable horizontally. This is not possible in the CSIDH graph's surface as discussed in the previous paragraph, so the next natural version to consider is $p \equiv 7 \pmod 8$, which is case 2(a) in Theorem 2.6.23. In this setting, we still have three 2-isogenies, but now two of those are horizontal, since the prime 2 splits in the ring of integers as $(2) = \mathfrak{l}_0 \bar{\mathfrak{l}}_0$ where

$$\mathfrak{l}_0 = \left(2, \frac{\sqrt{-p} - 1}{2}\right) \quad \text{and} \quad \bar{\mathfrak{l}}_0 = \left(2, \frac{\sqrt{-p} + 1}{2}\right).$$

It turns out, that just like in the case of CSIDH there is a one-to-one relationship between the vertices of $Ell_p(\mathcal{O}, \pi)$ and the Montgomery coefficient A of $E_A : y^2 = x^3 + Ax^2 + x$, there is a corresponding curve model for which the vertices on the surface of the CSURF isogeny graph correspond to a simple integer as well.

Definition 5.3.1. Let $E(\mathbb{F}_p)$ be an elliptic curve. It is called a **Montgomery$^-$ curve** if it is the form

$$E_A^- : y^2 = x^3 + Ax^2 - x.$$

The coefficient A is called the **Montgomery$^-$ coefficient.**

To avoid confusing, we will denote the "normal" Montgomery$^+$ curves as E_A^+. Note that as opposed to the definition of the Montgomery curve in the CSIDH case, we do not have to restrict the possible coefficients, because the discriminant is 0 if and only if -2 is a quadratic residue, but since CSURF uses $p \equiv 7 \pmod 8$, we have that -2 is not a square in $\mathbb{F}_p$.

Proposition 5.3.2. *Let $p \equiv 7 \pmod 8$ be a prime number, let $E(\mathbb{F}_p)$ be a supersingular elliptic curve over $\mathbb{F}_p$. Then we have that $\mathrm{End}_p(E) \cong \mathcal{O}_K = \mathbb{Z}[(1+\sqrt{-p})/2]$ if and only if there exists a coefficient $A \in \mathbb{F}_p$ for which E is $\mathbb{F}_p$-isomorphic to the curve $E_A^- : y^2 = x^3 + Ax^2 - x$. If such an A exists, then it is also unique.*

Proof. See [14, Proposition 4.]. The idea of the proof is to establish a bijection between the Montgomery$^\pm$ curves, and then use the result from CSIDH (Theorem 2.7.16). □

The consequence of this is great, because now the CSURF curves can be represented by a single integer A, just like CSIDH, keeping one of the most important advantage of having a tiny public key. Furthermore, the familiar Renes' formulae (Theorem 2.3.5) used in CSIDH can be amended by a couple sign flips as well [14, Proposition 2.]. Alternatively, the standard Montgomery$^+$ curves can be used for odd isogeny calculations as well, provided that the point $(0,0)$ corresponds to the specific point $P_\rightarrow$.

Lemma 5.3.3. *Let $p \equiv 7 \pmod 8$ and let $E \in Ell_p(\mathcal{O}_K)$. The 3 points of order 2 can be classified as*

- *a point $P_\rightarrow$, whose four halves belong to $E(\mathbb{F}_p)$,*
- *a point $P_\leftarrow$, whose four halves have x-coordinates in $\mathbb{F}_p$ and y-coordinates outside $\mathbb{F}_p$,*
- *a point $P_\downarrow$, whose four halves have x-coordinates outside $\mathbb{F}_p$.*

Furthermore, we have that $E[\mathfrak{l}_0] = \langle P_\rightarrow \rangle$ and $E[\bar{\mathfrak{l}}_0] = \langle P_\leftarrow \rangle$, while $\langle P_\downarrow \rangle$ defines the kernel of the descending 2-isogeny.

Proof. See [13, Lemma 3.] and the addendum to it [14, Lemma 4]. □

Lemma 5.3.3 gives an easy way to evaluate 2-isogenies in chains, since if E' is the codomain of a 2-isogeny from E, then we get the point $\langle P_\leftarrow \rangle$ for free, since that is the generator of the dual isogeny $[\bar{\mathfrak{l}}_0]$, thus we are left with a quadratic polynomial instead of a cubic. Then finding the roots gives us two candidates, but in light of Lemma 5.3.3, it is enough to whether one of the candidates has a half with x-coordinates in F_p. The formula to compute 2-isogenies in chain is given in Algorithm 8 below. An important note is that the square root in the algorithm denotes the canonical square root, which means that $\sqrt{a}$ is the unique square root which is again a square, which is equal to $a^{(p+1)/4}$ in $\mathbb{F}_p$. This algorithm performs roughly 10 times better than the usual scalar multiplication.

The instantiation of the starting curve depends on the implementation of the arithmetic. If the tweaked Montgomery form is used, then it is $E_0^- : y^2 = x^3 - x$, but if it relies on standard Montgomery form (that is mapping back and forth, see Algorithm 8), then the starting curve is set to $E : y^2 = 3x^2 + 2x$.

