SUPERSINGULAR ELLIPTIC CURVES, QUATERNION ALGEBRAS AND APPLICATIONS TO CRYPTOGRAPHY

EYAL Z. GOREN & JONATHAN LOVE

ABSTRACT. This paper contains a survey of supersingular isogeny graphs associated to supersingular elliptic curves and their various applications to cryptography. Within limitation of space, we attempt to address a broad audience and make this part widely accessible. For those graphs we also present three recent results and sketch their proofs. We then discuss a generalization to superspecial isogeny graphs associated to superspecial abelian varieties with real multiplication. These graphs were introduced by Charles, Goren and Lauter and so our discussion is brief. Motivated by their cryptographic applications, we prove a general theorem concerning generation of lattices over totally real fields by elements of specified norm. Throughout the paper we have attempted to clarify certain considerations that are either vaguely stated in the literature, or are folklore. We hope this paper will be useful both to a novice wishing to familiarize themselves with this very active area, and to the expert who may enjoy some vignettes and an overview of some new results.

CONTENTS

1. Introduction

In recent years the use of supersingular elliptic curves in cryptography has been an active field of research and at various times some cryptographic primitives based on supersingular elliptic curves, such as the CGL hash function, the SIKE key-exchange protocol or the SQIsign digital signature protocol, have been strong contenders for post-quantum cryptography protocols. This created a fertile playing field in which results from arithmetic geometry, some more than a century old, others very recent, were brought to bear on questions from mathematical cryptography. More recently, higher dimensional superspecial graphs have been brought into this area. For one, they offer a generalization of the cryptographic primitives previously mentioned; in addition, recent devastatingly successful attacks on the SIKE protocol can be viewed as happening within such superspecial graphs. The introduction of superspecial graphs raises and

motivates many interesting questions concerning the arithmetic of higher dimensional abelian varieties and the study of their moduli spaces and endomorphism rings. Although this is a well-developed area already, the questions raised in this context are new and of much mathematical interest.

This paper offers a fairly thorough introduction to supersingular isogeny graphs and provides an overview of several applications of supersingular elliptic curves to cryptography. In this part of the paper, we have attempted to reach a somewhat general audience: we assume a background in algebraic number theory at the level of a first-year graduate course, as well as some basic exposure to algebraic geometry. As our goal is to reach people who are new to the subject, the presentation includes a considerable amount of background material; many proofs are omitted, but in these cases references are provided at the beginning of each relevant section.

We also present three recent theorems on these topics, motivated by cryptographic applications but also of independent interest. As the results have appeared in print already with complete proofs, we only provide a sketch of the proofs here, for completeness. These are Theorems 5.1.1, 5.2.2 and 5.3.1 in the article, where the first is due to S. Mayo, and the last two to the authors of the present paper. Theorem 5.1.1 states, roughly and under some conditions, that the only non-trivial automorphism of a supersingular isogeny graph is an involution that comes from the Frobenius map. Theorem 5.2.2 states that endomorphism rings of supersingular elliptic curves in characteristic p are generated by isogenies of order in the set $\{\ell^k : k = 0, 1, 2, 3 \ldots \}$, where ℓ is any fixed prime different from p. (In fact, Theorem 5.2.2 follows from Theorem 5.2.1, that treats generation of general lattices by elements of specified norms.) Finally, Theorem 5.3.1 asserts that supersingular elliptic curves E are determined up to Frobenius transform by their theta function $\Theta_E(q) := \sum_{n=0}^{\infty} r_E(n)q^n$, where $r_E(n)$ is the number of endomorphisms of E of norm n. The material in this part of the paper cannot be viewed as original anymore, but along the way we took care to clarify some points that are hard to find explicitly, or in clear form, in the literature. We do however include one new application of Theorem 5.2.1 to the theory of even unimodular lattices (Corollary 5.2.3).

Following that, we pick up the pace and assume more background from the reader, since keeping the same pace of exposition would have more than doubled the length of the paper. Following [CGL2], we introduce superspecial isogeny graphs associated to superspecial abelian varieties with real multiplication. For real quadratic fields, these so-called RM isogeny graphs seem to be interpolating between the graphs associated to elliptic curves and those associated to abelian surfaces; for instance, they retain many properties of the elliptic curve graphs, such as the Ramanujan property, that the more commonly studied graphs of abelian surfaces do not. They raise several attractive questions that we hope some of the readers will pick up; one of them being the actual implementation of cryptographic primitives based on superspecial abelian varieties with real multiplication; another is to compare the computational complexity of walking supersingular isogeny graphs of elliptic curves, of (say) superspecial abelian surfaces with real multiplication and of superspecial principally polarized abelian surfaces. Yet another is to generalize all of that for totally real fields of arbitrary class number.

In the last part of the paper, motivated by possible cryptographic applications, we generalize Theorem 5.2.1, and so Theorem 5.2.2, to lattices over totally real fields. These appear in the text as Theorems 6.6.1 and 6.6.3, respectively. Despite generalizing our results over $\mathbb{Q}$, the argument is constructed differently. We make use of inhomogeneous theta functions and the theta correspondence, as well as results on the asymptotic growth of Fourier coefficients. We thus stand on the shoulders of giants, as we use deep results by Shimura, Siegel-Weil, and Deligne. Besides their application to endomorphism rings of superspecial abelian varieties with real multiplication, we expect such theorems to have other lovely applications, for example of the sort we provide for even unimodular lattices over $\mathbb{Z}$.

A precise description of the contents of this paper can be gleaned from the detailed table of contents.

2. Quaternion algebras and supersingular elliptic curves

2.1. **Quaternion algebras.** In this discussion we provide only the mere definitions and results that would assist in reading the rest of the paper. For more background, a general discussion of quaternion algebras can be found in [Vig] and [Voi].

Let L be a number field, or one of its completions; thus, $\mathbb{R}$, $\mathbb{C}$, or a finite extension of $\mathbb{Q}_p$ for some prime p. A **quaternion algebra** B over L is an L-algebra that, as a vector space of dimension 4 over L, has a basis $1, i, j, k$:

$$B = L1 \oplus Li \oplus Lj \oplus Lk,$$

where there exists some non-zero $\alpha, \beta \in L$ such that

$$i^2 = \alpha, \quad j^2 = \beta, \quad ij = -ji = k.$$

One refers to such an algebra by the symbol $\left(\frac{\alpha,\beta}{L}\right)$.

On any such quaternion algebra we have the **canonical involution** which can be written in terms of the basis as

$$b = x + yi + zj + wk \quad \mapsto \quad \bar{b} = x - yi - zj - wk$$

and satisfies $\overline{ab} = \bar{b} \cdot \bar{a}$ for $a, b \in B$. We also define the reduced **trace**, which is an L-linear map $\mathrm{Tr}\colon B \to L$, and the reduced **norm**, which is a multiplicative homomorphism $\mathrm{Nm}\colon B^\times \to L^\times$. These are defined for $b = x + yi + zj + wk \in B$ by

$$\begin{aligned}
\mathrm{Tr}(b) &= b + \bar{b} &= 2x, \\
\mathrm{Nm}(b) &= b \cdot \bar{b} &= x^2 + \alpha y^2 + \beta z^2 + \alpha\beta w^2.
\end{aligned}$$

The norm is a quadratic form on B, and we can associate to it a bilinear form $(a,b) = \frac{1}{2}\mathrm{Tr}(a\bar{b})$ satisfying the relations

$$\mathrm{Nm}(a) = (a,a), \qquad (a,b) = \frac{1}{2}\left(\mathrm{Nm}(a+b) - \mathrm{Nm}(a) - \mathrm{Nm}(b)\right).$$

Note that some sources use a different normalization for the bilinear form.

2.1.1. *Ramification and splitting.* A quaternion algebra B over L is called **split** if B is isomorphic as an L-algebra to the 2×2 matrix ring $M_2(L)$. Otherwise it is called **ramified** and in this case it is a division algebra. If $L = \mathbb{C}$, then every quaternion algebra is split; if L is equal to $\mathbb{R}$, then the only ramified quaternion algebra is the Hamilton quaternions $\left(\frac{-1,-1}{\mathbb{R}}\right)$. If L is a finite extension of $\mathbb{Q}_p$ and L_2 its unramified quadratic extension with $\langle\sigma\rangle = \mathrm{Gal}(L_2/L)$, then the only ramified quaternion algebra over L is isomorphic to

$$\left\{ \begin{pmatrix} a & \pi b \\ b^\sigma & a^\sigma \end{pmatrix} : a, b \in L_2 \right\},$$

where π is a uniformizer of L.

If $\varphi \colon L \to M$ is a field homomorphism and B is a quaternion algebra over L, then $B \otimes_{L,\varphi} M$ is a quaternion algebra over M. For example, $\left(\frac{\alpha,\beta}{L}\right) \otimes_{L,\varphi} M \cong \left(\frac{\varphi(\alpha),\varphi(\beta)}{M}\right)$. In particular, if L is a number field, v is a place of L with corresponding completion L_v, and B a quaternion algebra over L, then v is called ramified (respectively, split) relative to B if $B \otimes_L L_v$ is ramified (resp., split) over L_v.

A fundamental theorem states that (1) the number of ramified places for B is finite and even; (2) B is determined up to isomorphism by its set of ramified places; (3) given any set S consisting of an even number of places of L, there is a quaternion algebra B over L with S as its set of ramified places. As such a B is unique up to isomorphism, one may write $B = B_S$.

Suppose that M/L is an extension of number fields, V is a place of M, and v is the restriction of V to L. If B is a quaternion algebra over L that is split at v then $B \otimes_L M$ is split at V. If B is ramified at v then $B \otimes_L M$ is ramified at V if and only if $[M_V : L_v]$ is odd.

2.1.2. *The quaternion algebra $B_{p,\infty}$.* We shall use the notation $B_{p,\infty}$ to denote that quaternion algebra over $\mathbb{Q}$ ramified precisely at p and ∞, where p is a finite prime. If $p = 2$ we obtain the rational Hamilton quaternions $B_{2,\infty} = \left(\frac{-1,-1}{\mathbb{Q}}\right)$. For $p \equiv 3 \pmod 4$, $B_{p,\infty}$ is the quaternion algebra $\left(\frac{-1,-p}{\mathbb{Q}}\right)$ [Vig, p. 79]. This is the algebra

$$\mathbb{Q} \otimes \mathbb{Q}i \oplus \mathbb{Q}j \oplus \mathbb{Q}k,$$

with

$$i^2 = -1, \ j^2 = -p, \ ij = -ji = k,$$

and so $k^2 = -p$ and the norm form is $x^2 + y^2 + pz^2 + pw^2$. Similar models can be given for all primes p, and they are rather explicit (loc. cit. p. 98).

2.2. **Orders and ideals.** We now assume that L is either a number field or a finite extension of $\mathbb{Q}_p$, and we denote $\mathcal{O}_L$ its ring of integers. As before, let B be quaternion algebra over L.

An **order** $\mathcal{O}$ in B is a rank 4 $\mathcal{O}_L$-module that contains a basis for B over L and is also a subring of B; it therefore contains $\mathcal{O}_L$. Since $\mathcal{O}_L$ need not have class number 1, it is possible that $\mathcal{O}$ is not a free $\mathcal{O}_L$-module, but it always contains a free rank 4 $\mathcal{O}_L$-submodule of finite index. Every element b of an order is integral over $\mathcal{O}_L$ and, in particular, $\mathrm{Tr}(b), \mathrm{Nm}(b)$ lie in $\mathcal{O}_L$. As $\bar{b} = b - \mathrm{Tr}(b)$, any order is closed under the canonical involution. An order is called **maximal** if it is not strictly contained in any other order.

Let (x,y) denote the bilinear form $\frac{1}{2}\mathrm{Tr}(x\bar{y})$ on B, and hence $2(a,b) \in \mathcal{O}_L$ for any $a, b \in \mathcal{O}$. For any free $\mathcal{O}_L$-submodule $\oplus_{i=1}^4 \mathcal{O}_L v_i$ of $\mathcal{O}$, we consider the value

$$d(v_1, \ldots, v_4) := \det(2(v_i, v_j))_{1 \le i,j \le 4} \in \mathcal{O}_L.$$

This determinant depends on the choice of the basis v_i for $\oplus_{i=1}^4 \mathcal{O}_L v_i$ only up to squares of units of $\mathcal{O}_L^\times$. We define $\mathfrak{d}_\mathcal{O}$, the **discriminant** of $\mathcal{O}$, to be the $\mathcal{O}_L$-ideal whose square is the ideal generated by $d(v_1, \ldots, v_4)$ as $\{v_1, \ldots, v_4\}$ runs over all L-bases of B in $\mathcal{O}$.[1]

If $\mathcal{O}' \subset \mathcal{O}$ is a suborder then $\mathfrak{d}_{\mathcal{O}'} = [\mathcal{O} : \mathcal{O}']^2 \cdot \mathfrak{d}_\mathcal{O}$. It follows that every order is contained in a maximal order. It is a theorem that the discriminant of any maximal order is the product of all the finite ramified places of the algebra B.

It is possible for the discriminant to be $\mathcal{O}_L$ and yet for B not to be isomorphic to $M_2(L)$. For example, take L to be a real quadratic field in which p is inert. Then $B_{p,\infty} \otimes_\mathbb{Q} L$ is ramified precisely at the two infinite primes ∞_1, ∞_2 of L and at no finite prime.

---

[1]The usual definition of discriminant uses the form $\frac{1}{2}\mathrm{Tr}(ab)$ and not $\frac{1}{2}\mathrm{Tr}(a\bar{b})$, [Vig, p. 23 ff.], [Voi, §15.2]. However, using the form $\frac{1}{2}\mathrm{Tr}(a\bar{b})$ gives the same discriminant ideal. See [Voi, Ex. 13, p. 254].

2.2.1. *$B_{p,\infty}$ again.* For $p \equiv 3 \pmod 4$, an example of a maximal order of $B_{p,\infty} = \left(\frac{-1,-p}{\mathbb{Q}}\right)$ is given by

$$\mathbb{Z} \oplus \mathbb{Z} \cdot i \oplus \mathbb{Z} \cdot \frac{i+j}{2} \oplus \mathbb{Z} \cdot \frac{1+k}{2}$$

([Vig, p. 98]). The matrix $(2(v_i, v_j))_{1 \le i,j \le 4}$ associated to this basis is

$$\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & (p+1)/2 & 0 \\ 1 & 0 & 0 & (p+1)/2 \end{pmatrix} ;$$

its determinant is p^2, and so the discriminant of the order is (p) as expected.

If $\mathcal{O}$ is any maximal order of a quaternion algebra over a number field L, then, for any finite place v of L, the base change to $\mathcal{O}_{L_v}$ of the quadratic space $(\mathcal{O}, \mathrm{Nm})$ is isomorphic to $(M_2(\mathcal{O}_{L_v}), \det)$ if v is unramified. If v is ramified then the corresponding base change is isomorphic to

$$(2.1) \qquad \left(\left\{ \begin{pmatrix} a & \pi_v b \\ b^\sigma & a^\sigma \end{pmatrix} : a, b \in \mathcal{O}_{L_{v,2}} \right\}, \det \right),$$

where π_v is a uniformizer of L_v and $L_{v,2}$ is the quadratic unramified extension of L_v. In other words, this base change is isomorphic to $((\mathcal{O}_{L_{v,2}} \oplus \mathcal{O}_{L_{v,2}}), aa^\sigma - \pi_v bb^\sigma)$. Choosing a basis for $\mathcal{O}_{L_{v,2}}$ over $\mathcal{O}_{L_v}$, one may write this as a quadratic rank 4 $\mathcal{O}_{L_v}$-module.

2.2.2. *Ideals.* Let $\mathcal{O}$ be an order in a quaternion algebra B over L, a number field or a finite extension of $\mathbb{Q}_p$. A **right** (respectively, left) **ideal** I of $\mathcal{O}$ is a rank 4 $\mathcal{O}_L$-module that contains a basis of B over L, such that

$$I \cdot \mathcal{O} \subseteq I \quad (\text{respectively, } \mathcal{O} \cdot I \subseteq I).$$

For a rank 4 $\mathcal{O}_L$-module I that contains a basis of B over L, define its **left** (respectively, right) **order** by

$$\mathcal{O}_\ell(I) = \{b \in B : bI \subseteq I\}, \quad (\text{respectively, } \mathcal{O}_r(I) = \{b \in B : Ib \subseteq I\}).$$

These are indeed orders of B.

If I is a left $\mathcal{O}$-ideal then $\mathcal{O}_\ell(I) \supseteq \mathcal{O}$. If $\mathcal{O}$ is a maximal order and I is a left ideal of $\mathcal{O}$, then $\mathcal{O}_\ell(I) = \mathcal{O}$, while $\mathcal{O}_r(I)$ is a maximal order that in general is not even conjugate to $\mathcal{O}$ in B. Analogous statements hold for right ideals. An ideal J of $\mathcal{O}$ is called a **two-sided ideal** if $\mathcal{O} = \mathcal{O}_\ell(J) = \mathcal{O}_r(J)$.

2.2.3. *Ideal classes.* Let B be a quaternion algebra over L (still a number field or a finite extension of $\mathbb{Q}_p$) and $\mathcal{O}$ a maximal order of B. If I is a right $\mathcal{O}$-ideal, then so is λI for any $\lambda \in B^\times$. The relation $I \sim \lambda I$ is an equivalence relation on the set of right $\mathcal{O}$-ideals. The equivalence classes are called the **ideal classes** and we denote this by $\mathrm{Cl}(\mathcal{O})$; it is a set with a distinguished element, namely the class of $\mathcal{O}$, but in general is not a group in any natural way. A fundamental theorem states that the number of ideal classes is finite, and we refer to this number as the **class number** of $\mathcal{O}$. It is a theorem that the class number is independent of the choice of maximal order $\mathcal{O}$. We shall denote it h_B, or h if the quaternion algebra B is clear from context. More precisely:

(1) If L is a finite extension of $\mathbb{Q}_p$ and $B = M_2(L)$ then a maximal order is given by $M_2(\mathcal{O}_L)$. All maximal orders in B are conjugate in B, and each has class number 1. If B is the ramified quaternion algebra over L (recall that such B is unique up to isomorphism), then there is a unique maximal order, given by (2.1), and again the class number is 1.

(2) If L is a number field, in general the class number is not 1, but sometimes it is. If $L = \mathbb{Q}$ and $B = B_{p,\infty}$ then the class number $h_{B_{p,\infty}}$ is equal to
   (a) $\lfloor \frac{p}{12} \rfloor$ if $p \equiv 1 \pmod{12}$;
   (b) $\lfloor \frac{p}{12} \rfloor + 1$ if $p \equiv 5, 7 \pmod{12}$;

(c) $\lfloor \frac{p}{12} \rfloor + 2$ if $p \equiv 11 \pmod{12}$;

(d) 1 if $p = 2, 3$.

(This follows from [Voi, Theorem 30.1.5], as explicated by [Voi, Chapter 30 Exercise 6].)

Given a maximal order $\mathcal{O}$, let $I_1, \ldots, I_h$ be representatives for the right ideal classes of $\mathcal{O}$. Let $\mathcal{O}_i := \mathcal{O}_\ell(I_i)$. Then every maximal order of B is conjugate to one of the orders $\mathcal{O}_i$. In general, the number of maximal orders in B, up to conjugation in B (equivalently, up to isomorphism of rings) is called the **type number** t_B of B. Therefore, given these results, we see that the type number is smaller or equal to the class number. One can be more precise:

Consider the collection of two-sided ideals of $\mathcal{O}$. These ideals, considered up to equivalence $I \sim \lambda I$, where now $\lambda \in L^\times$, form a group $\mathrm{biCl}(\mathcal{O})$ that for L a number field of class number 1 is an abelian group of order 2^f, where f is the number of finite ramified places of B. In fact, for every choice of $\epsilon_v \in \{0, 1\}$ for the finite ramified places v of B, there is a two sided ideal I of $\mathcal{O}$ characterized by the fact that for each finite ramified place of B we have

$$I_v = \left\{ \begin{pmatrix} a & \pi_v b \\ b^\sigma & a^\sigma \end{pmatrix} : a, b \in \mathcal{O}_{L_v}, \ \pi_v^{\epsilon_v} | a \right\}$$

and we have $I_v = \mathcal{O}_v$ for all other places. Each equivalence class of two-sided ideals contains exactly one of these ideals, and this provides an isomorphism $\mathrm{biCl}(\mathcal{O}) \cong (\mathbb{Z}/2\mathbb{Z})^f$. This holds for every choice of maximal order $\mathcal{O}$. For any fixed $\mathcal{O}$, there is a map of pointed sets

$$\mathrm{biCl}(\mathcal{O}) \to \mathrm{Cl}(\mathcal{O}),$$

that may have a kernel, and this kernel may very well depend on the choice of $\mathcal{O}$. For example, for $\mathcal{O}$ a maximal order in $B_{p,\infty}$ we have $\mathrm{biCl}(\mathcal{O}) \cong \mathbb{Z}/2\mathbb{Z}$, and its image in $\mathrm{Cl}(\mathcal{O})$ is trivial if and only if $\mathcal{O}$ has an element of norm p.