The prime p in CSURF is kept very similar to how the CSIDH prime looks like. The CSURF prime is of the form

$$p = 4 \cdot 3 \cdot (\ell_0 \cdot \ell_1 \cdots \ell_{74}) - 1 = 4 \cdot 3 \cdot \underbrace{(2 \cdot 3 \cdot \ldots \cdot 389)}_{\substack{\text{75 consecutive primes,} \\ \text{skipping 347 and 359}}} - 1 \approx 2^{512.880}.$$

With the corresponding exponent vectors sampled randomly from the key space

$$[-137; 137] \times [-4; 4]^3 \times [-5; 5]^{46} \times [-4; 4]^{25},$$

Algorithm 8 Computing e lots of 2-isogenies on Montgomery$^-$ coefficient A with $p \equiv 7 \pmod 8$

if $e = 0$ **then return** A
else
  $A \leftarrow \text{sign}(e) \cdot A$
  $A \leftarrow 2\dfrac{A - 3\sqrt{A^2+4}}{A + \sqrt{A^2+4}}$
  for i from 2 to e **do**
    $A \leftarrow 2(3 + A(\sqrt{A^2-4} - A))$
  end for
  $A \leftarrow \dfrac{A + 3\sqrt{A^2-4}}{\sqrt{2\sqrt{A^2-4}(A + \sqrt{A^2-4})}}$
end if
return $\text{sign}(e) \cdot A$

resulting in $275 \cdot 9^{28} \cdot 11^{46} \approx 2^{255.995}$. The extra multiple of 2 is there to change to the $p \equiv 7 \pmod 8$ setting, the extra multiple of 3 helps with generating the kernel point P for the 3-isogeny faster, since if both $[(p+1)/9]Q$ and $[(p+1)/3]Q$ are computed, then the chance of failure is only 1/9. Another speedup comes from the fact that the largest prime is 389 instead of 587 as in CSIDH, but the biggest speedup is obviously due to the possibility of using 2-isogenies. This is taken advantage of in the keyspace, since CSURF allows the exponent of the 2-isogeny to be as large as 137. This also allows the largest primes to be evaluated at most 4 times, instead of 5.

According to the performance tests of CSURF-512, its Magma implementation performed 5.68% faster than CSIDH-512. This estimate is based on generating 25000 public keys in both implementations. They note, however, that advantage gained in the lowest bits are expected to worsen as the prime p increases, which it very well must do based on the discussion in Subsection 4.2.5, since that attack applies to CSURF too.

Since the publishing of the original CSURF paper, there have been (at least) two major news regarding CSURF. The first one is a paper by Fan, Tian, Li and Xu [33]. They independently considered the same idea as CSURF, but they have also showed in their paper that CSURF has a collision [33, Corollary 1] in the representation of the ideal classes, namely the key exponent vectors

$$(e_0, e_1, e_2, \ldots, e_{74}) = (e_0 + 1, e_1 + 2, e_2 + 1, \ldots, e_{74} + 1).$$

This follows from the fact that in their setup, we have that

$$\begin{aligned}
\mathfrak{l}_0\mathfrak{l}_1^2\mathfrak{l}_3 \ldots \mathfrak{l}_{74} &= \langle 2, \frac{\pi-1}{2}\rangle \cdot \langle \ell_3, \frac{\pi-1}{2}\rangle^2 \cdots \langle \ell_{74}, \frac{\pi-1}{2}\rangle \\
&= \langle 2 \cdot \ell_1^2 \cdots \ell_{74}, \frac{\pi-1}{2}\rangle \\
&= \langle \frac{1-\pi^2}{4}, \frac{\pi-1}{2}\rangle \\
&= \langle \frac{\pi-1}{2}\rangle .
\end{aligned}$$

therefore $\mathfrak{l}_0\mathfrak{l}_1^2\mathfrak{l}_3 \ldots \mathfrak{l}_{74}$ is a principal ideal in $\mathcal{O}_K$, so the action of the ideal class of such an exponent vector is trivial. They suggested an updated key space

$$[-141; 141] \times [-4; 4]^3 \times [-6; 6]^{13} \times [-5; 5]^{33} \times [-4; 4]^{24},$$

where the key space is still covered heuristically, but they also drop the largest prime $\ell_{74} = 389$.

In their own performance test, they implemented this new key space, together with an inversion of the isogeny evaluation queue, i.e. they changed the direction of the loop to go through the primes in descending order. This reportedly achieved a marked increase in speed of 28.69%.

In 2022, the authors of CSURF teamed up with Houben and Vercauteren [16] to develop the theory

of radical isogenies and applying to CSURF. The details of these isogenies are quite involved, but at a high level they can be explained as a much more efficient way of computing cyclic chains of ℓ-isogenies by extracting ℓ'th roots in $\mathbb{F}_p$. This helps with the expensive task of generating kernel generating points in the usual multiplication-by-$(p+1)/\ell$ way. In this paper, they developed a new way of finding these radical isogenies, optimized formulae for numbers up to $\ell = 19$, and they also found a way to ensure the horizontality of 2-isogenies. They have also generated a new prime CRAD-513 [16, pp. 24], which seems to have gained a 12% speedup compared to CSURF-512.

5.4 Conclusion – Future Work

We mentioned briefly that a paper titled "Optimization and Practicality of High-Security CSIDH" [10] by PQShield and other cryptography researches have tried successfully to speed up the performance of CSIDH-based improvements for larger primes, as a response to the threats from Kuperberg's algorithms. In this paper, they also implemented their improved versions, dCSIDH and CTIDH, in the TLS protocol to see how it performs compared to other, existing implementations. Specifically, they compared their Post-Quantum TLS without signatures implementation to TLS 1.3 and KEMTLS [64]. Their full results are available [10, Table 6], and the fastest speeds for all types is presented in Table 5.6.

Type of TLS Protocol	**Specific Implementation**	**Transmission (bytes)**		**Handshake latencies** (RTT, link speed)			
				30.9 ms, 1000 Mbps		**195.5 ms, 10 Mbps**	
		KEX	**Auth**	**SFIN recv**	**CFIN sent**	**SFIN recv**	**CFIN sent**
OPTLS (pregen)	CTIDH p2048	512	922	7.346	7.346	7.203	7.203
OPTLS (ephemeral)	CTIDH p2048	512	922	10.042	10.042	9.882	9.882
TLS	Kyber512–Falcon512	1568	2229	0.064	0.064	0.428	0.428
KEMTLS	Kyber512	1568	2234	0.094	0.063	0.593	0.396

Table 5.6: Comparison of the fastest implementations for OPTLS, TLS and KEMTLS, with the following columns: **KEX** is key exchange size, **Auth** is the authentication payload size, and Handshake latencies are shown for high-speed and low-speed links, **SFIN recv** is Server Finished received and **CFIN sent** is Client Finished sent. [10, Table 6.]

As can be seen from the table, the CSIDH implementations are magnitudes slower than the Kyber512-based protocols. It does seem like, as it stands, that CSIDH is not a competitive option in general, but still might be useful in very niche scenarios with very low bandwidths, where the protocol's time is near-equal to the time it takes to send several packages for the key exchange message.

This result is, however, no reason to stop researching the potentials of CSIDH-related cryptography ideas within isogeny-based cryptography. There might be other ideas to explore, for example combining previous ideas to produce something new, just like the CTIDH team has created new algorithms and key spaces that have symbiotically improved the performance.

One might try to combine isogeny-based theory with other ideas, as there is a potential alternative being developed as the writing of this thesis. Posted in July 3rd, 2026 in a PQC forum, there seems to be talks of a NIKE, which is called "Module Isogeny Key Exchange" (MIKE), with promised key size of 64B for NIST Level 1, constant-time implementation with ¡5 ms time for the full key exchange. The MIKE consists of Andrea Basso, Pierrick Dartois, Max Duparc, Jonathan Komada Eriksen, Sabrina Kunzweiler, Michael Meyer, Giacomo Pope, Krijn Reijnders, Damien Robert, Ryan Rueger, and Sina Schaeffler. The keen reader might have noticed that some of these names have already appeared in the references of this thesis, so there is good hope in the validity of MIKE. The white paper is planned to be released at the end of August 2026, so unfortunately no citation can be given as of yet.

They claim that their solution's cryptographic security relies on the supersingular endomorphism ring problem, which is the standard assumption across isogeny-based models. The Post-Quantum Cryptography world is volatile as ever, since on July 23rd, 2026, Benjamin Wesolowski [77] has published a paper showing that the supersingular isogeny problem can be solved in a better time complexity, namely $O(p^{1/3+o(1)})$. The previous best estimate was $O(p^{1/2} \cdot \log(p)^{O(1)})$, and that was the basis taken during the last years, and in this thesis. The implications of this paper are new, and it would be definitely worth

exploring the impact of it on all isogeny-based cryptographic models. CSIDH is not impacted by this too much, but mostly because Kuperberg's algorithm is a much bigger threat.

Bibliography

[1] G. Banegas, D. J. Bernstein, F. Campos, T. Chou, T. Lange, M. Meyer, B. Smith, and J. Sotáková. "CTIDH: faster constant-time CSIDH". In: *IACR Transactions on Cryptographic Hardware and Embedded Systems* **2021**.4 (2021), pp. 351–387. DOI: 10.46586/tches.v2021.i4.351-387.

[2] G. Banegas, A. Hellenbrand, and M. Saldanha. "Hardened CTIDH: Dummy-Free and Deterministic CTIDH". In: *Progress in Cryptology – INDOCRYPT 2025*. Vol. 16372. Springer, Cham, 2025, pp. 194–215. DOI: 10.1007/978-3-032-13301-4_9.

[3] D. Bernstein, J. De Feo, A. Leroux, and B. Smith. "Faster computation of isogenies of large prime degree". In: *Proceedings of the Fourteenth Algorithmic Number Theory Symposium – ANTS XIV*. Vol. 4. 1. Mathematical Sciences Publishers, 2020, pp. 39–55. DOI: 10.2140/obs.2020.4.39.

[4] T. Bernstein D. J. Lange, C. Martindale, and L. Panny. "Quantum Circuits for the CSIDH: Optimizing Quantum Evaluation of Isogenies". In: *Advances in Cryptology – EUROCRYPT 2019*. Springer International Publishing, 2019, pp. 409–441. DOI: 10.1007/978-3-030-17656-3_15.

[5] W. Beullens, T. Kleinjung, and F. Vercauteren. "CSI-FiSh: Efficient Isogeny Based Signatures Through Class Group Computations". In: *Advances in Cryptology -– ASIACRYPT 2019*. Springer-Verlag, 2019, pp. 227–247. DOI: 10.1007/978-3-030-34578-5_9.