Returning to our choice of representatives $I_1, \ldots, I_h$ for $\mathrm{Cl}(\mathcal{O})$, suppose that O_i is conjugate to O_j for some $i \neq j$, say $\mathcal{O}_i = \alpha \mathcal{O}_j \alpha^{-1}$. By changing the ideal I_j to αI_j (which does not change the right ideal class), we may assume without loss of generality that $\mathcal{O}_i = \mathcal{O}_j$. One can prove that this is the case if and only if $I_i = J I_j$ for some two-sided ideal J of $\mathcal{O}_j$ (see [Piz1, Proposition 2.18], [Vig, p. 87 ff.], [Voi, §§18.4-18.5]). Note that necessarily J is not principal in $\mathrm{Cl}(\mathcal{O})$. Thus, such a coincidence $\mathcal{O}_i = \mathcal{O}_j$ can happen only if the image of $\mathrm{biCl}(\mathcal{O}_j)$ in $\mathrm{Cl}(\mathcal{O}_j)$ is not trivial, and conversely, elements of $\mathrm{biCl}(\mathcal{O}_j)$ that are not principal ideals create such coincidences. For $B = B_{p,\infty}$ one concludes the formula

$$h_B = 2t_B - u_B,$$

where u_B is the number of maximal orders of B, considered up to conjugation, containing an element of norm p. These maximal orders counted by u_B are discussed further in §3.2.2.

2.3. **Supersingular elliptic curves.** General references for this section are [Sil1], [Was]. We also assume some familiarity with algebraic geometry, mostly at the level of [Sil1], but occasionally a bit more.

We fix an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$, and denote by $\mathbb{F}_{p^r}$ the unique finite field with p^r elements in $\overline{\mathbb{F}}_p$.

2.3.1. *Weierstrass equations.* Let k be a field. An **elliptic curve** over k is a proper non-singular geometrically-irreducible curve E of genus 1 with a specified k-rational point denoted 0_E. Using the Riemann-Roch theorem, one can prove that if $\mathrm{char}(k) \neq 2$ then such a curve can be given by an equation in $\mathbb{P}_k^2$ of the form

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

for some separable polynomial $f(x) \in k[x]$ of degree 3. In fact, if $\mathrm{char}(k) \neq 2, 3$ one may take

(2.2) $$f(x) := x^3 + ax + b \in k[x].$$

This is called a **short Weierstrass equation**. To be precise, it is the closure of this curve in the projective plane $\mathbb{P}^2_k$ that is the elliptic curve, and we take 0_E to be the unique point at infinity given by $(x : y : z) = (0 : 1 : 0)$. A slightly more complicated equation, likewise obtained by using the Riemann-Roch theorem, may be needed over fields of characteristic 2 or 3.

An elliptic curve E has a j-**invariant** $j(E)$; if E is given by the equation $y^2 = x^3 + ax + b$, then

$$(2.3) \qquad\qquad j(E) = 1728 \cdot \frac{4a^3}{27b^2 + 4a^3}.$$

If there is an isomorphism over k between E_1 and E_2 then $j(E_1) = j(E_2)$, and when k is algebraically closed the converse also holds. Furthermore, for any field k and element $j \in k$ there is an elliptic curve E over k with j-invariant $j(E) = j$. In fact, if $j \neq 0, 1728$, one may take

$$E \colon y^2 + xy = x^3 - \frac{36}{j - 1728} - \frac{1}{j - 1728}$$

[Sil1, §III, Proposition 1.4]. See §2.3.4 for the cases $j = 0, 1728$.

An elliptic curve is an abelian algebraic group; that is, there are morphisms

$$m \colon E \times E \to E, \quad inv \colon E \to E,$$

that make $E(k')$ into a group for any extension k'/k, where $m(P, Q)$ is the sum of $P, Q \in E(k')$ (we shall also write this as $P + Q$) and $inv(P)$ is the negative of P (written as $-P$). The zero point of this group is 0_E, and three points $P, Q, R \in E(k')$ are the intersection points of E with a line (counted with multiplicity if necessary) if and only if $P + Q + R = 0_E$.

2.3.2. *Isogenies, degrees, kernels.* It is a theorem that every morphism $f \colon E_1 \to E_2$ of elliptic curves satisfying $f(0_{E_1}) = 0_{E_2}$ is also a group homomorphism. If f is not the zero map then it is called an **isogeny**. An isogeny f – being a non-constant morphism of irreducible curves – must be surjective as a map of varieties (though the induced map $E_1(k') \to E_2(k')$ for k'/k need not be surjective unless k' is algebraically closed). Given an isogeny f, we may define its **kernel** $\mathrm{Ker}(f)$ to be the fiber product

$$
\begin{array}{ccc}
\mathrm{Ker}(f) & \longrightarrow & \{0_{E_2}\} \\
\downarrow & & \downarrow \\
E_1 & \xrightarrow{\ f\ } & E_2.
\end{array}
$$

As f is a non-constant morphism of curves, this fiber product is Spec of a k-algebra which is finite-dimensional as a k-vector space; this dimension is called the **rank** of the finite group scheme $\mathrm{Ker}(f)$. The **degree** $\deg(f)$ of the isogeny f is defined to be the rank of $\mathrm{Ker}(f)$. This is also equal to the degree of the extension of function fields $[k(E_1) : f^*k(E_2)]$. If $\bar{k}$ is an algebraically closed field then the $\bar{k}$-points $\mathrm{Ker}(f)(\bar{k})$ of $\mathrm{Ker}(f)$ form an abelian group of cardinality dividing $\deg(f)$, but equality does not always hold. Equality does hold in the cases where $\mathrm{char}(k) = 0$, or where $\mathrm{char}(k) = p > 0$ and $\gcd(p, \deg(f)) = 1$.

As E is an abelian group, for any integer n the **multiplication by** n map, denoted $[n]$, is an **endomorphism** (a morphism from E to itself fixing 0_E) and, in fact, an isogeny if $n \neq 0$. The kernel is denoted in this case $E[n]$ and it is always a group scheme of degree n^2. If $n = -1$ this map is given by $(x, y) \mapsto (x, -y)$ in terms of the short Weierstrass equation (2.2). Since a point P is of order 2 if and only if $P = -P$, we see that

$$E[2] = \{0_E, (\alpha_1, 0), (\alpha_2, 0), (\alpha_3, 0)\},$$

where $\alpha_1, \dots, \alpha_3$ are the roots of $f(x)$.

For an extension k'/k, the collection of homomorphisms $\mathrm{Hom}_{k'}(E_1, E_2)$ (that is, the collection of isogenies $E_1 \to E_2$ defined over k' together with the zero map) is an abelian group, and the collection of endomorphisms $\mathrm{End}_{k'}(E_1)$ is a ring with multiplication given by composition.

If $\mathrm{Hom}_{k'}(E_1, E_2)$ is nontrivial then it is a projective rank 1 module over $\mathrm{End}_{k'}(E)$. The degree map is a quadratic form $\mathrm{Hom}_k(E_1, E_2) \to \mathbb{Z}$. When we write $\mathrm{Hom}(E_1, E_2)$ or $\mathrm{End}(E_1)$ without subscripts, we mean $\mathrm{Hom}_{\bar{k}}(E_1, E_2)$ and $\mathrm{End}_{\bar{k}}(E_1)$, where $\bar{k}$ is an algebraic closure of k.

If H is a finite subgroup scheme of E then the quotient scheme E/H exists and is an elliptic curve whose $\bar{k}$ points are $E(\bar{k})/H(\bar{k})$. The morphism $E \to E/H$ is an isogeny with kernel H. If $f: E_1 \to E_2$ is an isogeny with kernel H, then there is a natural isomorphism $E_1/H \to E_2$ such that the following diagram commutes:

(2.4)
$$E_1 \xrightarrow{\ \ f\ \ } E_2$$
$$E_1/H.$$

with $\cong$ indicated.

Similarly, analogues of all the main isomorphism theorems from group theory hold for elliptic curves.

To every isogeny $f: E_1 \to E_2$ there is an associated **dual isogeny** $f^\vee: E_2 \to E_1$ uniquely determined by the property

$$f^\vee \circ f = [\deg(f)]$$

(meaning, the composition is the multiplication by the integer $\deg(f)$ on E_1). It is also true that $\deg(f) = \deg(f^\vee)$ and $f^{\vee\vee} = f$. Moreover, mapping an isogeny to its dual gives an isomorphism of quadratic modules $\mathrm{Hom}(E_1, E_2) \to \mathrm{Hom}(E_2, E_1)$. On $\mathrm{End}(E)$ this map is an anti-isomorphism of rings, meaning $(f \circ g)^\vee = g^\vee \circ f^\vee$. In particular, these statements imply the fact that $[n]^\vee = [n]$.

2.3.3. *Complex multiplication.* If k is a characteristic zero field and $E: y^2 = x^3 + ax + b$ is an elliptic curve over k, one can embed $\mathbb{Q}(a, b) \hookrightarrow \mathbb{C}$ and so one may view E as a complex elliptic curve. Let $\mathcal{T}_E$ denote that tangent space of E at 0_E. By the theory of complex Lie groups, the exponential map $\mathcal{T}_E \to E(\mathbb{C})$ is a surjective analytic homomorphism whose kernel is a full lattice Λ in $\mathcal{T}_E$; that is, a rank 2 discrete subgroup of $\mathcal{T}_E \cong \mathbb{C}$. By a homothety of $\mathbb{C}$, one may assume that $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$ where $\tau \in \mathbb{C}$ is such that $\mathrm{Im}(\tau) > 0$. Since we have an isomorphism

$$\mathbb{C}/\Lambda \cong E(\mathbb{C}),$$

we see that $E[n] \cong \frac{1}{n}\Lambda/\Lambda$, a group of cardinality n^2. Drawing further from the theory of complex Lie groups, we find that for elliptic curves $E_i \cong \mathbb{C}/\Lambda_i$,

$$\mathrm{Hom}(E_1, E_2) = \{\alpha \in \mathbb{C} : \alpha\Lambda_1 \subset \Lambda_2\}.$$

From this, it is not hard to conclude that if $E \cong \mathbb{C}/\mathbb{Z} \oplus \mathbb{Z}\tau$ is an elliptic curve over a field k of characteristic 0, then

$$\mathrm{End}(E) \cong \begin{cases} \text{an order } R \text{ of } \mathbb{Q}(\tau) & \tau \text{ quadratic over } \mathbb{Q} \\ \mathbb{Z} & \text{otherwise.} \end{cases}$$

In case $\mathrm{End}(E) = R$ is a quadratic order, one says that E has **complex multiplication**.

2.3.4. *The j-invariants 0 and 1728.* Over $\mathbb{C}$, the elliptic curve

$$E_{1728}: y^2 = x^3 + x,$$

has an automorphism $\varphi: (x, y) \mapsto (-x, iy)$. Note that $\varphi^2 = [-1]$ and so, given that $\mathrm{End}(E_{1728})$ is at most a quadratic order, we must have $\mathrm{End}(E) \cong \mathbb{Z}[i]$ (where φ corresponds to i). The j-invariant of E_{1728} is 1728. Note that the same equation defines an elliptic curve over $\mathbb{F}_p$ for any $p \neq 2$ (but the endomorphism ring may become larger).

Over $\mathbb{C}$, the elliptic curve

$$E_0 \colon y^2 = x^3 + 1,$$

has an automorphism $\varphi \colon (x, y) \mapsto (\omega x, y)$, where $\omega = e^{2\pi i/3}$ is a primitive third root of unity. Note that $\varphi^3 = [1]$ and so, given that $\mathrm{End}(E_0)$ is at most a quadratic order, we must have $\mathrm{End}(E) \cong \mathbb{Z}[\omega]$ (where φ corresponds to ω and $\omega^2 + \omega + 1 = 0$). The j-invariant of E_0 is 0. This equation is good also over $\mathbb{F}_p$ for any $p > 3$ (though again the endomorphism ring may be larger).

In a field of characteristic 2 or 3 we have $0 = 1728$. The equation $y^2 + y = x^3$ defines an elliptic curve of j-invariant 0 over $\mathbb{F}_2$, and $y^2 = x^3 + x$ defines an elliptic curve of j-invariant 0 over $\mathbb{F}_3$.

2.3.5. *Endomorphisms of elliptic curves over fields of positive characteristic.* Let k be a field of positive characteristic p and let E be an elliptic curve over k. In this case, there are exactly 3 possibilities for $\mathrm{End}_k(E)$.

$$\mathrm{End}_k(E) \cong \begin{cases} \mathbb{Z}, & \\ R, & \text{a quadratic imaginary order,} \\ \mathcal{O}, & \text{a maximal order of } B_{p,\infty}. \end{cases}$$

In fact, the quadratic imaginary orders R that arise have conductor prime to p and p splits in R, and, conversely, every such R arises as $\mathrm{End}_k(E)$ for some elliptic curve E over some finite field k. Likewise, every maximal order of $B_{p,\infty}$ is isomorphic to some $\mathrm{End}_k(E)$. For elliptic curves E over a field k of positive characteristic, it is a fact that $\mathrm{End}_k(E) \neq \mathbb{Z}$ if and only if E can be defined over a finite field (i.e. there exists a finite subfield $\mathbb{F}_{p^r}$ in k and an elliptic curve E' over $\mathbb{F}_{p^r}$ such that E is the base change of E' to k). If $\mathrm{End}_{\bar{k}}(E)$ is a maximal order of $B_{p,\infty}$ then E is called **supersingular**, and otherwise E is called **ordinary**.

To fix ideas, assume that E is given by a Weierstrass equation $y^2 = x^3 + ax + b$. Then $E^{(p)}$, the **Frobenius base-change** of E, is the elliptic curve given by

$$(2.5) \qquad\qquad E^{(p)} \colon y^2 = x^3 + a^p x + b^p.$$

In fact, scheme-theoretically, $E^{(p)} \cong E \otimes_{\mathrm{Spec}(k),\varphi} \mathrm{Spec}(k)$, where $\varphi(z) = z^p$ is the Frobenius homomorphism of k. This observation shows that the construction can be carried in complete generality for any k-scheme. We may likewise define $E^{(p^n)}$ for any $n \geq 1$ by applying Frobenius base-change n times.

The canonical morphism

$$(2.6) \qquad\qquad \mathrm{Fr} \colon E \to E^{(p)}, \quad \mathrm{Fr}(x, y) = (x^p, y^p),$$

takes 0_E to $0_{E^{(p)}}$ and is therefore an isogeny. Its degree is p. Its dual isogeny $\mathrm{Fr}^\vee$ is called the **Verschiebung** isogeny,

$$\mathrm{Ver} \colon E^{(p)} \to E,$$

and it satisfies $\mathrm{Ver} \circ \mathrm{Fr} = [p]$ on E and $\mathrm{Fr} \circ \mathrm{Ver} = [p]$ on $E^{(p)}$. When we need to be clear about the starting curve we write $\mathrm{Fr}_E \colon E \to E^{(p)}$, and we define

$$\mathrm{Fr}_E^n := \mathrm{Fr}_{E^{(p^{n-1})}} \circ \cdots \circ \mathrm{Fr}_{E^{(p)}} \circ \mathrm{Fr}_E,$$

an isogeny of degree p^n from E to $E^{(p^n)}$.

A supersingular elliptic curve E_1 over k has a unique subgroup scheme of rank p (while an ordinary elliptic curve has two). This implies that any isogeny $f \colon E_1 \to E_2$ of degree p fits into a

commutative diagram

(2.7)

where γ is an isomorphism.

2.3.6. *Equivalent definitions of supersingularity.* There are many ways to characterize supersingularity. We assume that E is defined over $\mathbb{F}_{p^r} \subset \overline{\mathbb{F}}_p$. The following are equivalent:

   (1) E is supersingular ($\mathrm{End}_{\overline{\mathbb{F}}_p}(E)$ is isomorphic to a maximal order $\mathcal{O}$ of $B_{p,\infty}$).
   (2) $E[p](\overline{\mathbb{F}}_p) = \{0_E\}$ (there are no non-trivial p-torsion points).
   (3) $E[p^n](\overline{\mathbb{F}}_p) = \{0_E\}$ for some (or all) $n \geq 1$.
   (4) The group scheme $\alpha_p := \mathrm{Ker}(\mathrm{Fr} : \mathbb{G}_a \to \mathbb{G}_a)$ embeds into E over $\overline{\mathbb{F}}_p$.[2]
   (5) E can be defined over $\mathbb{F}_{p^2}$ and, identifying E with $E^{(p^2)}$, we have $\mathrm{Fr}^2 = \epsilon \circ [p]$ for some automorphism ϵ of E.
   (6) The morphism $\mathrm{Ver} : E^{(p)} \to E$ is inseparable, meaning the extension of fields $\mathrm{Ver}^* k(E) \subset k(E^{(p)})$ is inseparable (of degree p).
   (7) Assuming $p > 2$, let $m = (p-1)/2$. Writing the elliptic curve E in **Legendre form** $E : y^2 = x(x-1)(x-\lambda)$ (which is always possible if one allows a finite extension of the base field), λ is a root of the polynomial

$$H_p(t) = \sum_{i=0}^{m} \binom{m}{i}^2 t^i.$$

   (If $p = 2$ then E is supersingular if and only if $j(E) = 0$.)
   (8) For every curve E' that is isogenous to E over $\overline{\mathbb{F}}_p$, $j(E') \in \mathbb{F}_{p^2}$.[3]
   (9) For $p > 2$, if E is defined over $\mathbb{F}_q$ and $\sharp(E(\mathbb{F}_q)) = q + 1 - A$, then $A \equiv 0 \pmod{p}$.[4]

The equivalence of most of these conditions is proven in [Sil1, Theorem V.3.1, Theorem V.4.1]. For condition (4), which will be transparent to the reader versed in the theory of finite commutative group schemes, see for instance [Ulm, Proposition 2.1].

Note that the degree of H_p in definition (7) is $(p-1)/2$. On the other hand, we have [Sil1, §III, Proposition 1.7]

$$j(E) = 2^8 \frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}$$

and so the association $\lambda \mapsto j$ is "usually" $6 : 1$. This suggests that the number of supersingular j-invariants, namely, the number of supersingular elliptic curves up to isomorphism over $\overline{\mathbb{F}}_p$, is roughly $(p-1)/12$. This counting can be made precise ([Sil1, §V, Theorem 4.1]) and one obtains that the number of supersingular elliptic curves is precisely the class number $h_{B_{p,\infty}}$ of $B_{p,\infty}$ given in §2.2.3. In fact there is an explicit correspondence between isomorphism classes of supersingular curves and ideal classes of a maximal order in $B_{p,\infty}$, which is described in §3.2.7.

---

[2] The group scheme α_p is the group scheme whose R points, for every commutative $\mathbb{F}_p$-algebra R, are $\{x \in R : x^p = 0\}$, with the usual addition law.

[3] One can make this into an effective criterion for determining whether a given elliptic curve E is supersingular or not; see [Sut1].

[4] This condition yields an effective polynomial time test for supersingularity using Schoof's algorithm [Sch].

If E_1 is a supersingular elliptic curve and $f\colon E_1 \to E_2$ is an isogeny, then E_2 is also supersingular. This follows immediately from (8), but it may also be instructive to see how this follows from (2). Suppose there exists a nonzero point $Q \in E_2[p]$. For any $n \geq 1$, the map $[p^{n-1}]$ is an isogeny, and so Q can be written as $p^{n-1}P$ for some point $P \in E_2(\overline{\mathbb{F}}_p)$ of exact order p^n. Take n such that $p^n > \deg(f)$. Note that $f^\vee(P)$ is a point on E_1 that is annihilated by $[p^n]$, and therefore must be 0_{E_1} since E_1 is supersingular. But this implies $f(f^\vee(P)) = [\deg f](P) = 0_{E_2}$, a contradiction since P has order $p^n > \deg(f)$.

2.3.7. *Models of supersingular elliptic curves over* $\mathbb{F}_{p^2}$. The j-invariant of a supersingular elliptic curve E lies in $\mathbb{F}_{p^2}$ and so E has a model over $\mathbb{F}_{p^2}$. In such a model we have $\mathrm{Fr}^2 = \epsilon \circ [p]$ for some automorphism ϵ of E. It turns out [BGJ+] that one can always find such a model where $\mathrm{Fr}^2 = [-p]$.