[6] J-F. Biasse, D. Jao, and A. Sankar. "A quantum algorithm for computing isogenies between supersingular elliptic curves". In: *Progress in Cryptology – INDOCRYPT 2014*. Springer International Publishing, 2014, pp. 428–442. DOI: 10.1007/978-3-319-13039-2_25.

[7] B.J. Birch. "How the number of points of an elliptic curve over a fixed prime field varies". In: *Journal of London Mathematical Society* **43** (1968), pp. 57–60. DOI: 10.2307/2007968.

[8] X. Bonnetain and A. Schrottenloher. "Quantum security analysis of CSIDH". In: *Advances in Cryptology – EUROCRYPT 2020 – Part II*. Springer, 2020, pp. 493–522. DOI: 10.1007/978-3-030-45724-2_17.

[9] M. Cain, Q. Xu, R. King, L.R.B. Picard, H. Levine, M. Endres, J. Preskill, H-Y. Huang, and D. Bluvstein. *Shor's algorithm is possible with as few as 10,000 reconfigurable atomic qubits*. 2026. URL: https://arxiv.org/abs/2603.28627.

[10] F. Campos, J. Chávez-Saab, J-J. Chi-Domínguez, M. Meyer, K. Reijnders, F. Rodríguez-Henríquez, P. Schwabe, and T. Wiggers. "Optimizations and Practicality of High-Security CSIDH". In: *IACR Communications in Cryptology* **1**.1 (2024). DOI: 10.62056/anjbksdja.

[11] F. Campos, A. Hellenbrand, M. Meyer, and K. Reijnders. "dCTIDH: Fast Deterministic CTIDH". In: *IACR Transactions on Cryptographic Hardware and Embedded Systems* **2025**.3 (2025), pp. 516–541. DOI: 10.46586/tches.v2025.i3.516-541.

[12] F. Campos, M. J. Kannwischer, M. Meyer, H. Onuki, and M. Stöttinger. "Trouble at the CSIDH: Protecting CSIDH with Dummy-Operations against Fault Injection Attacks". In: *17th Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC)*. IEEE, 2020, pp. 57–65. URL: https://eprint.iacr.org/2020/1005.

[13] W. Castryck. *CSIDH ON THE SURFACE (CSURF) – Lecture Notes*. Isogeny-Based Cryptography School – Week 3, 2021. 2021. URL: https://homes.esat.kuleuven.be/~wcastryc/summer_school_csurf.pdf. (accessed: 2026-06-22).

[14] W. Castryck and T. Decru. "CSIDH on the surface". In: *Lecture Notes in Computer Science – PQCrypto 2020*. Springer, Cham, 2020, pp. 111–129. DOI: 10.1007/978-3-030-45724-2_16.

[15] W. Castryck and T. Decru. *An efficient key recovery attack on SIDH*. Cryptology ePrint Archive, Paper 2022/975. 2022. URL: https://eprint.iacr.org/2022/975.

[16] W. Castryck, T. Decru, M. Houben, and F. Vercauteren. "Horizontal Racewalking Using Radical Isogenies". In: *Advances in Cryptology – ASIACRYPT 2022*. Springer-Verlag, 2022, pp. 67–96. DOI: 10.1007/978-3-031-22966-4_3.

[17] W. Castryck, T. Lange, C. Martindale, L. Panny, and J. Renes. "CSIDH: An Efficient Post-Quantum Commutative Group Action". In: *Advances in Cryptology – ASIACRYPT 2018*. Springer International Publishing, 2018, pp. 395–427. DOI: 10.1007/978-3-030-03332-3_15.

[18] D. Cervantes-Vázquez, M. Chenu, J-J. Chi-Domínguez, L. De Feo, F. Rodríguez-Henríquez, and B. Smith. "Stronger and Faster Side-Channel Protections for CSIDH". In: *LATINCRYPT 2019*. Springer, Heidelberg, 2019, pp. 179–193. DOI: 10.1007/978-3-030-30530-7_9.

[19] D.X. Charles, E.Z. Goren, and K.E. Lauter. "Cryptographic Hash Functions from Expander Graphs". In: *Journal of Cryptology* **22**.1 (2009), pp. 93–113. DOI: 10.1007/s00145-007-9002-x.

[20] J. Chávez-Saab, J-J. Chi-Domínguez, S. Jaques, and F. Rodríguez-Henríquez. "The SQALE of CSIDH: Sublinear Vélu Quantum-resistant isogeny Action with Low Exponents". In: *Journal of Cryptographic Engineering (JCEN)* **12**.3 (2022), pp. 349–368. DOI: 10.1007/s13389-021-00271-w.

[21] J-J. Chi-Domínguez and F. Rodríguez-Henríquez. "Optimal strategies for CSIDH". In: *Adv. Math. Commun.* Vol. 16. Springer, Heidelberg, 2022, pp. 383–411. DOI: 10.3934/amc.2020116.

[22] A.M. Childs, D. Jao, and V. Soukharev. "Constructing elliptic curve isogenies in quantum subexponential time". In: *Journal of Mathematical Cryptology* **8**.1 (2013), pp. 1–29. DOI: 10.1515/jmc-2012-0016.

[23] I. L. Chuang and M. A. Nielsen. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN: 9781107002173.

[24] A. Costache, B. Feigon, K. Lauter, M. Massierer, and A. Puskás. *Ramanujan graphs in cryptography*. Cryptology ePrint Archive, Paper 2018/593. 2018. URL: https://eprint.iacr.org/2018/593.