For any such models E_1, E_2 we have

$$\mathrm{Hom}_{\mathbb{F}_{p^2}}(E_1, E_2) = \mathrm{Hom}_{\overline{\mathbb{F}}_p}(E_1, E_2),$$

that is, all isogenies are defined over $\mathbb{F}_{p^2}$. Indeed, if $f \in \mathrm{Hom}_{\overline{\mathbb{F}}_p}(E_1, E_2)$ then so is $f^{(p^2)}$, the base-change of f by the square of Frobenius. This base change is characterized by the property that $f^{(p^2)}(\mathrm{Fr}^2 P) = \mathrm{Fr}^2(f(P))$ for any point P on E_1. But,

$$f^{(p^2)}(\mathrm{Fr}^2 P) = f^{(p^2)}(-pP) = -pf^{(p^2)}(P)$$

and $\mathrm{Fr}^2(f(P)) = -pf(P)$. As $[-p]$ is an isogeny, and $-p(f^{(p^2)} - f) = 0$, it follows that $f = f^{(p^2)}$ and thus $f \in \mathrm{Hom}_{\mathbb{F}_{p^2}}(E_1, E_2)$.[5]

2.4. **Reducing curves with complex multiplication modulo primes.** [Lan, Ser, Sil2] may be consulted for a complete overview. A powerful method of finding supersingular elliptic curves is the reduction of elliptic curves with complex multiplication (as defined in §2.3.3). Let E be an elliptic curve over $\mathbb{C}$ such that $\mathrm{End}(E) = \mathcal{O}_K$, the ring of integers of a quadratic imaginary field K. The complex uniformization $\mathbb{C}/\Lambda \cong E(\mathbb{C})$ (see §2.3.3) shows that Λ must be an $\mathcal{O}_K$-module. Being of rank 2 and torsion free, Λ must be a projective $\mathcal{O}_K$-module of rank 1 and thus its isomorphism class is an ideal class of $\mathcal{O}_K$. Conversely, given an ideal I of $\mathcal{O}_K$ and an inclusion $K \to \mathbb{C}$, we may view I as an $\mathcal{O}_K$-invariant lattice in $\mathbb{C}$ and get an elliptic curve with complex multiplication as $\mathbb{C}/I$. Fixing an inclusion $K \to \mathbb{C}$, one obtains a bijection between the class group $\mathrm{Cl}(\mathcal{O}_K)$ and isomorphism classes of elliptic curves with complex multiplication such that action of K on their tangent space is given by the fixed inclusion $K \to \mathbb{C}$.

Choose representatives $I_1, \ldots, I_h$ for the ideal classes of K. Let E_i be the elliptic curves such that $E_i(\mathbb{C}) \cong \mathbb{C}/I_i$.

One of the main theorems of complex multiplication is that $j(E_i)$ belongs to the Hilbert class field H_K of K. The Hilbert class field is the maximal unramified extension of K with abelian Galois group $\mathrm{Gal}(H_K/K)$, and this Galois group is isomorphic to $\mathrm{Cl}(\mathcal{O}_K)$. This theorem states further that $\mathrm{Gal}(H_K/H)$ acts simply transitively on $j(E_1), \ldots, j(E_h)$ and it follows that $H_K = K(j(E_i))$ for any i. In particular, if E has CM by $\mathcal{O}_K$ then E has a model over H_K.

For a number field L, every elliptic curve E/L with complex multiplication has **potential good reduction** everywhere. One interpretation of this statement is that there is a finite extension M of L such that E has a general Weierstrass equation $y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$, $a_i \in \mathcal{O}_{k_1}$, such that modulo every finite place v of M, the reduction of this equation modulo v still defines a non-singular (elliptic) curve $E \pmod v$. Let p be the rational prime determined by v. Then $E \pmod v$ is an elliptic curve over a finite field $\mathbb{F}_{p^r}$ (with r depending on v) and we

---

[5]Taking an appropriate twist, each supersingular curve also has a model over $\mathbb{F}_{p^2}$ satisfying $\mathrm{Fr}^2 = [p]$, and for such models we also have $\mathrm{Hom}_{\mathbb{F}_{p^2}}(E_1, E_2) = \mathrm{Hom}_{\overline{\mathbb{F}}_p}(E_1, E_2)$. This gives us an alternate setup, but we will not use this.

can ask whether it is ordinary or supersingular. It is a theorem that $E \pmod{v}$ is supersingular if p is inert or ramified in K, and it is ordinary if p is split in K.

As an example, consider E_0 that has CM by $\mathbb{Z}[\omega]$ (see §2.3.4). The only prime ramifying in $\mathbb{Z}[\omega]$ is 3. The prime 2 is inert and a prime $p > 3$ is split if and only if there are non-trivial third roots of unity in $\mathbb{F}_p$, which is the case if and only if $p \equiv 1 \pmod 3$. Thus, the primes congruent to 1 mod 3 split, and those congruent to 2 mod 3 are inert. In particular, 0 is a supersingular j-invariant for $p = 2, 3$ and all $p > 3$ congruent to 2 mod 3.

Similarly, E_{1728} has CM by $\mathbb{Z}[i]$. The prime 2 ramifies, the primes $p \equiv 3 \pmod 4$ are inert and the primes $p \equiv 1 \pmod 4$ are split. Consequently, 1728 is a supersingular j-invariant if $p = 2$ or $p \equiv 3 \pmod 4$. Working mod 12 we see that we found explicit supersingular j-invariants if $p = 2, 3$ or $p = 5, 7, 11 \pmod{12}$. This leaves the primes p that are congruent to 1 mod 12. For such primes one uses ad hoc methods that depend on the prime, for instance by looking for an imaginary quadratic field with small discriminant in which p is inert. One could alternatively work with CM elliptic curves for the field $\mathbb{Q}(\sqrt{-p})$. Note though that if p is of cryptographic size $p \approx 2^{500}$, then by the Brauer-Siegel theorem the class number of $\mathbb{Q}(\sqrt{-p})$ is approximately of size $\sqrt{p} \approx 2^{250}$. Thus, the minimal field of definition of such a CM curve would be of degree $\approx 2^{250}$ over $\mathbb{Q}$, which is completely infeasible to do computations with.

2.4.1. *Lifting theorems.* We now discuss a converse to the reduction of elliptic curves with CM. Let E be an elliptic curve over $\mathbb{F}_{p^r}$ and let $f \in \operatorname{End}(E)$ be a non-integer endomorphism of E. Let $x^2 + bx + c$ be the characteristic polynomial of f as an element of the ring $\operatorname{End}(E) \subset \mathbb{Z}$. Deuring's lifting theorem says that there is an elliptic curve $\tilde{E}$ over the Witt ring $W(\mathbb{F}_{p^r})$, or a ramified quadratic extension of it, and $\tilde{f} \in \operatorname{End}(\tilde{E})$ with characteristic polynomial $x^2 + bx + c$, such that the reduction of $(\tilde{E}, \tilde{f})$ modulo the uniformizer of $W(\mathbb{F}_{p^r})$ is (E, f); see [GZ, Proposition 2.7], [Lan, 13, §5]. A simpler statement, which the above theorem can be used to prove, is that every elliptic curve E over a finite field is a reduction of an elliptic curve over a number field with complex multiplication.

One can ask a related question: consider a quadratic imaginary field K of class number h_K, and elliptic curves $\tilde{E}_1, \ldots, \tilde{E}_{h_K}$ with CM by $\mathcal{O}_K$, all defined over a number field k and having good reduction at a prime v of k over p. Let $E_i = \tilde{E}_i \pmod{v}$ and assume that p is inert in K. If $h_K \gg p$, is every supersingular elliptic curve over $\overline{\mathbb{F}}_p$ isomorphic to some E_i? By Deuring's lifting theorem, this is equivalent to asking whether every maximal order $\mathcal{O}$ of $B_{p,\infty}$ admits an embedding of $\mathcal{O}_K \to \mathcal{O}$. This is true for h_K sufficiently large, and follows from a stronger result due to Elkies, Ono and Yang [EOY].

2.5. $\operatorname{Hom}(E_1, E_2)$ **as a quadratic module.** Let E_1, E_2 be supersingular elliptic curves over $\mathbb{F}_{p^2}$, and recall that we denote $\operatorname{Hom}(E_1, E_2) := \operatorname{Hom}_{\overline{\mathbb{F}}_p}(E_1, E_2)$. If we set the degree of the zero homomorphism to be 0, then the degree function

$$\deg \colon \operatorname{Hom}(E_1, E_2) \to \mathbb{Z},$$

is a quadratic form with associated bilinear form

$$(f, g) = \frac{1}{2} \left(\deg(f + g) - \deg(f) - \deg(g) \right).$$

Note that we can interpret $\deg(f)$ as $f^\vee f$, where $f^\vee$ is the dual isogeny. Then, using also that $(f + g)^\vee = f^\vee + g^\vee$, we find

$$(f, g) = \frac{1}{2} \left(g^\vee f + f^\vee g \right).$$

It is easy to check that this is a bilinear form when written in this way.

Fixing a $\mathbb{Z}$-basis $v_1, \ldots, v_4$ for $\operatorname{Hom}(E_1, E_2)$, the matrix $A = (2(v_i, v_j))_{i,j=1}^4$ has determinant p^2, so the quadratic form $\deg$ has discriminant p^2. We say that $\deg$ has **level** (or *stufe*) p because p is

the smallest positive integer with the property that pA^{-1} has integer entries. For example, take $E_1 = E_2$ such that $\operatorname{End}(E_i)$ is the maximal order given in §2.2.1. The determinant of the matrix given there is p^2 and its inverse is

$$\frac{1}{p} \begin{pmatrix} \frac{p+1}{2} & 0 & 0 & -1 \\ 0 & \frac{p+1}{2} & -1 & 0 \\ 0 & -1 & 2 & 0 \\ -1 & 0 & 0 & 2 \end{pmatrix}.$$

Recall that $\operatorname{End}_{\overline{\mathbb{F}}_p}(E_1)$ is isomorphic to a maximal order $\mathcal{O}$ in $B_{p,\infty}$. As $\operatorname{Hom}(E_1, E_2)$ is a right projective rank 1 module over $\operatorname{End}_{\overline{\mathbb{F}}_p}(E_1)$, it is isomorphic as a right $\mathcal{O}$-module to some ideal I that is only determined up to equivalence, $I \sim \lambda I$ for some $\lambda \in B_{p,\infty}^{\times}$. One can show that under such an isomorphism deg corresponds to $\frac{1}{\operatorname{Nm}(I)} \operatorname{Nm}(\cdot)$, a formula that indeed depends only on the ideal class of I. Here $\operatorname{Nm}(I)$ is a positive generator of the fractional ideal in $\mathbb{Q}$ generated by $\operatorname{Nm}(\alpha)$ for $\alpha \in I$.

An important fact is that the quadratic forms $(\operatorname{Hom}(E_1, E_2), \deg)$, for varying supersingular elliptic curves E_i, all belong to the same genus; that is, their localizations at every place are isomorphic. This is easy to check, because locally at a prime $\ell \neq p$ we may assume that $\operatorname{End}(E_1) = M_2(\mathbb{Z}_\ell)$, the ideal I corresponding to $\operatorname{Hom}(E_1, E_2)$ is $M_2(\mathbb{Z}_\ell)$, and the degree is the determinant. At $\ell = p$ we likewise have a description of $\mathcal{O} \otimes \mathbb{Z}_p$ (2.1) and we may again take $I = \mathcal{O} \otimes \mathbb{Z}_p$ with the degree function being induced by the determinant on $M_2(\mathbb{Q}_{p^2})$. Incidentally, this gives another way to find the discriminant and level. Less obvious, but true, is that every integral quadratic form in 4 variables lying in this genus is realized by $(\operatorname{Hom}(E_1, E_2), \deg)$ for some E_1, E_2 [Koh, Corollary 63].

3. SUPERSINGULAR ISOGENY GRAPHS

Supersingular isogeny graphs are directed graphs associated to supersingular elliptic curves and the isogenies between them. Similar constructions can be made for the minimal Ekedahl-Oort strata of the mod p fiber of a Shimura variety, for example for superspecial abelian varieties with real multiplication – an example we discuss below and that we hope to return to in future work. Supersingular isogeny graphs where considered already by Mestre [Mes], if not before, as a tool for computing modular forms on $\Gamma_0(p)$, but have come to broader attention more recently for their cryptographic applications.

3.1. **Definition of supersingular isogeny graphs.** Let p be a fixed prime and $\ell \neq p$ another prime. In many applications one thinks of p as "big", of "cryptographic size" (e.g. $p \approx 2^{500}$), and of ℓ as "small"; very often $\ell = 2$ is a completely adequate choice. The **supersingular isogeny graph** is a directed graph $\mathscr{G}(p, \ell)$ with one vertex for each supersingular j-invariant. To every such j-invariant j we choose once and for all an elliptic curve $E(j)$, defined over $\mathbb{F}_{p^2}$, for which $\operatorname{Fr}^2 = [-p]$. Given vertices j_1, j_2, the edges from j_1 to j_2 correspond to isogenies $f \colon E(j_1) \to E(j_2)$ of degree ℓ, considered up to automorphisms of $E(j_2)$. That is, the edges out of j_1 correspond to subgroups $H \subset E(j_1)[\ell]$ of order ℓ, and such an edge points to j_2 if $E(j_1)/H \cong E(j_2)$. See Figure 1 for an example of two supersingular isogeny graphs with the same prime p and distinct values of ℓ.

It is confusing, but within tolerance, to denote the vertices by $1, \dots, h$ (with $h = h_{B_{p,\infty}}$), or by the supersingular j-invariants $j_1, \dots, j_h \in \mathbb{F}_{p^2}$, or by the corresponding supersingular elliptic curves $E(j_1), \dots, E(j_h)$.

3.1.1. *Variants.* Fix a positive integer n, not divisible by p or ℓ. One can enrich the elliptic curves by a choice of a point of order n and construct a supersingular isogeny graph $\mathscr{G}(p, \ell, \Gamma_1(n))$. The vertices correspond to $\overline{\mathbb{F}}_p$-isomorphism classes of pairs (E, P), where E is supersingular and P a point of exact order n, and edges correspond to isogenies $f \colon (E_1, P_1) \to (E_2, P_2)$, such that f
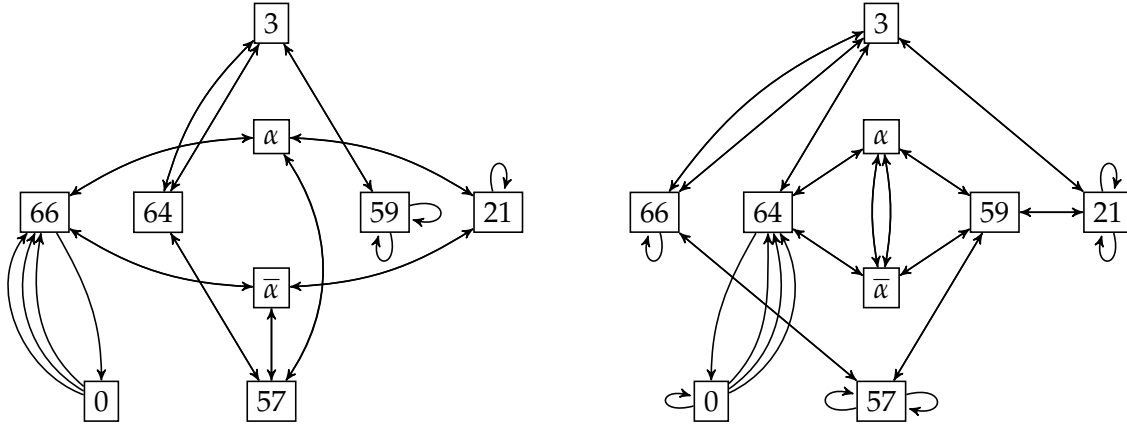
FIGURE 1. $\mathscr{G}(101, 2)$ on the left and $\mathscr{G}(101, 3)$ on the right. Vertices are labelled by j-invariants of supersingular curves over $\mathbb{F}_{101^2}$, where $\alpha, \overline{\alpha} \in \mathbb{F}_{101^2}$ are the roots of $x^2 + 27x + 54$. Each single-headed arrow is a directed edge of the graph, and each double-headed arrow corresponds to two directed edges, one in each direction.

is an isogeny of degree ℓ such that $f(P_1) = P_2$, considered up to automorphisms of (E_2, P_2). The reader familiar with modular curves will have no difficulty imagining other variants such as $\mathscr{G}(p, \ell, \Gamma_0(n))$ and $\mathscr{G}(p, \ell, \Gamma(n))$; collectively these graphs are referred to as supersingular isogeny graphs **with level structure**. These are interesting variants on which there has been some work recently [Arp, GK, PW, Rod]. Note that there is a natural morphism from any such graph to $\mathscr{G}(p, \ell)$.

In another direction, for the supersingular graph $\mathscr{G}(p, \ell)$ one can extract the vertices defined over $\mathbb{F}_p$. There are about $\sqrt{p}$ of these; more precisely, letting h be the class number of $\mathbb{Q}(\sqrt{-p})$, there are $h, 2h$ or $h/2$ if $p \equiv 7 \pmod 8, 3 \pmod 8$ or $1 \pmod 4$, respectively. For two vertices E_i, E_j in this set, we include all the same edges from E_i to E_j as in the original graph $\mathscr{G}(p, \ell)$. This gives the so-called **spine** of the supersingular graph $\mathscr{G}(p, \ell)$. Another variant is to consider supersingular elliptic curves defined over $\mathbb{F}_p$ and $\mathbb{F}_p$-rational isogenies of degree ℓ between them, up to $\mathbb{F}_p$-rational automorphisms of the target. Such studies are motivated by cryptographic considerations. These have been studied in the literature, see [Adj, ACN+, DG] and references therein.

3.2. **Properties of supersingular isogeny graphs.**

3.2.1. *Degree.* Since $\ell \neq p$, for every elliptic curve E over $\overline{\mathbb{F}}_p$ we have $E[\ell] \cong (\mathbb{Z}/\ell\mathbb{Z})^2$. As a result, E has exactly $\ell + 1$ subgroups of order ℓ and so $\mathscr{G}(p, \ell)$ has out-degree $\ell + 1$. Note that by the definition of edges, if there are a edges from E_1 to E_2, then there are in fact $a \cdot \sharp \mathrm{Aut}(E_2)$ isogenies $f \colon E_1 \to E_2$ of degree ℓ. Every such isogeny f has a dual isogeny $f^\vee \colon E_2 \to E_1$. This shows that if there is an edge from E_1 to E_2 then there is an edge from E_2 to E_1, and in fact

(3.1) $\sharp\{\text{edges } E_1 \to E_2\} \cdot \sharp \mathrm{Aut}(E_2) = \sharp\{\text{edges } E_2 \to E_1\} \cdot \sharp \mathrm{Aut}(E_1).$

Thus, if $\mathrm{Aut}(E_1) = \{\pm 1\}$, and every neighbouring vertex E_2 of E_1 also has $\sharp \mathrm{Aut}(E_2) = \{\pm 1\}$, then E_1 has in-degree $\ell + 1$ as well. If $p \equiv 1 \pmod{12}$ then every supersingular elliptic curve E has $\sharp \mathrm{Aut}(E) = \{\pm 1\}$, and so $f \mapsto f^\vee$ establishes a bijection on edges. Identifying each isogeny with its dual allows us to consider $\mathscr{G}(p, \ell)$ as undirected graph of degree $\ell + 1$.

In general there may be vertices E with extra automorphisms. By Deuring's lifting theorem, these would be the elliptic curves with $j = 0$ or 1728, provided they are supersingular; as we have seen, this depends only on the congruence class of p modulo 12. For $p = 2, 3$ there is

only one supersingular j-invariant, which is the j invariant 0. Whenever $j = 0$ or 1728 is a supersingular j-invariant, there is no longer a one-to-one correspondence between inward edges and outward edges at j or at neighbours of j; the in-degree at j, or its neighbours, can be computed using (3.1). See Figure 1 for an example of this behaviour; note that $j = 0$ is the unique supersingular j-invariant in $\mathbb{F}_{101^2}$ with extra automorphisms.

3.2.2. *Brandt matrices.* The reference [Gro] is very useful here. Enumerate the supersingular curves as $E_1, \dots, E_h$, where $h = h_p$ is the class number of $B_{p,\infty}$ and $E_i := E(j_i)$. For every integer $n \geq 0$, define the **Brandt matrix** $B(n)$ to the $h \times h$ matrix such that

$$(3.2) \qquad B(n)_{ij} = \frac{1}{\sharp \operatorname{Aut}(E_j)} \cdot \sharp\{f \colon E_i \to E_j : \deg(f) = n\}.$$

The $n = 0$ case uses our convention that the degree of the zero map is 0. For $n \geq 1$, we can also say that $B(n)_{ij}$ is the number of group schemes H of rank n of E_i such that $E_i/H \cong E_j$. Note in particular that $B(1)$ is the identity matrix, and that $B(p)_{ii}$ is 1 if $j(E_i) \in \mathbb{F}_p$ and is equal to 0 otherwise (this follows from (2.7)). That is, a supersingular j-invariant is in $\mathbb{F}_p$ if and only if $E(j)$ has an endomorphism of degree p. In particular,

$$\operatorname{Tr}(B(p)) = \text{number of supersingular } j\text{-invariants in } \mathbb{F}_p.$$

Note that this trace is exactly the quantity u_B discussed at the end of §2.2.3. More generally, $B(p)_{ij} = 1$ if $j(E_j) = j(E_i)^p$ and is 0 otherwise. For prime $\ell \neq p$, $B(\ell)_{ij}$ equals the number of edges from i to j in $\mathscr{G}(p, \ell)$; that is, $B(\ell)$ is the adjacency matrix of the directed graph $\mathscr{G}(p, \ell)$.

In general, the traces of Brant matrices can be computed explicitly as follows [Gro, Proposition 1.9]: For any integer $n \geq 0$,

$$(3.3) \qquad \operatorname{Tr}(B(n)) = \sum_{s \in \mathbb{Z}, s^2 \leq 4n} H_p(4n - s^2),$$

where the $H_p(4n - s^2)$ are p-modified class Hurwitz class numbers; cf. loc. cit. (1.8). In loc. cit. (1.11), Gross uses this formula to deduce an explicit expression for the number of supersingular j-invariants over $\mathbb{F}_p$ and over $\mathbb{F}_{p^2}$. For further applications to the structure of supersingular isogeny graphs see §3.2.4.