[25] C. Costello and H. Hisil. "A Simple and Compact Algorithm for SIDH with Arbitrary Degree Isogenies". In: *Advances in Cryptology – ASIACRYPT 2017*. Springer International Publishing, 2017, pp. 303–329. DOI: 10.1007/978-3-319-70697-9_11.

[26] C. Costello and B. Smith. *Montgomery curves and their arithmetic: The case of large characteristic fields*. Cryptology ePrint Archive, Paper 2017/212. 2017. URL: https://eprint.iacr.org/2017/212.

[27] J-M. Couveignes. *Hard Homogeneous Spaces*. IACR Cryptology ePrint Archive. 2006. URL: https://ia.cr/2006/291.

[28] D.A. Cox. *Primes of the form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication*. 2nd ed. John Wiley Sons, Inc, 20013. ISBN: 9781118400722.

[29] L. De Feo. "Mathematics of Isogeny Based Cryptography". In: *CoRR* (2017), pp. 1–44. DOI: 10.48550/arXiv.1711.04062.

[30] L. De Feo, D. Jao, and J. Plût. *Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies*. Cryptology ePrint Archive, Paper 2011/506. 2011. URL: https://eprint.iacr.org/2011/506.

[31] C. Delfs and S.D. Galbraith. "Computing isogenies between supersingular elliptic curves over $\mathbb{F}_p$". In: *Designs, Codes and Cryptography* **78** (2016), pp. 425–440. DOI: 10.1007/s10623-014-0010-1.

[32] W. Diffie and M. Hellman. "New directions in cryptography". In: *IEEE Transactions on Information Theory* **22**.6 (1976), pp. 644–654. DOI: 10.1109/TIT.1976.1055638.

[33] X. Fan, T. Tian, B. Li, and X. Xu. *CSIDH on Other Form of Elliptic Curves*. Cryptology ePrint Archive, Paper 2019/1417. 2019. URL: https://eprint.iacr.org/2019/1417.

[34] M. Fouquet and F. Morain. "Isogeny Volcanoes and the SEA Algorithm". In: *Algorithmic Number Theory*. Springer, 2002, pp. 276–291. DOI: 10.1007/3-540-45455-1_23.

[35] E. Fujisaki and T. Okamoto. “Secure integration of asymmetric and symmetric encryption schemes”. In: *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '99)*. Springer-Verlag, 1999, pp. 537–554. DOI: 10.1007/3-540-48405-1_34.

[36] P. et al Gaborit. *Hamming Quasi-Cyclic*. 2025. URL: https://pqc-hqc.org/doc/hqc_specifications_2025_08_22.pdf.

[37] S. Galbraith. *Eurocrypt 2020*. Wordpress. 2020. URL: https://ellipticnews.wordpress.com/2020/05/19/eurocrypt-2020/. (accessed: 2026-05-12).

[38] S. Galbraith. *The Ideal Class Group Action on Supersingular Elliptic Curves*. From series of lecture notes at a summer school in Akaroa, New Zealand. 2022.

[39] J. Grout, W. Stein, I. Hanson, S. Johnson, and A. Kramer. *SageMathCell – Sage Cell Server*. SageMath, Inc. 2011. URL: https://sagecell.sagemath.org/. (accessed: 2026-06-25).

[40] L.K. Grover. “A fast quantum mechanical algorithm for database search”. In: *Proceedings of the TwentyEighth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 1996, pp. 212–219. DOI: 10.1145/237814.237866.

[41] D. Gulotta. *Ramanujan graphs, quaternions, and number theory – Lecture 1*. Canada/USA Mathcamp. 2020. URL: https://dgulotta.github.io/ramanujan1.pdf. (accessed: 2026-07-22).

[42] T. Hoang. *Cerebras Systems Smashes the 2.5 Trillion Transistor Mark with New Second Generation Wafer Scale Engine*. Cerebras. 2021. URL: https://www.cerebras.ai/press-release/cerebras-systems-smashes-the-2-5-trillion-transistor-mark-with-new-second-generation-wafer-scale-engine.

[43] A. Hutchinson, J.T. LeGrow, B. Koziel, and R. Azarderakhsh. “Further Optimizations of CSIDH: A Systematic Approach to Efficient Strategies, Permutations, and Bound Vectors”. In: *ACNS 20, Part I*. Springer, Heidelberg, 2020, pp. 481–501. DOI: 10.1007/978-3-030-57808-4_24.

[44] D. Jao, S.D. Miller, and R. Venkatesan. “Expander graphs based on GRH with an application to elliptic curve cryptography”. In: *Journal of Number Theory* **129**.6 (2009), pp. 1491–1504. DOI: 10.1016/j.jnt.2008.11.006.

[45] D.R. Kohel. “Endomorphism rings of elliptic curves over finite fields”. PhD thesis. University of California at Berkley, 1996.

[46] G. Kuperberg. “A subexponential-time quantum algorithm for the dihedral hidden subgroup problem”. In: *SIAM J. COMPUT.* **35**.1 (2005), pp. 170–188. DOI: 10.1137/S0097539703436345.

[47] S. Lang. *Elliptic Functions*. 2nd ed. Graduate Texts in Mathematics. Springer New York, NY, 1987. DOI: 10.1007/978-1-4612-4752-4.