The Brandt matrices are multiplicative in the sense that for $m, n \geq 1$,

$$B(m)B(n) = B(mn), \quad \text{if } \gcd(m, n) = 1,$$

and they satisfy

$$B(\ell^{a+1}) = B(\ell)B(\ell^a) - \ell B(\ell^{a-1}), \quad \ell \neq p \text{ prime}, a \geq 1.$$

Furthermore, $B(p)^2 = Id$.

3.2.3. *The modular polynomial.* Let $n \geq 1$ be an integer. The **modular polynomial**, initially defined over $\mathbb{C}$, is the polynomial in two variables j, j'

$$\Phi_n(j, j'),$$

that vanishes on $(j_1, j_2) \in \mathbb{C}^2$ precisely when there is a cyclic degree n isogeny $E_1 \to E_2$ for elliptic curves $E_1, E_2/\mathbb{C}$ with $j(E_i) = j_i$. Thus, the polynomial $\Phi_n(j_1, j_2)$ defines a (typically singular) curve in the plane $\mathbb{A}^2_{j_1, j_2}$ that is birationally equivalent to the modular curve $X_0(n)$. The definition implies that $\Phi_n(j_1, j_2)$ is a symmetric polynomial, and when considered as a polynomial in j_1 it has degree equal to the number of cyclic subgroups of $(\mathbb{Z}/n\mathbb{Z})$, which is nothing but the Dedekind ψ-function

$$\psi(n) = n \prod_{p \mid n}(1 + \frac{1}{p}).$$

Further, suppose that for $\tau \in \mathbb{C}$ with $\text{Im}(\tau) > 0$ we define $j(\tau)$ to be the j-invariant of the elliptic curve $\mathbb{C}/(\mathbb{Z} \oplus \mathbb{Z}\tau)$. Then for every such τ, Φ_n vanishes at the pair $(j(\tau), j(n\tau))$, and in fact every zero of Φ_n can be written in this way for appropriate τ. See [Lan, 5, §§2-3].

A fact that is less straightforward to show is that normalizing $\Phi_n(j_1, j_2)$ to be monic, we have $\Phi_n(j_1, j_2) \in \mathbb{Z}[j_1, j_2]$. We can therefore reduce Φ_n modulo p, and for every $p \neq n$ this reduction modulo p vanishes on pairs $(j_1, j_2) \in \overline{\mathbb{F}}_p^2$ such that there is a cyclic degree n isogeny from $E(j_1)$ to $E(j_2)$. In particular, if $E(j_1)$ is a supersingular elliptic curve then its neighbours in $\mathscr{G}(p, \ell)$ are the solutions of the degree $\ell + 1$ monic polynomial $\Phi_\ell(j_1, j_2) \in \mathbb{F}_p[j_2]$.

The modular polynomials in $\mathbb{Z}[j_1, j_2]$ have enormous coefficients. For example,

$$\Phi_2(j_1, j_2) =$$
$$j_1^3 + j_2^3 + 1488(j_1^2 j_2 + j_2^2 j_1) - 162000(j_1^2 + j_2^2) - j_1^2 j_2^2 + 8748000000(j_1 + j_2) + 40773375 j_1 j_2 - 157464000000000,$$

and one of the coefficients of Φ_3 is greater than 10^{21}. For more examples, see [Sut2]. Interesting work has been done on the calculation of Φ_n, for instance see [BLS].

3.2.4. *Loops and multiple edges.* Fix a prime ℓ. For $p \neq \ell$, the number of loops (edges from a vertex to itself) and multiple edges (pairs of edges with the same start and end vertices) of $\mathscr{G}(p, \ell)$ is bounded as a function of ℓ alone. To see why this is so, consider first loops. If $f : E \to E$ is a loop then f is an endomorphism of degree ℓ of E and its characteristic polynomial is of the form $x^2 + ax + \ell$ for some $a \in \mathbb{Z}$. This f generates a quadratic imaginary order and so we must have $a^2 < 4\ell$. Thus, there are finitely many such orders. By Deuring's lifting theorem (E, f) can be lifted to $(\tilde{E}, \tilde{f})$ and, considered as complex elliptic curves there are only finitely many possibilities for $\tilde{E}$; in fact, at most the sum of the class numbers of $\mathbb{Z}[x]/(x^2 + ax + \ell)$ and of all the orders of $\mathbb{Q}[x]/(x^2 + ax + \ell)$ containing it. One can make this precise: the number of loops in $\mathscr{G}(p, \ell)$ is none other then the trace $\text{Tr}(B(\ell))$ of the ℓ-th Brandt matrix, as given in (3.3).

The situation for multiple edges is similar. Suppose $f_1, f_2 : E_1 \to E_2$ define multiple edges. Assume for simplicity that $\text{Aut}(E_1) = \{\pm 1\}$, as the analysis is more involved when E_1 has extra automorphisms. Then $f_2^\vee f_1$ is an endomorphism of E_1 of degree ℓ^2. If $\text{Ker}(f_2^\vee f_1) = E_1[\ell]$ then $f_2^\vee f_1 = [\pm \ell]$, and so $f_1 = \pm f_2$, which means that they define the same edge, contrary to hypothesis. Thus, $f_2^\vee f_1$ must be a non-integer endomorphism of E_1 with degree ℓ^2. By the same technique as for loops, the number of curves with a non-integer endomorphism of degree ℓ^2 is bounded as a function of ℓ alone. For a detailed study see [ACN+, Lemma 2.7],[Gha].

3.2.5. *Graph automorphisms.* Given (any) two elliptic curves E_1, E_2 over a perfect field k, base change along the Frobenius map provides a homomorphism of rings

$$\text{Hom}_k(E_1, E_2) \to \text{Hom}_k(E_1^{(p)}, E_2^{(p)}), \quad f \mapsto f^{(p)}.$$

This is entirely obvious when phrased in more generality: for an injection of fields $k \to k_1$ there is a ring homomorphism

$$\text{Hom}_k(E_1, E_2) \to \text{Hom}_{k_1}(E_1 \otimes_k k_1, E_2 \otimes_k k_1), \quad f \mapsto f \otimes Id,$$

only that here the field injection is $\text{Fr} : k \to k$ and we denote the resulting homomorphism $f^{(p)}$. Since k is perfect, we likewise have a base-change relative to Fr^{-1}, and so there is an isomorphism

$$\text{Hom}_k(E_1, E_2) \cong \text{Hom}_k(E_1^{(p)}, E_2^{(p)}).$$

We conclude that there is an automorphism of the graph $\mathscr{G}(p, \ell)$ induced by Frobenius: it takes a vertex $E(j)$ to $E(j)^{(p)} \cong E(j^p)$, and takes an edge corresponding to $f : E_1 \to E_2$ (up to $\text{Aut}(E_2)$) to the edge $f^{(p)} : E_1^{(p)} \to E_2^{(p)}$ (up to $\text{Aut}(E_2^{(p)})$). Further, as all our objects – both the elliptic curves and their isogenies – are defined over $\mathbb{F}_{p^2}$, this automorphism of $\mathscr{G}(p, \ell)$ is an involution;

we will denote it also by Fr. As we shall explain in §5.1 below, under quite general conditions, this is essentially the only non-trivial automorphism of $\mathcal{G}(p, \ell)$.

3.2.6. *Connectedness and expansion.* The graphs $\mathcal{G}(p, \ell)$ are connected. We sketch two proofs of this fact; each relies on a non-trivial theorem. The first method uses a local-global principle for representability of integers by quadratic forms, as in the proof of Theorem 5.2.1; the second uses modular forms, starting a thread that will be continued further in §3.2.8–3.2.9. These two perspectives are deeply related, as will be explored in §6.

A. Let Q be a positive-definite integral quadratic form in $n \geq 4$ variables and discriminant d. Assume that q represents an integer N everywhere locally and that $\gcd(N, d) = 1$. That is, one can solve $q(x_1, \ldots, x_n)$ in q-adic integers for every prime q. Then, if $N \gg 0$, Q represents N over $\mathbb{Z}$. For example, one may use [Han1, Theorem 6.3] and that a quadratic form can be anisotropic only at primes dividing its discriminant [Han1, Remark 3.8.1].

Let E, E' be supersingular elliptic curves. We apply the theorem for $Q = \deg$ on $\mathrm{Hom}(E, E')$ and $N = \ell^k$. The local conditions require us to check that det on $M_2(\mathbb{Z}_q)$ represents N, which is obvious (take the matrix $\mathrm{diag}(1, N)$) and that det on the maximal order $\left\{ \begin{pmatrix} a & pb \\ b^\sigma & a^\sigma \end{pmatrix} : a, b \in \mathcal{O}_{\mathbb{Q}_{p,2}} \right\}$ represents N; this amounts to solving $aa^\sigma - pbb^\sigma = N$ for $a, b \in \mathcal{O}_{\mathbb{Q}_{p,2}}$, which is possible by local class field theory but can also be done "by hand". Thus, for every pair of vertices E, E', there is an isogeny $f \colon E \to E'$ of degree ℓ^k for some k. Let $K_k = \mathrm{Ker}(f)$ and choose subgroups such that

$$K_k \supset K_{k-1} \supset \cdots \supset K_1 \supset K_0 = \{0\}, \quad \sharp K_i = \ell^i.$$

We obtain a sequence of ℓ-isogenies

$$E = E/K_0 \to E/K_1 \to \cdots \to E/K_{k-1} \to E/K_k \cong E',$$

which defines a path of length k from E to E' in $\mathcal{G}(p, \ell)$. Incidentally, note that if $k \gg 0$ then for every supersingular E, E' there is an isogeny $E \to E'$ of degree ℓ^k and also an isogeny of degree ℓ^{k+1}; this proves that the graphs $\mathcal{G}(p, \ell)$ are not bipartite.

B. The adjacency matrix of the oriented graph $\mathcal{G}(p, \ell)$ is conjugate to the matrix of the Hecke operator T_ℓ acting on modular forms of weight 2 on $\Gamma_0(p)$ [Eic1, II §6], [Eic2, §9]. Now, for a finite oriented graph with out-degree $\ell + 1$ the number of connected components of the graph is the multiplicity of the eigenvalue $\ell + 1$ of the adjacency matrix $A = (a_{ij})$, where a_{ij} is the number of edges from the vertex i to the vertex j (and so $\sum_j a_{ij} = \ell + 1$) at least if the graph has the property that when i is connected to j then also j is connected to i, and so there is no ambiguity about the meaning of a connected component. This property holds for our graphs. Thus, one needs to show that the multiplicity of the eigenvalue $\ell + 1$ for T_ℓ acting on modular forms $M_2(\Gamma_0(p))$ is 1. In fact, there is a decomposition of Hecke modules,

$$M_2(\Gamma_0(p)) = \mathbb{C} \cdot \mathscr{E}_2 \oplus S_2(\Gamma_0(p)),$$

where $\mathscr{E}_2$ is an **Eisenstein series** for which $T_\ell \mathscr{E}_2 = (\ell + 1) \mathscr{E}_2$. For any eigenform $f \in S_2(\Gamma_0(p))$, say $T_\ell f = \lambda_f f$, one has the Deligne bound $|\lambda_f| < 2\sqrt{\ell}$. Thus, the eigenspace for the eigenvalue $\ell + 1$ is 1-dimensional, and one concludes that $\mathcal{G}(p, \ell)$ is connected. In fact, this also shows that the graph $\mathcal{G}(p, \ell)$ is Ramanujan in the sense of [LPS]; for this statement it is perhaps best to assume that $p \equiv 1 \pmod{12}$ in which case we can think about $\mathcal{G}(p, \ell)$ as an undirected graph.

The Eisenstein series $\mathscr{E}_2$ is known explicitly[Gro, (5.7)]:

$$(3.4) \qquad \mathscr{E}_2 = \frac{p-1}{24} + \sum_{m \geq 1} \sigma(m)_p q^m,$$

where $\sigma(m)_p = \sum_{d | m, \, p \nmid d} d$. For any $m \geq 1$, $T_m(\mathscr{E}_2) = \sigma(m)_p \mathscr{E}_2$, and in particular $T_\ell \mathscr{E}_2 = (\ell + 1)\mathscr{E}_2$.

3.2.7. *The Deuring correspondence.* Some good references for this section include [Koh] and [Voi, Chapter 42]. As above, we list the supersingular j-invariants $j_1, \ldots, j_h$ with $h = h_{B_{p,\infty}}$, and let $E_i := E(j_i)$ for each i. We will fix a base point j_1; note that by reordering, any supersingular j-invariant can be used as j_1. We let $\mathcal{O} := \text{End}(E_1)$, a maximal order in the quaternion algebra $\mathcal{O} \otimes \mathbb{Q} \cong B_{p,\infty}$.

As described in §2.5, for every $i = 1, \ldots, h$, $\text{Hom}(E_1, E_i)$ is isomorphic as a right $\mathcal{O}$-module to some right ideal of $\mathcal{O}$. Explicitly, if we take any isogeny $\phi \colon E_i \to E_1$, then

$$I_i := \phi \circ \text{Hom}(E_1, E_i)$$

is a subset of $\mathcal{O} = \text{End}(E_1)$, and in fact a right ideal of $\mathcal{O}$; the right ideal class does not depend on the choice of ϕ. For $i = 1$ we can take $\phi = \text{id}$ and therefore $I_1 = \mathcal{O}$. The **Deuring correspondence** says that:

(1) $I_1, \ldots, I_h$ form a complete set of representatives for the right ideal classes of $\mathcal{O}_i$;
(2) The degree map on $\text{Hom}(E_1, E_i)$ corresponds to $\frac{1}{\text{Nm}(I_i)} \text{Nm}(\cdot)$ on I_i;
(3) The left order of I_i is isomorphic to $\text{End}(E_i)$.

From (2) we see that there is an edge from j_1 to j_i in $\mathscr{G}(p, \ell)$ if and only if the quadratic form $\frac{1}{\text{Nm}(I_i)} \text{Nm}(\cdot)$ on I_i represents ℓ. Moreover, $j_1 \in \mathbb{F}_p$ if and only if the norm form on $I_1 = \mathcal{O}$ represents p. For $i \neq 1$, we have $\text{End}(E_i) \cong \text{End}(E_1)$ if and only if the elliptic curves E_i and E_1 are the Frobenius base-change of each other, which is the case if and only if $\frac{1}{\text{Nm}(I_i)} \text{Nm}(\cdot)$ represents p. One concludes that the class number $h_{B_{p,\infty}}$ is the number of vertices of $\mathscr{G}(p, \ell)$, while the type number $t_{B_{p,\infty}}$ (as in §2.2.3) is the number of orbits for the automorphism Fr of the graph.

Define the **supersingular polynomial**

$$S_p(x) = \prod_{\substack{j \in \overline{\mathbb{F}}_p \text{ supersingular}}} (x - j).$$

This is a polynomial of degree $h_{B_{p,\infty}}$ belonging to $\mathbb{F}_p[x]$ that factors as a product of linear and quadratic terms over $\mathbb{F}_p$. The number of linear terms is the number of supersingular j-invariants in $\mathbb{F}_p$ and the number of quadratic terms is the number of pairs $\{j, j^p\}$ where j is a supersingular j-invariant not in $\mathbb{F}_p$. Thus, the number of irreducible factors of S_p is the type number $t_{B_{p,\infty}}$. Here are a few examples:

$$S_{37}(x) = (x - 8)(x^2 - 6x - 6)$$
$$S_{47}(x) = x(x - 1728)(x - 9)(x - 10)(x + 3)$$
$$S_{73}(x) = (x - 9)(x + 17)(x^2 - 5x + 9)(x^2 - 16x + 8)$$
$$S_{101}(x) = x(x - 3)(x - 21)(x - 57)(x - 59)(x - 64)(x - 66)(x^2 + 27x + 54).$$

The primes p for which $S_p(x)$ splits into linear factors – in other words, for which all supersingular j-invariants are defined over $\mathbb{F}_p$ – are $p \leq 31$ and $p = 41, 47, 59, 71$ (these so-called "supersingular primes" are exactly the primes for which the quotient of the modular curve $X_0(p)$ by the Atkin-Lehner involution w_p has genus 0, and are also exactly the prime factors of the order of the monster group). For all other primes, $S_p(x)$ has at least one quadratic factor.

Based on the Deuring correspondence, one can define the supersingular graphs $\mathscr{G}(p, \ell)$ and the Brandt matrices without any reference to elliptic curves. Namely, start with a choice of a maximal order $\mathcal{O}_1$ of $B_{p,\infty}$ and let $I_1, \ldots, I_h$ be representatives for its right ideal classes. Let $\mathcal{O}_i := \mathcal{O}_\ell(I_i)$. Every maximal order in $B_{p,\infty}$ is conjugate to some $\mathcal{O}_i$ (though most conjugacy classes of maximal orders appear twice in the list $\mathcal{O}_1, \ldots, \mathcal{O}_h$). The vertex 1 is connected to the vertex i in $\mathscr{G}(p, \ell)$ by the number of times the form $\frac{1}{\text{Nm}(I_i)} \text{Nm}(\cdot)$ on I_i represents ℓ, divided by $\sharp \mathcal{O}_i^\times$. The same procedure can now be done for the order $\mathcal{O}_i$ to determine the edges from

the vertex i to any vertex j – there is no need to coordinate the choice of representatives for right ideal classes with those already made. Thus, the construction of the Brandt matrices can be done without any reference to supersingular elliptic curves. This allows the construction to be significantly generalized; for instance, Brandt matrices can also be defined for quaternion algebras over totally real fields [Eic2]. See further comments in §6.2.

However, it is important to stress that even if the graph $\mathscr{G}(p,\ell)$ can be defined equally well using elliptic curves or quaternion ideals, these two implementations have drastically different security analyses when considered for cryptographic applications. In particular, path-finding on $\mathscr{G}(p,\ell)$ is believed to be hard when the vertices are labelled by elliptic curves or their j-invariants, but is known to be easy when the vertices are labelled by quaternion ideals. See §4.1.2 for more detail on this.

3.2.8. *Theta functions.* For two supersingular elliptic curves E_1, E_2 we have a rank 4 quadratic module $(\mathrm{Hom}(E_1, E_2), \deg)$ that can be identified with the quadratic module $(I, \frac{1}{\mathrm{Nm}(I)} \mathrm{Nm}(\cdot))$ for a suitable ideal I in $B_{p,\infty}$. Since all maximal orders are everywhere locally conjugate and all maximal orders in a quaternion algebra over $\mathbb{Q}_p$ have class number one, one concludes (as we have already mentioned) that the quadratic modules $(\mathrm{Hom}(E_1, E_2), \deg)$ have discriminant p^2 and level p, and are all in the same genus. Further, any integral quadratic form in this genus is realized as $(\mathrm{Hom}(E_1, E_2), \deg)$ for some E_1, E_2.

We define the **theta function** of this quadratic module by

$$\Theta_{E_1,E_2}(q) := \sum_{n=0}^{\infty} \sharp\{f \in \mathrm{Hom}(E_1, E_2) : \deg(f) = n\} \cdot q^n = \sum_{n=0}^{\infty} r_{E_1,E_2}(n) \cdot q^n.$$

If E_1 and E_2 correspond to ideal classes I_i and I_j respectively under the Deuring correspondence, then $r_{E_1,E_2}(n)$ is nothing other than $2w_j B(n)_{ij}$ (this holds for all $n \geq 0$ because of the way we chose to define $B(0)$). By the general theory of theta functions associated to quadratic forms, the theta function Θ_{E_1,E_2} evaluated at $q = e^{2\pi i z}$ is a weight 2 modular form on $X_0(p)$. A version of Eichler's basis problem (which is a theorem in this case) asserts that the set of weight 2 modular forms

$$\{\Theta_{E_1,E_2} : E_1, E_2 \text{ supersingular elliptic curves}\}$$

spans $M_2(\Gamma_0(p))$. Note that this is a set of roughly $\binom{p/12}{2}$ elements, while the dimension of $M_2(\Gamma_0(p))$ is about $p/12$. Thus, there must be many linear relations between these theta functions. There are some obvious relations that hold for all supersingular E_1, E_2:

- $\Theta_{E_1,E_2} = \Theta_{E_2,E_1}$, because $f \mapsto f^\vee$ is an isometry $\mathrm{Hom}(E_1, E_2) \to \mathrm{Hom}(E_2, E_1)$.
- $\Theta_{E_1,E_2} = \Theta_{E_1^{(p)},E_2^{(p)}}$, because $f \mapsto f^{(p)}$ is an isometry $\mathrm{Hom}(E_1, E_2) \to \mathrm{Hom}(E_1^{(p)}, E_2^{(p)})$.

Aside from these relations and those that can be derived from them, there is no reason to expect any of the linear relations to take the very simple form of an equality between two theta functions. However, such coincidences do occur (in fact, already for $p = 67$); cf. [GLo] and references therein.

Fix a supersingular elliptic curve E_1 and let $E_1, \ldots, E_h$ be representatives for the isomorphism classes of supersingular elliptic curves. Then letting

$$2w_j := \sharp \mathrm{Aut}(E_j),$$

we have

$$\Theta(q) := \sum_{j=1}^{h} \frac{1}{2w_j} \Theta_{E_1,E_j}(q) = \mathscr{E}_2(q),$$

where $\mathscr{E}_2$ is the unique weight 2 Eisenstein series on $\Gamma_0(p)$ with leading coefficient $\frac{p-1}{24}$ appearing in §3.2.6 and for which we have an explicit formula. This result is a special case of the Siegel-Weil formula, as we explain in §6.5.2 below.

One also knows the action of Hecke operators on the theta functions. Simplifying the notation by writing $\Theta_{ij} = \Theta_{E_i, E_j}$, [Gro, Proposition 5.5] or [Eic2] prove that

$$(3.5) \qquad T_m \Theta_{ij} = \sum_k B(m)_{ik} \Theta_{kj} = \sum_k B(m)_{kj} \Theta_{ik}.$$

3.2.9. *Functions on the vertices of $\mathcal{G}(p, \ell)$ and modular forms.* The second proof that the supersingular graphs are connected (§3.2.6) used the fact that $B(\ell)$, the adjacency matrix of $\mathcal{G}(p, \ell)$, also represents the action of the Hecke operator T_ℓ in its action on $M_2(\Gamma_0(p))$ in *some* basis. Thus, there is an isomorphism of the Hecke algebra (with $\mathbb{C}$-coefficients, say) of $X_0(p)$ with the algebra generated over $\mathbb{C}$ by the Brandt matrices $\{B(n) : n \geq 0\}$, such that for n prime to p, $B(n)$ corresponds to the Hecke operator T_n, and $B(p)$ corresponds to the Atkin-Lehner involution. We will denote both algebras by $\mathbb{T}$.

There is a compatible isomorphism between the $\mathbb{T}$-module we call $\mathrm{Fun}(p, \ell)$, which is simply the complex vector space of $\mathbb{C}$-valued functions on the set of supersingular j-invariants under the action of the Brandt matrices, and the module $M_2(\Gamma_0(p))$ with the action of the Hecke operators. We can already find a connection between these two modules in the construction of the theta functions (3.5), but to provide an explicit isomorphism one needs the following construction:

Enumerate the supersingular j-invariants as $j_1, \ldots, j_h$ (in an arbitrary fashion) and let δ_i be the delta function at j_i, namely, taking value 1 on j_i and value 0 at any other j_n. One defines a Hermitian pairing on $\mathrm{Fun}(p, \ell)$, by defining

$$\langle \delta_i, \delta_j \rangle = w_i \cdot \delta_{ij}$$

(here δ_{ij} is Kronecker's delta, equal to 1 if $i = j$ and 0 otherwise). The dual basis $\delta_1^\vee, \ldots, \delta_h^\vee$ is simply $\delta_i^\vee = \frac{1}{w_i} \delta_i$. If $u = \sum \alpha_i \delta_i \in \mathrm{Fun}(p, \ell)$, let $\deg(u) = \sum \alpha_i$; if $v = \sum \beta_i \delta_i^\vee \in \mathrm{Fun}(p, \ell)^\vee$, let $\deg v = \sum \frac{\beta_i}{w_i}$. One lets $\mathbb{T}$ act on $\mathrm{Fun}(p, \ell)$ by the transpose action so that that

$$B(m)\delta_i = \sum_{j=1}^h B(m)_{ij} \delta_j.$$

Then the isomorphism of Hecke modules

$$\mathrm{Fun}(p, \ell) \otimes_{\mathbb{T}} \mathrm{Fun}(p, \ell)^\vee \to M_2(\Gamma_0(p))$$

is given by [Gro, Proposition 5.6]

$$\delta_i \otimes \delta_j^\vee \mapsto \frac{1}{2w_j} + \sum_{m \geq 1} \langle B(m)\delta_i, \delta_j^\vee \rangle q^m = \frac{1}{2w_j} + \sum_{m \geq 1} B(m)_{ij} q^m = \frac{1}{2w_j} \Theta_{E_i, E_j}.$$

More generally, for $u \in \mathrm{Fun}(p, \ell)$, $v \in \mathrm{Fun}(p, \ell)^\vee$,

$$u \otimes v \mapsto \frac{\deg u \cdot \deg v}{2} + \sum_{m \geq 1} \langle B(m)u, v \rangle q^m.$$

4. APPLICATIONS TO CRYPTOGRAPHY

There are several interesting applications of supersingular elliptic curves to cryptography. It was recently shown [CD, MM+] that one of the most prominent applications (SIKE) is not secure, at least under the original implementation (see however the recent preprint [Dar] that suggests that minor adjustments to the implementation will not be able to prevent the attack). Others, though very promising, are still under close scrutiny. That said, such applications instigated further research into the arithmetic of elliptic curves and even abelian varieties, and created much interest in the cryptographic community. This, in our view, is ample justification for further study and for reviewing some of these primitives.

4.1. **Cryptographic hash functions.** Let S be the set of all sequences of $0, 1$ of any finite length. Thus, $S = \{0,1\}^* := \cup_{n=0}^\infty \{0,1\}^n$. Fix a positive integer N. A **hash function** valued in $\{0,1\}^N$ is a function

$$h \colon S \to \{0,1\}^N.$$

One can think about it as associating to any given datum $s \in S$ a certain "trace", or "footprint", $h(s)$ of s. Of course, as S is infinite, such a function cannot be one-to-one and evidently there will be infinitely many $s \in S$ with the same value $h(s)$. More than that, for any $n > N$ there will be two data s_1, s_2 of length n with $h(s_1) = h(s_2)$.

Cryptographic hash functions have to satisfy several restrictions, the most essential of which is that for a given $n \gg N$ and $s_1 \in \{0,1\}^n$ it should be computationally infeasible to find a different $s_2 \in \{0,1\}^n$ such that $h(s_1) = h(s_2)$. Another property is that the number of preimages $h^{-1}(v) \cap \{0,1\}^n$ should converge quickly to $2^n/2^N$ (that is, uniform) as $n \to \infty$.

As a "pseudo-example" of an application of such functions, one can imagine the following scenario: Alice sends to Bob a large file $s \in S$, and Bob wants to verify that he received the file sent by Alice, and not a file that had been purposely manipulated by a third party. One option is for Alice to send the file again through a different communication channel (assuming that it is unlikely that a single third party will be able to compromise both channels), and for Bob to compare the files; unfortunately this uses a lot of extra time and bandwidth if s is very large. A more efficient option is for Alice and Bob to agree on a hash function h in advance (that need not be secret), and then for Alice to send the file s in one channel and its hash value $h(s)$ in another (perhaps this channel is much more secure but cannot handle as much data). Then Bob can apply the hash function h to the received file, and check that it agrees with the hash value sent by Alice. Since it is computationally infeasible to generate a file s' of the same length with the same hash value $h(s) = h(s')$, Alice and Bob can be reasonably confident that the file Bob received is the one sent by Alice.

4.1.1. *CGL hash functions.* The CGL hash function proposed in [CGL1] is based on using data $s \in S$ as instructions for traversing the supersingular graph $\mathscr{G}(p, \ell)$. The j-invariant that is the final vertex of this walk is the hash value; it is an element of $\mathbb{F}_{p^2}$. We explain the procedure in some detail for the case $\ell = 2$, as in the literature this precise description is almost always omitted. To simplify further, choose $p \equiv 1 \pmod{12}$ so that every supersingular elliptic curve has only ± 1 as automorphisms. We have in mind that p is of cryptographic size, say $p \approx 2^{500}$, and the file has size at least $10 \log(p)$, to err on the side of caution.

For each supersingular j-invariant $j \in \mathbb{F}_{p^2}$ we fix a model $E(j)$ over $\mathbb{F}_{p^2}$ satisfying $\mathrm{Fr}^2 = [-p]$ (§2.3.7). Note that for any such curve $E = E(j)$ the 2-torsion points $E[2](\overline{\mathbb{F}}_p)$ are all defined over $\mathbb{F}_{p^2}$: indeed, Fr^2 acts on them by $-p \equiv 1 \pmod 2$.

We fix once and for all an ordering on $\mathbb{F}_{p^2}$: for instance, we may fix an isomorphism $\mathbb{F}_{p^2} \cong \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}$ as groups, take on $\mathbb{Z}/p\mathbb{Z}$ the ordering $\bar{0} < \bar{1} < \ldots \overline{p-1}$, and take on $\mathbb{F}_{p^2}$ the lexicographic order. Fix an initial supersingular j-invariant j_0 and its corresponding model $E_0 := E(j_0)$. Fix also a 2-torsion point $P_0 = (x, 0) \in E_0(\mathbb{F}_{p^2})$.

Given a string $s = s_1 \cdots s_n$ of 0's and 1's, we define a walk on $\mathscr{G}(p, 2)$ as follows. Suppose we have reached j_i for some $i = 0, \ldots, n-1$, and assume we have a 2-torsion point $P_i \in E_i(\mathbb{F}_{p^2})$ for $E_i = E(j_i)$. We label the other 2-torsion points of E_i as $Q_0 = (x_0, 0)$ and $Q_1 = (x_1, 0)$, enumerated so that $x_0 < x_1$ according to the fixed order on $\mathbb{F}_{p^2}$. The bit s_i is either 0 or 1 and hence determines a 2-torsion point $Q_{s_i} \in \{Q_0, Q_1\}$. Using Vélu's formulas [Vél], we can explicitly compute an elliptic curve $E' := E_i/\langle Q_{s_i} \rangle$, together with a map $E_i \to E'$. Let j_{i+1} be the j-invariant of E', so there is an isomorphism $E' \to E_{i+1} := E(j_{i+1})$ that is uniquely determined up to ± 1. Since ± 1 conveniently acts trivially on 2-torsion, the map $E_i[2] \to E_{i+1}[2]$ is uniquely determined. The image of $E_i[2]$ in E_{i+1} (or equivalently, the kernel of the dual isogeny $E_{i+1} \to E_i$)

is generated by some $P_{i+1} \in E_{i+1}[2]$. Thus we obtain a curve E_{i+1} and a 2-torsion point P_{i+1} on it, and we may continue the process.

After n steps, having used $s_1, \ldots, s_n$, we arrive at a pair (E_n, P_n). The hash function value is $j(E_n) \in \mathbb{F}_{p^2}$; there are approximately $\frac{p-1}{12}$ possible values for the hash function. The Ramanujan property of $\mathscr{G}(p, \ell)$ ensures that the pre-image sizes converge to a uniform distribution extremely rapidly, and for large graphs of small degree one can hardly do better in this aspect.

4.1.2. *Security of CGL hash functions.* Not every Ramanujan graph will do for the creation of cryptographic hash functions and indeed some other suggestions made in [CGL1] were quickly discovered to be insecure [PLQ, TZ]. At this point in time, it is understood that the difficulty of breaking the CGL hash function described above rests on the difficulty of the following problem:

Key computational problem:
Given a supersingular elliptic curve E, calculate $\mathrm{End}(E)$*. More precisely, given a Weierstrass equation for a supersingular elliptic curve E, produce a list of elements of* $B_{p,\infty}$ *that generate a maximal order isomorphic to* $\mathrm{End}(E)$*.*

As we have already remarked, there is an efficient algorithm to verify that the given curve E is indeed supersingular [Sut1], and this can be done without computing any non-integer endomorphisms of E. The problem above is key because if one can perform such a calculation efficiently, then one can reduce the problem of finding a path from E_0 to E in $\mathscr{G}(p, \ell)$ – and thereby generating another input hashing to the same value – to a computation solely in the quaternion algebra $B_{p,\infty}$. Indeed, the problem reduces to the following: given two maximal orders, $\mathcal{O}_0 \cong \mathrm{End}(E_0)$ and $\mathcal{O} \cong \mathrm{End}(E)$, find an ideal I such that

$$\mathcal{O}_r(I) = \mathcal{O}_0, \quad \mathcal{O}_\ell(I) \cong \mathcal{O};$$

in addition, for any two-sided ideal J of $\mathcal{O}_\ell(I)$, find an element $\alpha \in JI$ with scaled norm $\frac{1}{\mathrm{Nm}(JI)} \mathrm{Nm}(\alpha)$ equal to a power of ℓ. Indeed, as J varies over two-sided ideals of $\mathcal{O}_\ell(I)$, JI attains all (one or two) right ideal classes of $\mathcal{O}_0$ that correspond to curves E' with $\mathrm{End}(E') \cong \mathcal{O}$ under the Deuring correspondence; in particular, the curve E is attained for some J. The element α then corresponds to an isogeny $E_0 \to E$ of degree a power of ℓ, which can be used to construct a path from E_0 to E in $\mathscr{G}(p, \ell)$. This quaternion algebra variant of the problem can be solved efficiently; see [EH+, KL+, LP, PW, Wes] for more detail.

Of course, one cannot say with absolute confidence whether a cryptographic protocol is secure until is has been broken. At best, one can hope for a proof showing that protocol is as secure as a computational problem that has withstood assaults for a longer time. The fact that the quaternionic algebra problem can be solved efficiently could be viewed as a sign that the CGL hash functions, or other protocols based on supersingular isogeny graphs, are not secure. However, consider the following analogy with the discrete logarithm problem, which is to solve $g^n = h$ for given $g, h \in \mathbb{F}_q^\times$. Choosing a generator γ of $\mathbb{F}_q^\times$, one has an easy isomorphism $f \colon \mathbb{Z}/(q-1)\mathbb{Z} \to \mathbb{F}_q^\times, a \mapsto \gamma^a$, and solving $n f^{-1}(g) = f^{-1}(h)$ in $\mathbb{Z}/(q-1)\mathbb{Z}$ is very easy if $f^{-1}(g)$ and $f^{-1}(h)$ are known. Thus the hardness of the discrete log problem depends on the assumption that the inverse function $f^{-1} \colon \mathbb{F}_q^\times \to \mathbb{Z}/(q-1)$ is hard to compute.

By analogy, choosing a supersingular elliptic curve E with known $\mathcal{O} := \mathrm{End}(E)$ is akin to choosing a primitive root γ, and this is not hard in practice. As already discussed, there is a bijection $f \colon \mathrm{Cl}(\mathcal{O}) \to \{\text{supersingular elliptic curves}\}/\cong$, based on the Deuring correspondence, from right ideal classes of $\mathcal{O}$ to supersingular elliptic curves; this bijection is easy to compute, and allows one to effectively navigate $\mathscr{G}(p, \ell)$. As in the case of the discrete log problem, the difficulty of computing f^{-1} is the basis the security of cryptographic applications of $\mathscr{G}(p, \ell)$.

$$E$$

Alice knows how to $\swarrow$ $\searrow$ Bob knows how to

$$E/H_A \qquad\qquad\qquad\qquad E/H_B$$

Bob knows how to $\searrow$ $\swarrow$ Alice knows how to

$$E/(H_A + H_B)$$

Public: E, P_A, Q_A, P_B, Q_B

Public: $E/H_A, E/H_B, \tilde{P}_A, \tilde{Q}_A, \tilde{P}_B, \tilde{Q}_B$
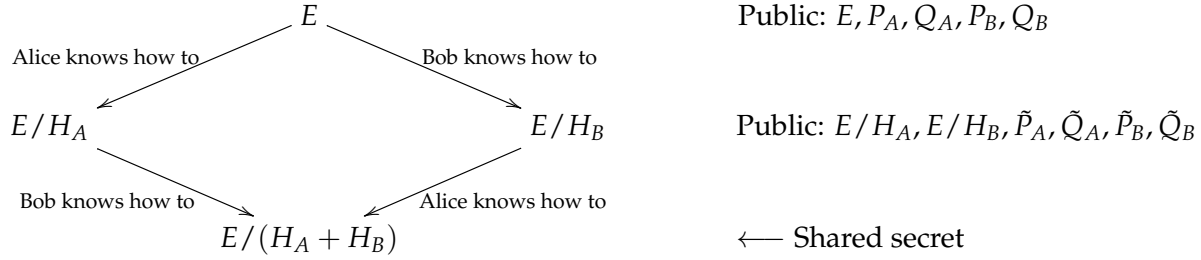
$\longleftarrow$ Shared secret

FIGURE 2. A summary of the SIDH protocol.

In [BCN+] one finds an elegant criterion as to when two closed paths based in E in $\mathscr{G}(p,\ell)$ generate a subring of $\mathrm{End}(E)$ of finite index. First, note that a closed path γ gives, by composing the isogenies corresponding to its edges, an endomorphism f of E. To be precise, f is not uniquely defined by γ (since edges correspond to isogenies only up to automorphisms of the image), but under the hypothesis $p \equiv 1 \pmod{12}$, f is well-defined up to a sign. The degree of f is $\ell^{\mathrm{length}(\gamma)}$. The criterion says (under the simplifying assumption $p \equiv 1 \pmod{12}$) that if two closed paths γ, γ_1 at E, neither with any backtracking, pass through different sets of vertices – meaning one of the paths passes through a vertex through which the other does not – then the two corresponding endomorphisms generate a subring of $\mathrm{End}(E)$ of finite index.

It has been used as a heuristic in analyzing the security of the hash functions (e.g. in [EH+], §3.3]) that one can in fact generate the whole ring $\mathrm{End}(E)$ by using additional closed paths. This follows from Theorem 5.2.2 that we present in §5.2 below.

4.2. **SIDH: A Diffie-Hellman protocol.** De Feo, Jao and Plût suggested a protocol called **SIDH** (short for "Supersingular Isogeny Diffie-Hellman") for establishing a secret key between two parties using supersingular elliptic curves and their isogenies [dFJP]. In a sense, one can view their protocol as "completing squares" in the superposition of two supersingular graphs $\mathscr{G}(p, \ell_A)$ and $\mathscr{G}(p, \ell_B)$. The protocol proceeds roughly as follows: two distinct primes ℓ_A, ℓ_B and powers r_A, r_B are chosen in advance. (Additional conditions on these parameters are posed in [dFJP], but these are largely immaterial for explaining the main idea. Some of these constraints will become clear over the course of the explanation; for instance we will see that r_A, r_B must be fairly large). A base supersingular elliptic curve E is also chosen.

The information $E, \ell_A^{r_A}, \ell_B^{r_B}$ is available for all, as well as a basis P_A, Q_A for $E[\ell_A^{r_A}]$ and P_B, Q_B for $E[\ell_B^{r_B}]$. Alice secretly chooses integers m_A, n_A not divisible by ℓ, and secretly calculates the subgroup $H_A = \langle m_A P_A + n_A Q_A \rangle$, a subgroup of order $\ell_A^{r_A}$ inside $E[\ell_A^{r_A}]$. Independently of Alice and also in secret, Bob likewise chooses secret integers m_B, n_B and computes a subgroup $H_B = \langle m_B P_B + n_B Q_B \rangle$ that is a subgroup of order $\ell_B^{r_B}$ inside $E[\ell_B^{r_B}]$.

Alice publishes the elliptic curve E/H_A and the images $\tilde{P}_B, \tilde{Q}_B$ of the points P_B, Q_B on the curve E/H_A. The data E/H_A is given as a Weierstrass equation, say, and so its publication reveals no information about H_A (if $r_A \gg 0$ there will be many H_A leading to the same elliptic curve). Bob does the same and publishes E/H_B and the image $\tilde{P}_A, \tilde{Q}_A$ of the points P_A, Q_A on E/H_B. In the next step, Alice uses the points $\tilde{P}_A, \tilde{Q}_A$ on the curve E/H_B and her secret m_A, n_A to create a subgroup $\tilde{H}_A = \langle m_A \tilde{P}_A + n_A \tilde{Q}_A \rangle$ on E/H_B; this is the image of H_A on E/H_B under the isogeny $E \to E/H_B$. Bob does the same using $\tilde{P}_B, \tilde{Q}_B$ on E/H_A to get a subgroup $\tilde{H}_B = \langle m_B \tilde{P}_B + n_B \tilde{Q}_B \rangle$ on E/H_A. It is elementary to check that

$$(E/H_A)/\tilde{H}_B \cong E/(H_A + H_B) \cong (E/H_B)/\tilde{H}_A$$

and the secret Alice and Bob share is the *j*-invariant of the elliptic curve $E/(H_A + H_B)$. See Figure 2 for a summary. Note that if, for example, $r_A = r_B = 1$ then we are completing a square

in the superposition of the graphs $\mathscr{G}(p, \ell_A)$ and $\mathscr{G}(p, \ell_B)$ such that parallel edges have the same degree.

A complete implementation of a cryptographic protocol based on this method was submitted under the name **SIKE** (short for "Supersingular Isogeny Key Encapsulation") to the NIST Post-Quantum Cryptography Standardization process in 2017 [Jao]. Unfortunately, this elegant idea is not cryptographically secure, as shown by Castryck - Decru [CD] and, independently, by Maino - Martindale, see [MM+] and references therein: they were able to show that the data of the auxiliary points P_A, Q_A on E together with their images $\tilde{P}_A, \tilde{Q}_A$ on E/H_B is enough to efficiently compute an isogeny $E \to E/H_B$. Interestingly, these attacks use in a very strong way what one may call superspecial graphs associated to principally polarized abelian surfaces and $(2, 2)$-isogenies between them. This suggests that the study of higher dimensional analogues of the supersingular isogeny graphs is both arithmetically rich and has relevance to cryptographic protocols based on elliptic curves. Already there are many cryptographic protocols that take advantage of this extra structure. See for example §4.3.1 and [Dar].