[48] L. Maino, C. Martindale, L. Panny, G. Pope, and B. Wesolowski. “A Direct Key Recovery Attack on SIDH”. In: *Advances in Cryptology – EUROCRYPT 2023*. Ed. by Carmit Hazay and Martijn Stam. Vol. 14008. Lecture Notes in Computer Science. Springer, Heidelberg, 2023, pp. 448–471. DOI: 10.1007/978-3-031-30589-4_16.

[49] M. Meyer, F. Campos, and S. Reith. “On Lions and Elligators: An efficient constant-time implementation of CSIDH”. In: *10th International Conference – PQCrypto 2019*. Springer, 2019, pp. 307–325. URL: https://eprint.iacr.org/2018/1198.

[50] J. S. Milne. *Algebraic Number Theory (v3.08)*. Available at www.jmilne.org/math/. 2020.

[51] T. Morrison, L. Panny, J. Sotáková, and M. Wills. *The SEA algorithm for endomorphisms of supersingular elliptic curves*. arXiv. 2026. URL: https://arxiv.org/abs/2501.16321.

[52] K. Nakagawa, H. Onuki, A. Takayasu, and T. Takagi. *L_1-Norm Ball for CSIDH: Optimal Strategy for Choosing the Secret Key Space*. Cryptology ePrint Archive, Paper 2020/181. 2020. URL: https://eprint.iacr.org/2020/181.

[53] NIST. *Post-quantum cryptography standardization*. National Institute of Standards and Technology. 2016. URL: https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/call-for-proposals-final-dec-2016.pdf. (accessed: 2026-04-11).

[54] NIST. *Module-Lattice-Based Key-Encapsulation Mechanism Standard*. National Institute of Standards and Technology. 2024. DOI: 10.6028/NIST.FIPS.203.

[55] H. Onuki, Y. Aikawa, T. Yamazaki, and T. Takagi. "(Short Paper) A Faster Constant-Time Algorithm of CSIDH Keeping Two Points". In: *IWSEC 19*. Springer, Heidelberg, 2019, pp. 23–33. DOI: 10.1007/978-3-030-26834-3_2.

[56] C. Peikert. "He Gives C-Sieves on the CSIDH". In: *Advances in Cryptology – EUROCRYPT 2020*. Springer, Cham, 2020, pp. 463–492. DOI: 10.1007/978-3-030-45724-2_16.

[57] A.K. Pizer. "Ramanujan Graphs". In: *Computational Perspectives on Number Theory (Chicago, IL, 1995)*. Vol. 7. American Mathematical Society, 1998, pp. 159–178. DOI: 10.1090/amsip/007.

[58] Damien R. "Breaking SIDH in Polynomial Time". In: *Advances in Cryptology – EUROCRYPT 2023*. Vol. 14008. Lecture Notes in Computer Science. Springer, Heidelberg, 2023, pp. 472–503. DOI: 10.1007/978-3-031-30589-4_17.

[59] J. Renes. "Computing Isogenies Between Montgomery Curves Using the Action of (0, 0)". In: *Post-Quantum Cryptography – PQCrypto 2018*. Springer International Publishing, 2018, pp. 229–247. DOI: 10.1007/978-3-319-79063-3_11.

[60] M. Roetteler, M. Naehrig, K. M. Svore, and K. Lauter. "Quantum Resource Estimates for Computing Elliptic Curve Discrete Logarithms". In: *Advances in Cryptology – ASIACRYPT 2017*. Springer International Publishing, 2017, pp. 241–270. URL: https://eprint.iacr.org/2017/598.

[61] A. Rostovtsev and A. Stolbunov. *Public-key cryptosystem based on isogenies*. IACR Cryptology ePrint Archive. 2006. URL: https://ia.cr/2006/145.

[62] R. Schoof. "Elliptic curves over finite fields and the computation of square roots mod p". In: *Mathematics of Computation* **44**.170 (1985), pp. 483–494. DOI: 10.2307/2007968.

[63] R. Schoof. "Counting points on elliptic curves over finite fields". In: *Journal de théorie des nombres de Bordeaux* **7** (1995), pp. 219–254. DOI: 10.1007/BF01404549.

[64] P. Schwabe, D. Stebila, and T. Wiggers. "Post-Quantum TLS Without Handshake Signatures". In: *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. Association for Computing Machinery, 2020, pp. 1461–1480. DOI: 10.1145/3372297.3423350.

[65] D. Shanks. "Class number, a theory of factorization, and genera". In: *Proceedings of Symposium of Pure Mathematics* **XX** (1969), pp. 415–440. URL: https://mathscinet.ams.org/mathscinet/relay-station?mr=316385.

[66] C. Shao. *Introduction to Quantum Computing*. Isogeny-Based Cryptography School – Week 8, 2021. 2021. URL: https://isogenyschool2020.co.uk/schedule/Introduction-to-quantum-computing-lecture-notes.pdf. (accessed: 2026-06-22).

[67] P.W. Shor. "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer". In: *SIAM J. Comput.* **26**.5 (1997), pp. 1484–1509. URL: https://arxiv.org/abs/quant-ph/9508027.

[68] C. Siegel. "Über die Classenzahl quadratischer Zahlkörper". ger. In: *Acta Arithmetica* **1**.1 (1935), pp. 83–86. URL: http://eudml.org/doc/205054.

[69] J.H. Silverman. *The Arithmetic of Elliptic Curves*. 2nd ed. Graduate Texts in Mathematics. Springer New York, NY, 2009. DOI: 10.1007/978-0-387-09494-6.

[70] A.V. Sutherland. *Elliptic Curves: Lecture Notes and Worksheets*. MIT OpenCourseWave. 2021. URL: https://ocw.mit.edu/courses/18-783-elliptic-curves-spring-2021/pages/lecture-notes-and-worksheets/. (first accessed: 2026-02-11).

[71] S. Tani. "Claw finding algorithms using quantum walk". In: *Theoretical Computer Science* **410**.50 (2009), pp. 5285–5297. DOI: 10.1016/j.tcs.2009.08.030.

[72] T. Tao. *Pappus' theorem and elliptic curves*. Wordpress. 2011. URL: https://terrytao.wordpress.com/2011/07/15/pappuss-theorem-and-elliptic-curves/. (accessed: 2026-04-19).

[73] J. Tate. "Endomorphisms of abelian varieties over finite fields". In: *Inventiones mathematicae* **2** (1966), pp. 134–144. DOI: 10.1007/BF01404549.

[74] J. Vélu. "Isogénies entre courbe elliptiques". In: *C. R. Acad. Sci. Paris Séries A* **273** (1971), pp. 238–241. URL: https://aghitza.github.io/publications/translation-velu.pdf. (English translation by Alex Ghitza).

[75] L.C. Washington. *Elliptic Curves: Number Theory and Cryptography*. 2nd ed. Discrete Mathematics and Its Applications. CRC Press, 2008. ISBN: 9781040066195.

[76] W. C. Waterhouse. "Abelian varieties over finite fields". In: *Annales scientifiques de l'École Normale Supérieure* **2**.4 (1969), pp. 521–560. URL: https://www.numdam.org/article/ASENS_1969_4_2_4_521_0.pdf.

[77] B. Wesolowski. *The supersingular isogeny problem in time and memory* $p^{1/3+o(1)}$. Cryptology ePrint Archive, Paper 2026/1486. 2026. URL: https://eprint.iacr.org/2026/1486.

CORRECTIONS FOR
SeaSide Surfing on High Tide: An overview of CSIDH and its recent improvements in Post-Quantum Cryptography

This document is meant to give extra reasoning / fix faulty logic in my master thesis, and is meant to contain only bigger issues. Therefore, I will not mention the simple typos in text form, only typos in mathematical expressions that would render the argument incorrect. When I write "line $-k$" for the line reference, I mean that it is the k'th from the bottom.

Page 8, after Lemma 2.1.12, line 1-4 [Statement correction, Proof extension]

We assume that

$$\infty, \quad A, B, C, \quad A+B, B+C, -(A+B), -(B+C)$$

are eight distinct points on the curve. Furthermore, assume that $-((A+B)+C)$ and $-(A+(B+C))$ are also distinct from all these eight points, but not necessarily from each other. This is called being in the "general position". The proof is correct for this case, but there is nothing mentioned about the degenerate cases, i.e. when some of the points are equal.

This proof can be extended by a perturbation argument, which relies on the fact that the secant of two points are continuous when varying the distance between them, and their limit is the tangent at say the point A. The nonsingular nature of the elliptic curve means that the tangent is well-defined everywhere, so this limit actually exists for all points. The claimed effect of nonsigularity in the proof (**line 4**) is incorrect as it stands.

Page 13, proof of Lemma 2.2.3, line -3 [Typo]

"$q_2(x) \implies 0$" $\longrightarrow$ "$q_1(x) \implies 0$"

Page 13, proof of Theorem 2.2.6, line 1-2 [Statement correction]

"Since E' is supersingular" $\longrightarrow$ "Since E' is separable"

Page 20, Theorem 2.3.1, line 2-3 [Statement extension]

"The curve E' and the isogeny φ are defined over a finite extension of k and unique up to isomorphism." $\longrightarrow$ "The curve E' and the isogeny φ are defined over **the same** finite extension of k **as the finite subgroup** G, and **are** unique up to isomorphism."

This statement is modified so that the proof of Theorem 2.7.16 is correct when referring to this theorem. This statement modification is mentioned in Corollary III.4.13 in Silverman's book.

Page 24, Definition 2.4.6, line 1 [Statement correction]

"Let D be a squarefree integer, satisfying a monic quadratic equation with integer coefficients" $\longrightarrow$ "Let D be a squarefree integer."

Page 27, Theorem 2.4.26, line 2 [Statement extension]

"If $\mathfrak{a}$ is invertible, then the inverse [...]" $\longrightarrow$ "If $\mathfrak{a}$ is invertible, then $\mathfrak{a}\overline{\mathfrak{a}} = (\mathrm{N}(\mathfrak{a}))$, and the inverse [...]"

This should be added to the statement of the theorem to make it more clear for the reader what I am referring to in **line 4** of the proof of Corollary 2.4.27.

Page 28, line 5 [Statement correction]

"if we let $\mathcal{O} = [1, \omega]$ be another $\mathcal{O}$-ideal," $\longrightarrow$ "if we let $\mathcal{O} = [1, \omega]$ be the basis of the imaginary quadratic order $\mathcal{O}$,"

Page 30, Definition 2.4.34 [Typo]

In the definition of $\left(\frac{a}{0}\right)$, the second condition should be "0" instead of "01".