There are other Diffie-Hellman protocols based on supersingular isogeny graphs that do not require the data of auxiliary points and their images under a secret isogeny. This means that so far, they are immune to attacks of the form that cracked SIDH. A few notable examples include **CSIDH** [CL+], which restricts to supersingular elliptic curves defined over $\mathbb{F}_p$, and **OS-IDH** [CK], which enriches each supersingular curve with the data of a marked quadratic order in $\mathrm{End}(E)$. However, the additional structure does lead to some potential vulnerabilities, which have been explored in several works, for instance in [BS, BIJ, BL+, DdF]. More research is needed to determine whether these algorithms are viable for applications.

4.3. **SQIsign: a digital signature protocol.** SQIsign (short for "Short Quaternion and Isogeny Signature") is another cryptographic protocol based on supersingular isogeny graphs. It was introduced by de Feo, Kohel, Leroux, Petit and Wesolowski [dFK+]. To explain the idea we draw on the landmark papers [DH, GMR]. The main challenge is to send messages together with additional information – the signature – that authenticates the sender. To achieve this, the sender uses a one-way function f; that is, an invertible function f that is easy to compute and hard to invert (at least without additional secret information). Given a message M, the signing party publishes the function f and sends the message M together with a signature $f^{-1}(M)$. The receiver, and anyone else, can authenticate the signature by verifying that $f(f^{-1}(M)) = M$. No one else can feasibly calculate $f^{-1}(M)$, and so the signature $f^{-1}(M)$ must be from the party that knows the secret inverse to f, namely the original sender. It is technically possible to "imitate" the original sender by coming up with a signature S first and declaring the message to be $N := f(S)$, but such a message N is likely to be gibberish; in practice, no one else can feasibly come up with both a "reasonable" message N together with a corresponding signature $f^{-1}(N)$, and thus it is not feasible to imitate the sender.

The security of SQIsign relies on the hardness of the following computation problem:

Given a cyclic isogeny $f \colon E \to E'$, find any other cyclic isogeny $g \colon E \to E'$ with $\mathrm{Ker}(f) \neq \mathrm{Ker}(g)$.

(If one specifies that both isogenies have degree a power of ℓ, this is exactly the security assumption for the CGL hash functions.)

The key idea of SQIsign is the following. Suppose one is given an explicit cyclic isogeny from a curve E_A to a curve E_1. By the hardness assumption, if all one knows is this one isogeny, it is believed to be extremely difficult to find any other cyclic isogeny from E_A to E_1. However, if a party knows the endomorphism ring of E_A, then using quaternion methods they can produce a rich collection of isogenies from E_A to E_1. So if Alice wants to sign a message, she will do so by proving that she has access to this rich collection of isogeny paths, and can create distinct cyclic isogenies between the two curves.
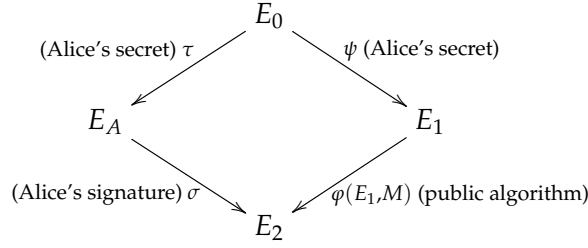
FIGURE 3. A summary of the SQIsign protocol. All four curves are public; only E_0 has public endomorphism ring.

The SQIsign protocol progresses roughly as follows; see [dFK+] for details. First some public information is agreed on: a prime p and a supersingular elliptic curve E_0 over $\mathbb{F}_p$ with known endomorphism ring $\mathcal{O}_0 = \text{End}(E_0)$. If $p \equiv 3 \pmod 4$, one may take the elliptic curve $E : y^2 = x^3 + x$ of j-invariant 1728, whose endomorphism ring can be calculated to be $\mathbb{Z}[i, \frac{i+j}{2}, \frac{1+k}{2}]$, where $i^2 = -1, j^2 = -p$ as in §2.2.1. To see this, on the one hand, we can calculate the discriminant of this order to check that it is maximal. On the other hand, it visibly contains $\mathbb{Z}[i]$. But $\mathbb{Z}[i]$ has class number 1, and so by Deuring's lifting theorem there is a unique maximal order of $B_{p,\infty}$, up to isomorphism, that contains $\mathbb{Z}[i]$. Since $\text{End}(E)$ clearly contains $\mathbb{Z}[i]$, it follows that the specified maximal order is isomorphic to $\text{End}(E)$.

In addition, there is a public algorithm that takes in a starting elliptic curve E and an integer M and produces a cyclic isogeny from E by "following the directions given by M." An example of such an algorithm would be to convert M to a sequence of $0, 1$ bits and apply the CGL hash function starting from E (§4.1.1), but in general this algorithm is allowed to (and, in fact, should) use isogenies of multiple small prime degrees. We will see at the end of the section that for security reasons the directions should not depend on M alone, but for simplicity we ignore this issue for now.

Before Alice signs any messages, she computes a random (secret) isogeny $\tau \colon E_0 \to E_A$ to some E_A, and publishes E_A. The curve E_A will be used to compute the verification function f, but since it is infeasible for anyone else to compute an isogeny from E_0 to E_A, τ will be used to compute the signature function f^{-1}.

Now suppose Alice wants to sign a message M. She first computes a secret random isogeny ψ, originating at E_0, and ending at some E_1, that at the time being is kept secret. Alice then starts from E_1 and follows the directions given by M to compute a cyclic isogeny $\varphi \colon E_1 \to E_2$. Alice now has the complete information of an isogeny $\varphi \circ \psi \circ \tau^\vee \colon E_A \to E_2$, but in fact she knows more: since the endomorphism ring of E_0 is known, she can use τ to compute the endomorphism ring of E_A, as well as a right ideal of $\text{End}(E_A)$ corresponding to E_2 under the Deuring correspondence. Thus she can use quaternion computations to find in this ideal any kind of isogeny $\sigma \colon E_A \to E_2$ she might want; for instance, she can ensure that $\varphi^\vee \circ \sigma \colon E_A \to E_1$ is cyclic. She then publishes the pair (E_1, σ) as the signature on M. Note that σ, and any other isogeny, are published as paths, meaning as a sequence of cyclic isogenies of low degree. This is forced on all such protocols, because there is no other feasible way to write down an isogeny of large degree.

Once Alice's signature has been published, anyone can reproduce E_2 and φ as these depend only on E_1 and M that are public knowledge. To authenticate the signature Bob verifies that $\varphi^\vee \circ \sigma$ is cyclic, which he can do, as Bob, and any other party, knows $\varphi \colon E_1 \to E_2$ and so also its dual $\varphi^\vee \colon E_2 \to E_1$. The protocol is summarized in Figure 3.

Note that once Alice has published her signature (E_1, σ), if an adversary intercepting the communication could find an isogeny from E_0 to *any* of the curves E_1, E_2, E_A, then they could compute the endomorphism rings of *all* the curves involved, revealing Alice's secret and enabling them to forge her signature! This is why it is crucial that Alice starts by producing a curve E_1 that is very far from E_0. It is also crucial that Alice is able to produce multiple isogenies from $E_A \to E_2$ that proceed in very different directions: if Alice just publishes $\varphi \circ \psi \circ \tau^\vee$ for instance, then anyone could follow this path partway to find an isogeny $E_A \to E_0$, again revealing her secret. The key observation is that quaternion methods allow one to choose an alternate path of isogenies $E_A \to E_2$ that completely avoids E_0, as well as any of the curves along the path φ from E_1 to E_2.

If a fraudster wanted to imitate Alice, they could produce a curve E_1 however they wanted, and follow the directions given by a (fake) message N of their choice to reach a curve E_2. But however they try, there will be difficulties in producing a valid signature.

- If they choose E_1 to have known endomorphism ring (for instance via an isogeny from E_0), then it will likely be impossible to find any isogeny connecting E_A to E_2.
- The fraudster could let E_1 be the endpoint of some isogeny μ from E_A; they would then be able to compute an isogeny $\sigma \colon E_A \to E_2$, namely the composition $\varphi \circ \mu$. But in this case $\varphi^\vee \circ \sigma = [\deg \varphi] \circ \mu$ would be very far from cyclic. It is infeasible in this case to find an isogeny σ for which $\varphi^\vee \circ \sigma$ is cyclic, as none of the endomorphism rings of the curves involved in this isogeny are known.
- The fraudster could start with an isogeny $\sigma \colon E_A \to E_2$, then backtrack through the algorithm to find the curve E_1 which, after following the instructions given by N, yields E_2; if they could do this, then they could successfully replicate the signature. To avoid this scenario, we must ensure in the initial setup that the algorithm can't be backsolved. For instance, instead of following directions given by M, the user might first be asked to plug E_1 and M into a hash function h, and follow the directions given by $h(E_1, M)$ instead. By ensuring that both M and E_1 are required before one can even write down the directions, it becomes impossible to start with a desired endpoint E_2 and backtrack.

4.3.1. *NIST submission and variants.* In the past few years there has been a flurry of renewed interest around the SQIsign protocol, owing to a confluence of two factors. On one hand, while the initial NIST Post-Quantum Cryptography Standardization process did not recommend any isogeny-based systems [NIS1], a new call for digital signature schemes was made in 2023 [NIS2] to which SQIsign was submitted. Thus SQIsign has now replaced SIKE as being the only isogeny-based algorithm that is actively being considered for widespread adoption.

On the other hand, it was discovered that the methods used in [CD] to break the SIDH protocol (§4.2) could be used as a new way to generate the isogenies needed for a signature scheme, leading to implementations that are faster, more scaleable, use less memory, and have stronger security analyses. The first of these was **SQISignHD** (HD for "Higher Dimension") [DL+], but similar variants continue to be developed on a regular basis.

4.4. **Insecure subsets of** $\mathscr{G}(p, \ell)$**.** Assume p is very large (say $p \approx 2^{500}$). For all but a vanishingly small proportion of supersingular elliptic curves E over $\mathbb{F}_{p^2}$, the smallest non-integer element of $\mathrm{End}(E)$ will have extremely large norm. To wit, the number of supersingular curves with a non-integer endomorphism of degree at most M grows like $O(M^{\frac{3}{2}})$; for example, of the roughly $\frac{p}{12}$ supersingular elliptic curves, only around $\sqrt{p}$ of them will have a non-integer endomorphism with degree less than $\sqrt[3]{p}$. One implication of this observation is that for the vast majority of supersingular curves over $\mathbb{F}_{p^2}$, one can pick any relatively small ℓ (say $\ell \le 100$) and the smallest non-backtracking cycle in $\mathscr{G}(p, \ell)$ containing E will have length on the order of $\log p$ (as such a cycle produces a non-integer endomorphism of any elliptic curve appearing as a vertex on the cycle). This means that a breadth-first search algorithm starting at this vertex will not be able

to distinguish $\mathscr{G}(p, \ell)$ from an $(\ell + 1)$-regular tree in polynomial time. This near-homogeneity provides a certain amount of security: for a typical E, it is very difficult to use the local structure of any of the graphs $\mathscr{G}(p, \ell)$ in a neighborhood of E to obtain any useful information.

In his thesis [Lov] (see also [LB]), the second author studied the structure of the set of elliptic curves for which this near-homogeneity is not guaranteed. Fix $M < \sqrt{p}/2$, considered as very small compared to p. Let $SS^M(p)$ denote the supersingular j-invariants $j \in \overline{\mathbb{F}}_p$ with the property that if $E(j)$ is a corresponding elliptic curve, then $E(j)$ has an endomorphism $\alpha \notin \mathbb{Z}$ such that $\mathrm{Nm}(\alpha) \leq M$. As said, the cardinality of $SS^M(p)$ has order of magnitude $O(M^{\frac{3}{2}}) \ll p$.

Given $j \in SS^M(p)$, the $\alpha \in \mathrm{End}(E(j)) \setminus \mathbb{Z}$ of small norm is almost unique. Namely, if there is another $\alpha' \in \mathrm{End}(E(j))$ with norm bounded by M, then $\alpha' \in \mathbb{Q}(\alpha)$ (otherwise, one calculates that $\mathbb{Z} \oplus \mathbb{Z}\alpha \oplus \mathbb{Z}\alpha' \oplus \mathbb{Z}\alpha\alpha'$ is a subring of $\mathrm{End}(E)$ with discriminant smaller than p). By abuse of notation we shall refer to any such α as α_j and say it is a **witness** that $j \in SS^M(p)$. Thus, for any $j \in SS^M(p)$ there is a well-defined quadratic field $\mathbb{Q}(\alpha_j)$, where α_j is a witness that $j \in SS^M(p)$. The discriminant D of such a quadratic field satisfies

$$-4M < D < 0.$$

We may thus partition the set $SS^M(p)$ into disjoint subsets called **clusters**:

$$SS^M(p) = \bigcup_D T_D, \qquad j \in T_D \Leftrightarrow \mathbb{Q}(\alpha_j) \text{ has discriminant } D.$$

(To simplify the exposition, we allow some T_D to be empty.) Setting $b := \frac{2}{\sqrt{3}}\sqrt{M}$, we can define a graph $\mathscr{G}'(p, b)$ whose vertices are Galois orbits of supersingular j-invariants and whose edges correspond to Galois orbits of isogenies $f \colon E_1 \to E_2$ of any prime degree $\ell \leq b$, as usual considered up to $\mathrm{Aut}(E_2)$. One may think of $\mathscr{G}'(p, b)$ as obtained by taking the quotient of $\mathscr{G}(p, \ell)$ by the Frobenius involution for all primes $\ell \leq b$, and then gluing together the corresponding vertices of all these $\mathscr{G}(p, \ell)$. Two elliptic curves are close to each other in $\mathscr{G}'(p, b)$ if it is possible to reach one from the other using isogenies of sufficiently small degree as well as the Frobenius isogeny; there are many more isogenies available than in $\mathscr{G}(p, \ell)$, which only uses ℓ-isogenies for a single prime ℓ.

The set $SS^M(p)$ forms a subgraph of $\mathscr{G}'(p, b)$. If M is small enough, $M < \sqrt[3]{3p/16}$, then one has the following [LB, Theorem 13.]: (1) The non-empty clusters are exactly the connected components of the full subgraph on $SS^M(p)$; (2) the shortest path in $\mathscr{G}'(p, b)$ between two vertices in distinct clusters has length greater than $\frac{1}{2}\log_b(p) - \log_b(2M)$. Otherwise said, for any two vertices in a single cluster there exists a path of short isogenies between them that stays within the cluster, but for two vertices in distinct clusters, there is no path between them that stays within $SS^M(p)$. See Figure 4 for an illustration of the graph $\mathscr{G}'(p, b)$ with the clusters in $SS^M(p)$ labeled.

Let ℓ be a prime, $\ell \leq M < \sqrt[3]{3p/16}$. While the clusters of $SS^M(p)$ form connected subgraphs of $\mathscr{G}'(p, b)$, the path between two curves in a common cluster may require isogenies of different prime degrees as well as a Frobenius map. For this reason, the cluster may not form a connected subgraph of $\mathscr{G}(p, \ell)$ in general. Each cluster in $SS^M(p)$ is a juxtaposition of certain very small "patches" from within the graphs $\mathscr{G}(p, \ell)$, $\ell \leq b$.

There are algorithms to compute the endomorphism rings of curves in $SS^M(p)$ with a runtime that is polynomial in M and $\log p$, and so, despite the large distance between clusters, it is always possible to find paths in $\mathscr{G}(p, \ell)$ between any two curves in $SS^M(p)$. In light of the previous sections, this means that $SS^M(p)$ constitutes a portion of $\mathscr{G}(p, \ell)$ that is particularly insecure for cryptographic applications. For instance, *the CGL hash function can be easily broken if the starting*
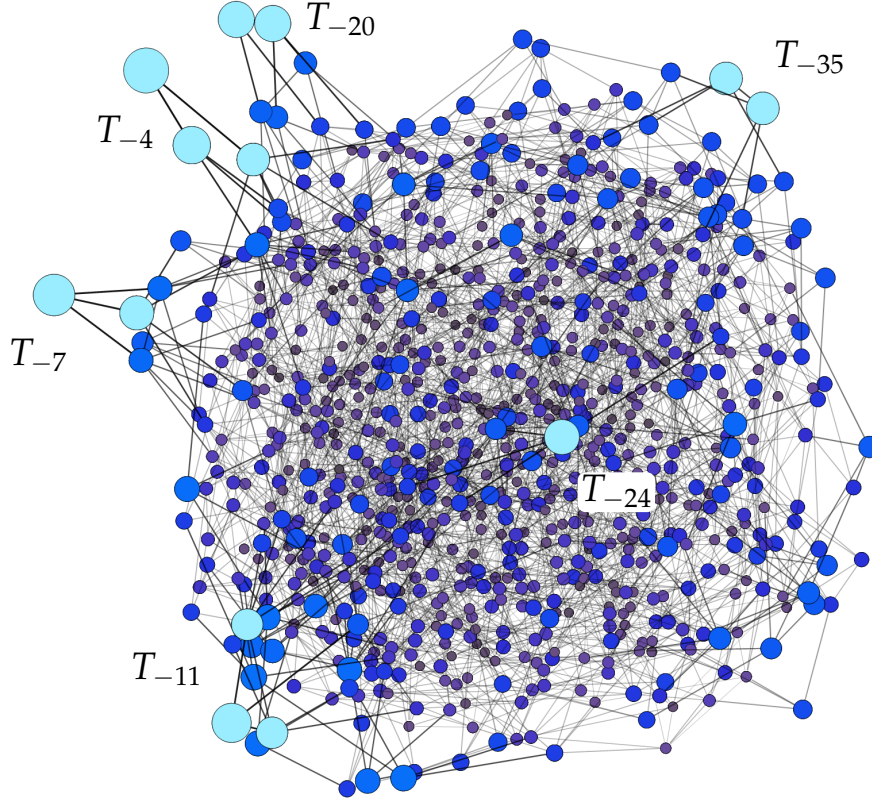
FIGURE 4. The graph $\mathscr{G}'(20011, \frac{2}{\sqrt{3}}(12))$. Vertices are supersingular curves in characteristic $p = 20011$ with conjugate pairs $\{E, E^{(p)}\}$ identified; two curves E, E' are connected by an edge if there is an isogeny $E \to E'$ or $E \to (E')^{(p)}$ of degree 2 or 3. Brighter, larger vertices correspond to curves with smaller non-integer endomorphisms; the light blue vertices are those in $SS^{12}(20011)$, and each cluster is labelled. Data computed using Magma, plotted using Mathematica. A version of this image first appeared in [LB].

and ending curves both happen to lie in $SS^M(p)$. Likewise, the security of the SQIsign protocol depends on E_1, E_2, E_A all avoiding $SS^M(p)$.

At present, the only known method to generate a supersingular curve in large characteristic p in a reasonable amount of time is to start with an elliptic curve with CM and reduce it mod p, as in §2.4; this reduction will necessarily land in $SS^M(p)$ for a value of M proportional to the discriminant of the CM order, and is therefore in the "insecure part" of the isogeny graph, as one must take M to be quite small in order to be able to perform the computation. One may then take such a curve and follow random walks in the isogeny graphs $\mathscr{G}(p, \ell)$ for various values of ℓ to approximate choosing a supersingular curve uniformly at random; however, the party that computes this path can use this information to compute the endomorphism ring of the output curve. This means that there is currently no known way to produce a supersingular curve in a feasible amount of time without at least one party knowing the endomorphism ring of this curve (see also [BB+]). This is a serious issue for some applications, such as [dFM+], which due to this limitation can only be implemented using "trusted setup". Other applications, including Diffie-Hellman protocols, avoid this issue by having multiple users each produce curves independently, ensuring that no individual user has complete information.

5. SOME NEW RESULTS CONCERNING $\mathcal{G}(p, \ell)$

We report here on three new results concerning the subjects of this article. Since the details have or will appear in print elsewhere, we only provide explanation, context and comments regarding the proofs, referring the interested reader to the articles themselves.

5.1. **Rigidity.** From the point of view of cryptographic applications, any extra structure the graphs $\mathcal{G}(p, \ell)$ may have could be viewed as potential vulnerability of their cryptographic applications. Thus, one would be happy to know that they have very few automorphisms.

Let $\mathrm{Aut}(\mathcal{G}(p, \ell))$ denote the **automorphism group** of the graph. Such an automorphism is a pair (f_V, f_E), where f_V is a bijection on the set on the set of vertices and f_E is a bijection on the set of oriented edges such that if v is the origin (respectively, end) of an edge e then $f_V(v)$ is the origin (respectively, end) of the edge $f_E(e)$. Among these automorphisms there are the trivial ones $A^{\mathrm{triv}}(\mathcal{G}(p, \ell))$, which are the automorphisms $(f_V, f_E) = (\mathrm{Id}, f_E)$. This is a very small and easy to describe normal subgroup of $\mathrm{Aut}(\mathcal{G}(p, \ell))$ that simply permutes the loops at a vertex, or the multiple edges between two vertices. Since the number of such loops and multiple edges is bounded in terms of ℓ alone and is completely understood in terms of embeddings of quadratic orders into maximal orders of $B_{p,\infty}$, the structure of $A^{\mathrm{triv}}(\mathcal{G}(p, \ell))$ is also grosso modo well-understood. We let the **essential automorphism group** be

$$\mathrm{Aut}^{\mathrm{ess}}(\mathcal{G}(p, \ell)) = \mathrm{Aut}(\mathcal{G}(p, \ell)) / A^{\mathrm{triv}}(\mathcal{G}(p, \ell)).$$

In the following theorem, one assumes $p > 71$ simply to guarantee that the automorphism Fr of $\mathcal{G}(p, \ell)$ is not an element of $A^{\mathrm{triv}}(\mathcal{G}(p, \ell))$. Furthermore, the assumption made in the theorem is now known to hold for density 1 of primes ℓ for a given p [KSW]. There are cases where this assumption is not satisfied and also $\mathrm{Aut}^{\mathrm{ess}}(\mathcal{G}(p, \ell))$ is bigger. Sam Mayo proved the following theorem during his undergraduate summer program with the first author of this article [May].

Theorem 5.1.1. *Let $p > 71$ and assume that T_ℓ generates the Hecke algebra $\mathbb{T}_\mathbb{C}$ of $X_0(p)$ over the complex numbers. Then*

$$\mathrm{Aut}^{\mathrm{ess}}(\mathcal{G}(p, \ell)) = \{Id, \mathrm{Fr}\}.$$

For the proof and references, see loc. cit. The first move is to note that an automorphism φ of the graph commutes with the adjacency matrix of $\mathcal{G}(p, \ell)$, thus is a linear operator on the functions $\mathrm{Fun}(p, \ell)$ that commutes with T_ℓ, hence with $\mathbb{T}$. However, by a result of Emerton, that implies that $\varphi \in \mathbb{T}$ (in fact, over $\mathbb{Z}$). As $\mathbb{T}_\mathbb{Q}$ is product of totally real fields, that already gives that φ belongs to an elementary 2 group of size at most 2^g, where g is the genus of $X_0(p)$. Further study puts further conditions on φ and in particular shows that it arises from an automorphism of $X_0(p)$ itself. Those have been well-studied, and one concludes that the only non-trivial option is for φ to be induced by the Atkin-Lehner involution.

5.2. **Generation by ℓ-isogenies.** In analyzing the security of hash functions, or other cryptographic protocols, associated with $\mathcal{G}(p, \ell)$, sometimes one requires the assumption that for a supersingular curve E, the ring $\mathrm{End}(E)$ is generated as a $\mathbb{Z}$-module (equivalently, as a ring) by homomorphisms f of degree $\{\ell^k : k \geq 0\}$; for example, in [EH+, §3.3]. It turns out that this is but a very special case of a general result concerning the generation of lattices of rank at least 4 by elements of norms in a specified set. In §6, we generalize this result to lattices over totally real fields.

Let $Q : \mathbb{Z}^n \to \mathbb{Q}$ be a quadratic form; that is, the function $(x, y) := \frac{1}{2}(Q(x + y) - Q(x) - Q(y))$ is a bilinear form. We say that Q is **integral** if $Q(x) \in \mathbb{Z}$ for all $x \in \mathbb{Z}^n$ (which implies that $(x, y) \in \frac{1}{2}\mathbb{Z}$ for $x, y \in \mathbb{Z}^n$), that Q is **nondegenerate** if for all $x \in \mathbb{Z}^n$ there exists $y \in \mathbb{Z}^n$ with $(x, y) \neq 0$, and that Q is **primitive** if the ideal of $\mathbb{Z}$ generated by the values $Q(x), x \in \mathbb{Z}^n$ is $\mathbb{Z}$ itself. In [GLo, Corollary 2.3], we proved the following theorem.

Theorem 5.2.1. *Let Q be an integral, non-degenerate, primitive quadratic form in $n \geq 4$ variables. Let S be an infinite set of positive integers. Assume that for all $s \in S$ and all primes r, there exists a basis of $\mathbb{Z}_r^n$ comprised of elements x with norm $Q(x) = s$. Then $\mathbb{Z}^n$ is generated by elements x with norm $Q(x) \in S$.*

One can apply this result to other lattices $\Lambda \neq \mathbb{Z}^n$ by picking a basis, defining an isomorphism $\Lambda \simeq \mathbb{Z}^n$. The proof given in [GLo] uses a strong approximation theorem of N. Sardari that allows finding integral points of sufficiently large norm satisfying local congruence conditions. The conditions required to apply Sardari's theorem hold by the assumptions in our theorem, but in the application to maximal orders of $B_{p,\infty}$ they reduce to a very easy verification about existence of bases of maximal orders $B_{p,\infty} \otimes \mathbb{Q}_r$ (for *all* primes r, including $r = p$) with given norms $s \in S$, that are easy to check given we have an explicit and simple description of the local maximal orders. We thus deduce:

Theorem 5.2.2. *Let $\ell \neq p$ be a prime. Every maximal order $\mathcal{O}$ of $B_{p,\infty}$ is generated by elements of norm in the set $\{\ell^k : k = 0, 1, 2, 3, \dots\}$.*

In §6, we would prefer to formulate our approach slightly differently. In truth, in both approaches the deus ex machina is the same, but a change of perspective will be enlightening and useful.

5.2.1. *An application to even unimodular lattices.* While somewhat tangential to our main goal of studying supersingular isogenies, the authors would like to use this opportunity to share another intriguing and charming corollary of Theorem 5.2.1.

Corollary 5.2.3. *Let S be any infinite set of positive even integers, and Λ an even unimodular lattice. Then Λ is generated by elements with norm in S.*

Proof. All even unimodular lattices are in the same genus as E_8^n for some n; see [Ome], especially §106 A. So by Theorem 5.2.1, it suffices to check that for any $s \in S$ and every prime p, there exists a basis of the localization of E_8 at p consisting of elements of norm s.

Locally at 2, the quadratic form associated to E_8 is equivalent to $2x_1x_2 + 2x_3x_4 + 2x_5x_6 + 2x_7x_8$ (loc. cit.). Writing $s = 2t$, the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & t \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & t \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & t \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & t \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & t \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & t \\ 1 & t & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

has determinant 1, and so the row vectors form a basis of vectors of norm s for this lattice.

Locally at $p \geq 3$, the quadratic form is equivalent to $x_1^2 + \cdots + x_8^2$ (loc. cit.). There exist p-adic integers x, y, z solving $x^2 + y^2 + z^2 = s - 1$ and such that z is a p-adic unit: explicitly, we can take $x, y \in \mathbb{Z}$ satisfying $x^2 + y^2 \equiv s - 2 \pmod{p}$ (possible by the pigeonhole principle because the sets $\{s - 2 - y^2 : y \in \mathbb{F}_p\}$ and $\{x^2 : x \in \mathbb{F}_p\}$ both have $\frac{p+1}{2}$ elements), so that $z_0 = 1$ solves $x^2 + y^2 + z_0^2 \equiv s - 1 \pmod{p}$, and we obtain $z \in \mathbb{Z}_p^\times$ with $x^2 + y^2 + z^2 = s - 1$ by Hensel lifting. The matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & x & y & z \\ -1 & 0 & 0 & 0 & 0 & x & y & z \\ 0 & 1 & 0 & 0 & 0 & x & y & z \\ 0 & 0 & 1 & 0 & 0 & x & y & z \\ 0 & 0 & 0 & 1 & 0 & x & y & z \\ 0 & 0 & 0 & 0 & 1 & x & y & z \\ x & y & z & 0 & 0 & 1 & 0 & 0 \\ x & y & z & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

has determinant $2z \in \mathbb{Z}_p^\times$ and so the row vectors form a basis of vectors of norm s for this lattice. $\square$

5.3. **Identifying an elliptic curve from its endomorphism ring.** Let E be a supersingular elliptic curve. One can ask whether E is determined up to isomorphism by $\mathrm{End}(E)$. The answer to that goes back more than a century and was already mentioned above: two supersingular elliptic curves with j-invariants j and j' have isomorphic endomorphism rings if and only if $j' \in \{j, j^p\}$. Thus, one may ask further if there is a way to generate the elliptic curve given its maximal order. If one knows a single base supersingular elliptic curve E_0 as a Weierstrass equation and also a description of its maximal order, then the problem can be solved by finding a path in the isogeny graph $\mathscr{G}(p, \ell)$ by working first completely at the level of quaternion algebras. This path is then transformed to a path in $\mathscr{G}(p, \ell)$, labelled with elliptic curves now, that starts with E_0 and ends in E. Vélu's formula can be used to derive a Weierstrass equation for E. There is thus a solution to this problem that is polynomial time in $\log(p)$; see [EH+, KL+].

For many primes we can find such a base elliptic curve E_0 quite easily. We have already given an example in §4.3 when $p \equiv 3 \pmod 4$.

We can consider another invariant for a supersingular elliptic curve E which is its theta function as in §3.2.8

$$\Theta_E(q) = \sum_{n=0}^{\infty} \sharp\{f \in \mathrm{End}_{\overline{\mathbb{F}}_p}(E) : \deg(f) = n\} \cdot q^n = \sum_{n=0}^{\infty} r_E(n) \cdot q^n.$$

It is natural to ask whether Θ_E determines E up to Frobenius transform. While the theta function is not a fine enough invariant to be able to distinguish between rank 4 lattices in general [CSl], the authors showed that it is strong enough to distinguish between endomorphism rings of elliptic curves [GLo].

Theorem 5.3.1. *Let E, E' be supersingular elliptic curves with j invariants j, j', respectively. Then*

$$\Theta_E = \Theta_{E'} \quad \Leftrightarrow \quad j' \in \{j, j^p\}.$$

This raises the question whether one can construct the order $\mathrm{End}(E)$ from Θ_E. The answer to that is yes, and we discuss that after discussing the proof Theorem 5.3.1.

5.3.1. *Discussion of the proof of Theorem 5.3.1.* The proof of Theorem 5.3.1 is hard to conceptualize, and that is one of the difficulties in extending it. Here are some of the main ideas used in the proof.

(1) Throughout the proof there is an interplay between a maximal order $\mathcal{O}$ of $B_{p,\infty}$ and its **Gross lattice** $\mathcal{O}^T$ used much in [Gro]. This is a ternary lattice which is the image of $\mathcal{O}$ under the map $x \mapsto 2x - \mathrm{Tr}(x)$. Its importance is that it classifies embeddings of quadratic imaginary orders into $\mathcal{O}$; this construction can easily be generalized to quaternion algebras B over number fields L and the classification of embeddings of monogenic orders containing $\mathcal{O}_L$ in quadratic extensions of L into maximal orders of B.

(2) *Peeling off the first layer:* the power series $\Theta_E(q) - \Theta_{\mathbb{Z}}(q)$ counts the number of non-integer elements in $\mathrm{End}(E)$ of each norm; thus if the first coefficient is cq^n, we know there is at least one non-integer element α of norm n. Since there are no elements of smaller norm, we know the trace of α must be 0 or ± 1 (otherwise we could subtract an integer to find an element of smaller norm); i.e. we have either $\alpha = \sqrt{-n}$ or $\alpha = \frac{1}{2}(\pm 1 + \sqrt{1 - 4n})$ in the corresponding quadratic field. Minkowski's bounds tell us that n is small in comparison to p.

(3) A key lemma, the original proof of which we can find in the literature is due to Kaneko [Kan], but it has been used by various other authors independently, e.g. [GL1], gives constraints not only on the norms of elements of a quaternion order generating distinct imaginary quadratic orders, but as proven in [GLo] also on the *angles* between such elements. This constraint is especially strong when both elements are relatively small compared to the discriminant of the quaternion order.

(4) Combining Minkowski's bounds with Kaneko's lemma, we find that the elements of norm n cannot lie in distinct-but-isomorphic quadratic orders. Thus we can learn the full minimal polynomials of the elements of norm n from the value of c alone (since $\mathbb{Z}[\sqrt{-n}]$ has two elements of norm n and $\mathbb{Z}[\frac{1}{2}(1+\sqrt{1-4n})]$ has four), giving us the full lattice structure of a rank 2 sub lattice $\mathbb{Z}[\alpha]$ of $\mathcal{O}$.

(5) *Peeling off the second layer:* Repeat the above steps but for $\Theta_E(q) - \Theta_{\mathbb{Z}[\alpha]}(q)$ to find the next independent short element β. Kaneko's lemma can again be used to identify the minimal polynomial of β as well as its angle with α, giving us now the full lattice structure of a rank 3 sub lattice $\mathbb{Z} + \alpha\mathbb{Z} + \beta\mathbb{Z}$ of $\mathcal{O}$.

(6) *Peeling off the third layer:* Unfortunately, the Minkowski bounds on the fourth successive minimum are not strong enough for Kaneko's lemma to apply directly anymore; that is, it is possible for there to be multiple shortest elements of $\mathrm{End}(E) - (\mathbb{Z} + \alpha\mathbb{Z} + \beta\mathbb{Z})$ that satisfy the same minimal polynomial. At this point a more careful analysis of the lattice geometry is needed: we prove that for each possible minimal polynomial of the last successive minimum, only certain initial sequences of coefficients of $\Theta_E(q) - \Theta_{\mathbb{Z}+\alpha\mathbb{Z}+\beta\mathbb{Z}}(q)$ are possible, and these possibilities do not overlap. Thus the theta function allows us to determine the minimal polynomial of a fourth independent vector γ.

(7) Once again, Kaneko's lemma allows us to determine the angles between α, β, γ, and from this also the angles between their images in the Gross lattice. This determines the isometry type of the Gross lattice, which uniquely determines the quaternion order up to isomorphism by a result of Chevyrev and Galbraith [CG].

5.3.2. *Determining the order from its theta function.* Let Θ be a weight 2 modular form on $\Gamma_0(p)$ that we know to be equal to Θ_E for some supersingular elliptic curve E. Let also $\mathcal{O} = \mathrm{End}(E)$. We can also say that Θ is associated to the maximal order $\mathcal{O}$ of $B_{p,\infty}$. The question we address here is how to determine $\mathcal{O}$.

The proof of Theorem 5.3.1 proceeds, as we have discussed, by determining from the theta function Θ the successive minima of the Gross lattice $\mathcal{O}^T$, and the angles between the vectors realizing them. Thus, one is able to reconstruct the Gross lattice $\mathcal{O}^T$ from Θ. One can then find the maximal order $\mathcal{O}$ by the taking the even integral Clifford algebra of $(\mathcal{O}^T)^\vee$, the dual lattice of $\mathcal{O}^T$. See [Voi, §§22.1 - 22.3].

5.3.3. *Determining the elliptic curve from the statistics of closed paths in $\mathscr{G}(p,\ell)$.* Consider a supersingular elliptic curve E as a vertex of the supersingular graph $\mathscr{G}(p,\ell)$ and consider the statistics of closed paths in $\mathscr{G}(p,\ell)$, based at E, of all lengths k. Those are precisely the diagonal entries $B(\ell^k)_{11}$ (where in the enumeration of the supersingular elliptic curves E is declared the first one). Those are also the coefficients $\{r_E(\ell^k) : k = 0,1,2,3\dots\}$ in Θ_E. We claim that if T_ℓ generates the Hecke algebra $\mathbb{T}$ (which, as noted above, occurs for a density 1 set of primes ℓ) then this information alone is sufficient to determine E up to Frobenius base change.

Corollary 5.3.2. *Assume that T_ℓ generates the Hecke algebra $\mathbb{T}$ of $X_0(p)$ and E, E' are supersingular elliptic curves such that in their theta functions the coefficients of the powers $q^{\ell^k}, k = 0,1,2,3\dots$ agree. Then, $\Theta_E = \Theta_{E'}$ and so $E' \cong E$ or $E' \cong E^{(p)}$ over $\overline{\mathbb{F}}_p$.*

In fact, a similar statement holds even if $\mathbb{T}$ is not generated by a single Hecke operator. In general, $\mathbb{T}$ is generated by some finite set of Hecke operators, say $T_{\ell_1}, \dots, T_{\ell_s}$. Using an argument like the one below, one can show that if the coefficients of the powers $q^{\ell_i^k}$ agree for all $i = 1, \dots, s$ and $k \geq 0$, then $\Theta_E = \Theta_{E'}$. But for simplicity we only consider the case $s = 1$ below.

By considering the form $F = \Theta_E - \Theta_{E'}$, which is a cusp form since the Eisenstein space of $M_2(\Gamma_0(p))$ is one-dimensional, it suffices to prove the following under the same assumption that T_ℓ generates $\mathbb{T}$: given a cusp form $F \in S_2(\Gamma_0(p))$, if the coefficient in F of q^{ℓ^k} vanishes for all $k = 0,1,2,3,\dots$, then $F = 0$.

Let $f_1, \ldots, f_n$ be a basis of normalized eigenforms in $S_2(\Gamma_0(p))$ (here n is one less than the class number of $B_{p,\infty}$), and write

$$F = \sum_{r=1}^{n} c_r f_r, \qquad T_\ell(f_r) = \lambda_r f_r.$$

If $\lambda_a = \lambda_b$ then, since T_ℓ generates the Hecke algebra, we have $T f_a = T f_b$ for every Hecke operator $T \in \mathbb{T}$. But by multiplicity one for $M_2(\Gamma_0(p))$, any two distinct eigenforms must have distinct eigenvalues for some Hecke operator. Thus we can conclude that the eigenvalues $\lambda_1, \ldots, \lambda_n$ of T_ℓ are distinct.

Now write a Fourier expansion for each cusp form, $f_r = \sum a_r(i) q^i$. By assumption on the coefficients of q^{ℓ^k} in F, we have

$$\sum_{r=1}^{n} c_r a_r(\ell^k) = 0$$

for all k. On the other hand, we can compute the coefficients $a_r(\ell^k)$ explicitly using the relations

$$a_r(\ell) = \lambda_r,$$
$$a_r(\ell^k) = a_r(\ell) a_r(\ell^{k-1}) - \ell^{k-1} a_r(\ell^{k-2})$$

that hold for any normalized eigenform. This allows us to write each $a_r(\ell^k)$ as a monic integer degree k polynomial in λ_r, for instance

$$a_r(\ell^2) = \lambda_r^2 - \ell,$$
$$a_r(\ell^3) = \lambda_r^3 - (\ell + \ell^2) \lambda_r,$$

and so on. Thus, using the fact that $\sum c_r a_r(\ell^k) = 0$, by induction on k we find that

$$\sum_{r=1}^{n} c_r \lambda_r^k = 0$$

for all $k \geq 0$. Since $\lambda_1, \ldots, \lambda_n$ are distinct, this implies that $c_1 = c_2 = \cdots = c_n = 0$ by a Vandermonde determinant calculation: that is, we have

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & \cdots & \lambda_n^2 \\ & & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \lambda_3^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{pmatrix} = 0,$$

where the square matrix has determinant $\prod_{1 \leq i < j \leq n} (\lambda_j - \lambda_i) \neq 0$. In fact, this shows that we only need to know the q^{ℓ^k} coefficient vanishes for $k = 0, \ldots, n-1$ in order to conclude $F = 0$.

6. GENERATION OF TOTALLY REAL LATTICES BY SPECIAL ELEMENTS

Our purpose in this section is to generalize Theorem 5.2.1 to lattices over totally real fields. Besides its intrinsic appeal, this result also has implications for a generalization of supersingular isogeny graphs $\mathscr{G}(p, \ell)$ that we define below. This generalization uses Hilbert modular varieties and should have interesting applications to cryptography.

Much more background is assumed of the reader in this section, as it would be unrealistic to introduce from scratch the rather high-powered theories from arithmetic geometry and automorphic forms that we need.

A natural, but different, extension of the theory is to superspecial principally polarized abelian varieties – see [ATY, CDS, CSm, FT, Jor, JZ] and references therein. In fact, the superspecial isogeny graphs coming from Hilbert modular varieties have implications to the study of those coming from Siegel modular varieties (see §6.2 below).

6.1. **A brutally short introduction to abelian varieties with real multiplication.** The books [Gor2, vdG] and the foundational papers [DP, Rap] provide general background on abelian varieties with real multiplication and their moduli spaces.

To simplify the exposition, and to rely on results already in the literature, we put ourselves in the following setting. Let L denote a totally real field of degree g over $\mathbb{Q}$, and let $\mathcal{O}_L$ be the ring of integers of L. We assume that L has strict class number 1; that is, all ideals are principal, and there exist units in $\mathcal{O}_L$ of all possible sign combinations. We denote by $\mathfrak{d}_L$ the different ideal, δ_L a totally positive generator of $\mathfrak{d}_L$, and $d_L := N_{L/\mathbb{Q}}(\delta_L)$ the discriminant of $L/\mathbb{Q}$. We also choose a rational prime ℓ that is inert in L, so that $\mathcal{O}_L/\ell\mathcal{O}_L \cong \mathbb{F}_{\ell^g}$. From the point of view of algebraic number theory the assumptions on the class number and decomposition of ℓ are rather severe restrictions, but much of what we discuss below can be generalized to arbitrary class numbers and arbitrary rational primes ℓ that are, say, unramified in L. We also fix a rational prime p, $p \neq \ell$, that is unramified in L. Using this data, we will construct analogues $\mathscr{G}_L(p, \ell)$ of the supersingular graphs $\mathscr{G}(p, \ell)$.