Page 30, proof of Theorem 2.4.35 [Citation extension]

"See [28]" $\longrightarrow$ "See [28, Proposition 5.16.]"

Page 30, Definition 2.4.36, line 2 [Typo]

"$\alpha \neq \beta \in \mathbb{F}_p$" $\longrightarrow$ "$a \neq b \in \mathbb{F}_p$"

Page 41, proof of Theorem 2.5.22, line -7 [Clarification]

The "same translations" here refer to the translations in the proof of Theorem 2.5.18(iii), which results in $\operatorname{tr}(\alpha) = 0$ and $\operatorname{tr}(\alpha\pi_E) = 0$. Luckily, the Frobenius endomorphism already has $\operatorname{tr}(\pi_E) = 0$ in our supersingular setting. Then, continuing the logic of the previous proof, we have that $\operatorname{tr}(\beta) = 0 \implies \beta = -\hat{\beta}$ in general, so we have that

$$\alpha = -\hat{\alpha}, \quad \pi_E = -\widehat{\pi_E} \quad \text{and} \quad \pi_E\alpha = -\hat{\alpha}\widehat{\pi_E}.$$

From this, substituting in the first two equalities into the third one yields the claimed equality $\alpha\pi_E = -\pi_E\alpha$.

Page 51, Lemma 2.7.9 [Statement extension]

"A non-zero principal $\mathcal{O}$-ideal $\mathfrak{a} = (\alpha)$ is associated with an endomorphism." $\longrightarrow$ "A non-zero principal $\mathcal{O}$-ideal $\mathfrak{a} = (\alpha)$, **where $\mathbf{N}(\alpha)$ is coprime to p,** is associated with an endomorphism."
This modification is necessary, because as the proof stands, it is only true for separable isogenies. To deal with the principal ideal (π), we would need to invoke scheme theory, which is not part of this thesis. This does not affect the validity of the rest of thesis, because it is always possible to choose a representative coprime to p from the ideal class $\mathfrak{a}$.

Page 51, proof of Lemma 2.7.9, line 2 [Typo]

"$E[I] \subseteq \ker(\alpha)$" $\longrightarrow$ "$E[\mathfrak{a}] \subseteq \ker(\alpha)$"

Page 53, proof of Theorem 2.7.13, line -6 [Proof extension]

The claim that I_φ is proper is not obvious. Notice that $I_\varphi = Hom(E_2, E_1)\varphi$, which follows from Corollary III.4.11 of Joseph Silverman's book titled "The Arithmetic of Elliptic Curves". It states that if $\varphi : E_1 \to E_2$ is a separable isogeny, and $\psi : E_1 \to E_3$ is any isogeny with $\ker\varphi \subseteq \ker\psi$, then there is a unique isogeny $\lambda : E_2 \to E_3$ satisfying $\psi = \lambda \circ \varphi$.
Next, if we define the fractional $\mathcal{O}$-ideal

$$J_\varphi = \frac{1}{\mathrm{N}(\varphi)}\hat{I}_\varphi = \left\{ \frac{\hat{\alpha}}{\mathrm{N}(\varphi)} \mid \alpha \in I_\varphi \right\} = \frac{\hat{\varphi}}{\mathrm{N}(\varphi)} Hom(E_1, E_2),$$

where the last equation follows from $\hat{\alpha} = \widehat{\lambda \circ \varphi} = \hat{\varphi} \circ \hat{\lambda}$ for all $\lambda \in Hom(E_2, E_1)$ and $\hat{\lambda} \in Hom(E_1, E_2)$. We then have that

$$\begin{aligned} I_\varphi J_\varphi &= (Hom(E_2, E_1)\varphi) \cdot \left(\frac{\hat{\varphi}}{\mathrm{N}(\varphi)} Hom(E_1, E_2) \right) \\ &= Hom(E_2, E_1) \circ \varphi \circ \frac{\hat{\varphi}}{\mathrm{N}(\varphi)} \circ Hom(E_1, E_2) \\ &= Hom(E_2, E_1) \circ \frac{[\deg\varphi]}{\mathrm{N}(\varphi)} \circ Hom(E_1, E_2) \\ &= Hom(E_2, E_1) \circ [1] \circ Hom(E_1, E_2) \\ &= \mathrm{End}(E_1) = \mathcal{O}. \end{aligned}$$

This shows that I_φ is invertible with inverse J_φ, and by a theorem in the thesis, it is also proper.

Page 54, below Table 2.1, line 7 [Typo]

"$[\mathfrak{a}] * E_0 = E_A$" $\longrightarrow$ "$[\mathfrak{a}] * E_0 = E$"

Page 78, line 2 [Typo]

"$\sum_{x=x_0}$" $\longrightarrow$ "$\sum_{x\neq x_0}$"

Page 81, line -7 [Missing reference]

"Algorithm **??**" $\longrightarrow$ "Algorithm 1"

Page 83, line 6 [Missing reference]

"SQALE [**sqale**]" $\longrightarrow$ "SQALE [20]"

Page 89, line 3 [Missing reference]

"See **??**" $\longrightarrow$ "See Section 5.4"

Page 89, Section 5.3, line 10 [Missing reference]

"hardness of the parallelization problem (see Problem ??)" $\longrightarrow$ "hardness of determining E_{AB} from only knowing E_0, E_A and E_B"