Let A be a g-dimensional abelian variety over a field. Suppose that A comes equipped with an action of $\mathcal{O}_L$ – that is, with an injective ring homomorphism $\iota: \mathcal{O}_L \to \mathrm{End}(A)$ – and a principal $\mathcal{O}_L$-linear polarization $\lambda : A \to A^\vee$. Here $\mathcal{O}_L$-linear means that if we equip the dual abelian variety $A^\vee$ with the dual $\mathcal{O}_L$-action $\iota^\vee: \mathcal{O}_L \to \mathrm{End}(A^\vee)$, where $\iota^\vee(\alpha) := (\iota(\alpha))^\vee$, then we require $\lambda \circ \iota = \iota^\vee \circ \lambda$. With this setup, the triple $\underline{A} = (A, \iota, \lambda)$ is called an abelian variety with real multiplication by $\mathcal{O}_L$, which we abbreviate as an **AV with RM** (the field L being understood from the context).

There is a coarse moduli scheme $\mathcal{M}_L$ for such objects. It is a quasi-projective scheme, flat of relative dimension g over $\mathrm{Spec}(\mathbb{Z}[d_L^{-1}])$. The scheme has quotient singularities that can be resolved by introducing any rigid level-N structure, for $N \geq 3$ and $(N, d_L) = 1$, at which point the moduli scheme becomes smooth over $\mathrm{Spec}(\mathbb{Z}[\zeta_N][(Nd_L)^{-1}])$. There is a natural isomorphism

$$\mathcal{M}_L(\mathbb{C}) \cong \mathrm{SL}_2(\mathcal{O}_L)\backslash\mathfrak{H}^g,$$

where $\mathfrak{H} = \{z : \mathrm{Im}(z) > 0\}$ is the complex upper half space and $\mathrm{SL}_2(\mathcal{O}_L)$ acts diagonally as Möbius transformations, using the g different embeddings $\sigma_i: L \to \mathbb{R}$.

If A is complex abelian variety with RM and $A \cong \mathbb{C}^g/\Lambda$, then Λ is a torsion-free $\mathcal{O}_L$-module of rank 2, and is hence projective. So for some $\mathcal{O}_L$-ideals $\mathfrak{a}, \mathfrak{b}$ we have $\Lambda \cong \mathfrak{a} \oplus \mathfrak{b}$ as an $\mathcal{O}_L$-module, and since we assume L has strict class number 1 this implies $\Lambda \cong \mathcal{O}_L^2$. However, the realization of Λ as a lattice in $\mathbb{C}^g$ depends on A (much as is the case for elliptic curves). This is the first step in parametrizing complex abelian varieties with RM by $(\tau_1, \ldots, \tau_g) \in \mathfrak{H}^g$ (loc. cit. and [BiL]).

The mod p fiber of $\mathcal{M}_L$, denoted $\overline{\mathcal{M}}_L$ (p being understood from the context), classifies such triples $\underline{A}$ over fields of characteristic p. In the Goren-Oort stratification [GO] of $\overline{\mathcal{M}}_L$, which is stable under prime-to-p Hecke operators, the smallest stratum consists of finitely many points called **superspecial points** that correspond to principally polarized superspecial abelian varieties with real multiplication (**ssAV with RM**).

There are several equivalent definitions of superspecial abelian varieties. For instance, over any algebraically closed field k of characteristic p, one can define a g-dimensional superspecial abelian variety as an algebraic group isomorphic to $E_1 \times \cdots \times E_g$, where the $\{E_i\}$ are supersingular elliptic curves. In fact, by a remarkable theorem of Deligne, if $g > 1$ then for any fixed supersingular elliptic curve E, $E_1 \times \cdots \times E_g \cong E^g$ [Sh]. Fix such an elliptic curve E and let $\mathcal{O}$ be its endomorphism ring, which is a maximal order of $B_{p,\infty}$. Thus, what distinguishes superspecial abelian varieties with RM is compatible data:

$$\iota: \mathcal{O}_L \to M_g(\mathcal{O}) = \mathrm{End}(E^g), \quad \Xi \in M_g(\mathcal{O}),$$

where Ξ, which encodes the polarization, is a matrix that (1) is symmetric and positive relative to the Rosati involution $(m_{ij}) \mapsto (\overline{m}_{ji})$ induced by the product polarization of E^g; (2) commutes with $\iota(\mathcal{O}_L)$; (3) is a unit of $M_g(\mathcal{O})$. All this taken modulo the action of $\mathrm{Aut}(E^g) = M_g(\mathcal{O})^\times$. Note that this is rather explicit data.

There are finitely many points corresponding to ssAVs with RM in the moduli space $\overline{\mathcal{M}}_L$; let $\eta = \eta(L)$ denote their number. Explicit mass formulae exist (see [BG] for $g = 2$ and [Yu] in general), but a rough estimate is

$$(6.1) \qquad\qquad \eta \approx 2^{1-g} p^g |\zeta_L(-1)|,$$

where ζ_L is the Dedekind zeta function of L [CGL2, Proposition 3.6]. Using the functional equation for ζ_L, we have $\zeta_L(-1) = \frac{(-1)^g}{2^g \pi^{2g}} d_L^{3/2} \zeta_L(2)$, from which one can conclude that

$$(6.2) \qquad\qquad |\zeta_L(-1)| \approx \frac{1}{2^g \pi^{2g}} d_L^{3/2}.$$

So when g is held constant, the number of ssAVs with RM by $\mathcal{O}_L$ in characteristic p grows like $p^g d_L^{3/2}$ in the sense that the ratio is bounded above and below by constants that can be explicitly computed. There is much more to say, especially for L quadratic; see [Zag] for a masterful exposition.

Equivalent characterizations of A being superspecial are

(1) The degree p^g isogeny $\mathrm{Ver}_A := \mathrm{Fr}^\vee_{(A^{(1/p)})^\vee} : A \to A^{(1/p)}$, satisfies $\mathrm{Ker}(\mathrm{Ver}_A) = \mathrm{Ker}(\mathrm{Fr}_A)$;
(2) There is an embedding $\alpha_p^g \hookrightarrow A[p]$.

The equivalence of (1) and (2) follows immediately from the theory of finite commutative group schemes, and the equivalence of (2) with the definition of superspecial given above is a theorem due to Oort [Oor].

We remark that for $g > 1$, these are stronger requirements than just requiring that $A[p]$ has no nontrivial k-points. Using these characterizations, it is not hard to prove that if A is superspecial and $f \colon A \to B$ is an isogeny of g-dimensional abelian varieties of degree prime to p, then B is superspecial too; this statement may fail if $p \mid \deg(f)$.

Suppose that k is a field of characteristic p and $\underline{A}, \underline{B}$ are ssAV with RM over k; we let

$$\mathrm{Hom}_k(\underline{A}, \underline{B}) := \mathrm{Hom}_{\mathcal{O}_L, k}(A, B).$$

Namely, this is the $\mathcal{O}_L$-module of homomorphisms $A \to B$ that commute with the action of $\mathcal{O}_L$ and are defined over k. If k is algebraically closed, we typically omit the subscript k in $\mathrm{Hom}_k(\underline{A}, \underline{B})$.

If k is algebraically closed, it was proven in [Nic1] that $\mathrm{End}(\underline{A})$ is an order of discriminant ideal $p\mathcal{O}_L$ in the quaternion algebra $B_{p,L} := B_{p,\infty} \otimes_{\mathbb{Q}} L$. Note that the discriminant of the quaternion algebra $B_{p,L}$ may strictly contain $p\mathcal{O}_L$, in fact it may be as large as $\mathcal{O}_L$. Thus, $\mathrm{End}(\underline{A})$ need not be a maximal order, but it is always an Eichler order.

Example 6.1.1. Let E be a supersingular elliptic curve over $\mathbb{F}_{p^2}$ with endomorphism ring $\mathcal{O}$, a maximal order of $B_{p,\infty}$. The functor $E \otimes_{\mathbb{Z}} \mathcal{O}_L$ is the functor that takes a characteristic p commutative ring R to

$$(E \otimes_{\mathbb{Z}} \mathcal{O}_L)(R) := E(R) \otimes_{\mathbb{Z}} \mathcal{O}_L.$$

This functor is representable by E^g, where an isomorphism is established by choosing a $\mathbb{Z}$-basis of $\mathcal{O}_L$, but writing it as a tensor product as above endows it with a canonical action of $\mathcal{O}_L$: $\iota(\alpha)(P \otimes \beta) = P \otimes \alpha\beta$. There is also a compatible $\mathcal{O}_L$-linear polarization on $E \otimes_{\mathbb{Z}} \mathcal{O}_L$ because we have canonical identifications

$$(E \otimes_{\mathbb{Z}} \mathcal{O}_L)^\vee = E^\vee \otimes_{\mathbb{Z}} \mathfrak{d}_L^{-1} = E \otimes_{\mathbb{Z}} \mathfrak{d}_L^{-1},$$

and the map $\mathrm{Id} \otimes \delta_L^{-1}$ is the $\mathcal{O}_L$-linear principal polarization. We thus obtain the structure of an ssAV with RM, and one can verify directly that

$$\mathrm{End}\left(\underline{E \otimes_{\mathbb{Z}} \mathcal{O}_L}\right) = \mathrm{End}_{\mathcal{O}_L}(E \otimes_{\mathbb{Z}} \mathcal{O}_L) = \mathcal{O} \otimes_{\mathbb{Z}} \mathcal{O}_L \subset B_{p,\infty} \otimes_{\mathbb{Q}} L = B_{p,L}.$$

The paper [Con] contains useful results about such tensor constructions.

6.2. **RM superspecial isogeny graphs.** Our main reference is [CGL2], where many of the missing details can be found, and where the construction is given in much greater generality.

With the conventions of §6.1, let $j_1, \ldots, j_\eta$ ($\eta = \eta(L)$) denote the superspecial points of $\overline{\mathcal{M}}_L$ and let $\underline{A}(j_1), \ldots, \underline{A}(j_\eta)$ be the corresponding ssAVs with RM. In fact, we can choose for each a model over $\mathbb{F}_{p^2}$ such that $\mathrm{Fr}^2_{A(j_k)} = [-p]$, and so for any two we have

$$\mathrm{Hom}_{\overline{\mathbb{F}}_p}(\underline{A}(j_a), \underline{A}(j_b)) = \mathrm{Hom}_{\mathbb{F}_{p^2}}(\underline{A}(j_a), \underline{A}(j_b)).$$

In particular, $\mathrm{Hom}_{\mathbb{F}_{p^2}}(\underline{A}(j_a), \underline{A}(j_b))$ is a module of rank 4 over $\mathcal{O}_L \subset \mathrm{End}(A(j_a))$. Nicole [Nic1] shows that it is an integral quadratic module as follows:

Write $\underline{A}(j_a) = (A_a, \iota_a, \lambda_a)$ and similarly for $\underline{A}(j_b)$. Given $f \in \mathrm{Hom}_{\overline{\mathbb{F}}_p}(\underline{A}(j_a), \underline{A}(j_b))$ consider $\lambda_a^{-1} f^\vee \lambda_b f$. This homomorphism belongs to $\mathrm{End}(\underline{A}(j_a))$ and is fixed by the Rosati involution associated to λ_a (this is the involution that takes $h \in \mathrm{End}(\underline{A}(j_a))$ to $\lambda_a^{-1} h^\vee \lambda_a$), because, being polarizations, λ_a and λ_b are self-dual. The elements of $\mathrm{End}(\underline{A}(j_a))$ fixed by the Rosati involution are precisely $\mathcal{O}_L$, so we get a quadratic form

$$\mathrm{Hom}_{\overline{\mathbb{F}}_p}(\underline{A}(j_a), \underline{A}(j_b)) \to \mathcal{O}_L, \qquad f \mapsto \deg_L(f) := \lambda_a^{-1} f^\vee \lambda_b f.$$

As in the case of elliptic curves, all these quadratic modules are in the same genus and so to establish some of their properties one may reduce to the case of a single ssAV with RM $\underline{A} = (A, \iota, \lambda)$ and its endomorphism ring $\mathrm{End}(\underline{A})$. It is convenient to take $\underline{A} = E \otimes \mathcal{O}_L$ and $\mathrm{End}(\underline{A}) = \mathcal{O} \otimes \mathcal{O}_L$ as in Example 6.1.1, and this can be made as explicit as one wants by a judicious choice of E. The quadratic form on this rank 4 $\mathcal{O}_L$-module is

$$(6.3) \qquad\qquad f \mapsto \lambda^{-1} f^\vee \lambda f = f^\dagger f,$$

where $f^\dagger := \lambda^{-1} f^\vee \lambda$ is the image of f under the Rosati involution $f \mapsto f^\dagger$ associated to λ. In the model $\mathcal{O} \otimes \mathcal{O}_L$ the quadratic form is just the map $f \otimes \alpha \mapsto \mathrm{Nm}(f) \otimes \alpha^2$. In this form we can see that the map is a totally definite quadratic form, by which we mean that under any embedding $\sigma_i \colon L \to \mathbb{R}$ this form is positive-definite as a real-valued form.

Even more concretely, $\mathcal{O}$ always contains a finite index lattice $\Lambda = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}k$, where $i^2 = a$ and $j^2 = b$ are negative integers, $k = ij = -ji$, and the norm form on this lattice is simply $x^2 - \alpha y^2 - \beta z^2 + \alpha\beta w^2$ for $x, y, z, w \in \mathbb{Z}$. Then $\mathcal{O} \otimes \mathcal{O}_L$ contains the $\mathcal{O}_L$ lattice $\Lambda \otimes \mathcal{O}_L = \mathcal{O}_L \oplus \mathcal{O}_L i \oplus \mathcal{O}_L j \oplus \mathcal{O}_L k$ and the norm form is

$$(6.4) \qquad\qquad x^2 - \alpha y^2 - \beta z^2 + \alpha\beta w^2, \quad x, y, z, w \in \mathcal{O}_L.$$

We now construct a graph $\mathscr{G}_L(p, \ell)$ (an **RM superspecial isogeny graph**) whose vertices $1, \ldots, \eta$ correspond to the superspecial points $j_1, \ldots, j_\eta$ of $\overline{\mathcal{M}}_L$ or, equivalently, to the ssAVs with RM $\underline{A}(j_1), \ldots \underline{A}(j_\eta)$ (which we relabel more simply as $\underline{A}_1, \ldots, \underline{A}_\eta$). Each $\underline{A}_j$ has a preferred model as chosen above. Consider a vertex $\underline{A} = (A, \iota, \lambda) \in \{\underline{A}_1, \ldots, \underline{A}_\eta\}$. Since ℓ is inert in $L/\mathbb{Q}$ we have $\underline{A}[\ell] \cong (\mathcal{O}_L/\ell\mathcal{O}_L)^2 \cong \mathbb{F}_{\ell^g}^2$, and so every proper nontrivial $\mathcal{O}_L$-invariant subgroup scheme H of $\underline{A}[\ell]$ is isomorphic to $\mathcal{O}_L/\ell\mathcal{O}_L$. There are $\ell^g + 1$ of these, and they are automatically isotropic relative to the polarization λ. Thus, by Mumford's descent theory, there is a unique $\mathcal{O}_L$-structure ι' and principal $\mathcal{O}_L$-polarization μ on A/H such that under the isogeny $f \colon A \to A/H$ we have $f^*\mu = \ell\lambda$. We connect the vertex $\underline{A}$ to all the vertices $\underline{A/H} = (A/H, \iota', \mu)$ constructed this way. Note also that any $\mathcal{O}_L$-homomorphism $\underline{A} \to \underline{A}'$ of degree ℓ^g and whose kernel is

contained in $\underline{A}[\ell]$ must arises this way. We may thus also think of edges as defined by such homomorphisms, up to automorphisms of the image $\underline{A}'$. This gives us a directed graph $\mathscr{G}_L(p, \ell)$ with η vertices, all with out-degree $\ell^g + 1$. These graphs were explored in [CGL2]; for related previous work see [Liv] and references therein.

Some important properties that these graphs have in common with supersingular isogeny graphs of elliptic curves are that (1) they are connected, (2) they have very few loops and multiple edges, and (3) they are Ramanujan [CGL2]. They are thus suitable candidates for generalizing the CGL hash function and other cryptographic applications. It will be interesting to consider which other known properties of the graphs $\mathscr{G}(p, \ell)$ generalize to $\mathscr{G}_L(p, \ell)$.

The RM superspecial isogeny graphs $\mathscr{G}_L(p, \ell)$ are related to another class of graphs, which have vertices corresponding to principally polarized superspecial abelian varieties of dimension g, and edges corresponding to isogenies whose kernel is a maximal isotropic subgroup of the ℓ-torsion. Call this graph $\mathscr{G}_S^g(p, \ell)$ ("S" for Siegel). In general, the graphs $\mathscr{G}_S^g(p, \ell)$ are *not* Ramanujan [JZ, Section 9], demonstrating that the presence of RM structure results in a qualitative difference in the properties of the resulting graph. However, when L is a degree g totally real field of strict class number 1 and ℓ is inert in $L/\mathbb{Q}$, there is a natural map of graphs

$$(6.5) \qquad\qquad\qquad \mathscr{G}_L(p, \ell) \to \mathscr{G}_S^g(p, \ell)$$

that takes $\underline{A} = (A, \lambda, \iota)$ to (A, λ). The edges $\underline{A} \to \underline{A}/H$ in $\mathscr{G}_L(p, \ell)$ go to edges in $\mathscr{G}_S^g(p, \ell)$, because any such H is automatically a maximal isotropic subgroup of the ℓ-torsion.

We now focus on the case $g = 2$ and present a heuristic to suggest that the map (6.5) is in fact surjective for infinitely many L. Let $d \equiv 1 \pmod 4$ be prime and set $L = \mathbb{Q}(\sqrt{d})$; note that for any such field the class number h_L and strict class number h_L^+ are equal ([Bue, Theorems 6.19 and 9.3]). To have ℓ inert in L is the condition that d is not a square modulo ℓ if $\ell > 2$, and $d \equiv 5 \pmod 8$ if $\ell = 2$. A positive proportion of primes d satisfy these conditions, and by the Gauss conjecture, or the Cohen-Lenstra heuristics, one expects the majority of real quadratic fields to have class number 1; taking the intersection, one also expects there to be infinitely many real quadratic fields L of strict class number 1 in which ℓ is inert.

Combining (6.1) and (6.2) one sees that the number of vertices in $\mathscr{G}_L(p, \ell)$ grows to infinity like $d_L^{3/2}$; on the other hand, the number of vertices of $\mathscr{G}_S^2(p, \ell)$ remains constant, and so one would expect the map $\mathscr{G}_L(p, \ell) \to \mathscr{G}_S^2(p, \ell)$ to be surjective on vertices for all L with d_L sufficiently large. This is indeed the case [GL2, Theorem 2.4]. Thus the security of the graphs $\mathscr{G}_L(p, \ell)$ and $\mathscr{G}_S^2(p, \ell)$ appears to be linked.

We return to our original setting where the degree g of $L/\mathbb{Q}$ is arbitrary. In [CGL2] more general versions of the graphs $\mathscr{G}_L(p, \ell)$ are defined, associated to any prime ideal $\mathfrak{l} \subset \mathcal{O}_L$ that is prime to p. The edges are related to isogenies $f : \underline{A} \to \underline{B}$ of ssAVs with RM for which $\mathrm{Ker}(f)$ and $\mathcal{O}_L/\mathfrak{l}$ are isomorphic over $\overline{\mathbb{F}}_p$ as $\mathcal{O}_L$-modules. The astute reader may wonder about the existence of a principal polarization on $\underline{B}$; this is one place where we use that the strict class number of L is 1. This assumption allows one to essentially ignore polarizations entirely, for if λ, μ are two principal $\mathcal{O}_L$-polarizations on (A, ι), then the fact that $\mathcal{O}_L$ has strict class number 1 implies that $\mu = \epsilon^2 \lambda$ for some $\epsilon \in \mathcal{O}_L^\times$, so multiplication by ϵ induces an isomorphism $(A, \iota, \lambda) \cong (A, \iota, \mu)$.

In his thesis [Nic1], see also [Nic2], Nicole has developed a Deuring correspondence for superspecial abelian varieties with RM under the strict class number 1 assumption, though there is little doubt that this assumption can be removed. Fixing one superspecial point, say $E \otimes \mathcal{O}_L$, there is a bijection between right ideal classes of the order $\mathrm{End}(E \otimes \mathcal{O}_L) = \mathcal{O} \otimes \mathcal{O}_L$ and isomorphism classes of ssAV with RM. Many of the properties of the Deuring correspondence carry over to this new context, and the theory fits perfectly with a theory of Brandt matrices as in [Eic2]; see [CGL2].
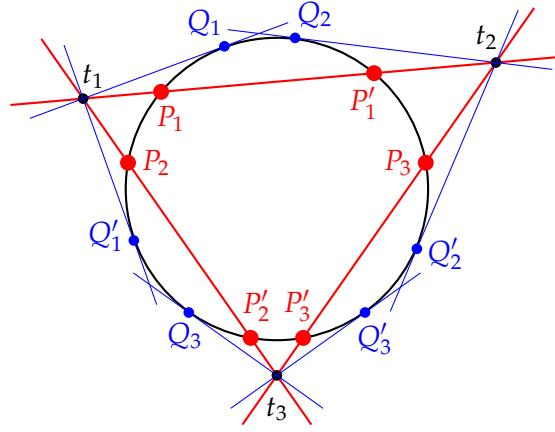
FIGURE 5. Defining a Richelot isogeny between genus 2 hyperelliptic curves defined by ramified covers of a conic.

6.2.1. *Structure and applications of RM superspecial isogeny graphs.* Setting a CGL hash function for such graphs is not immediately obvious; one needs a way to encode the vertices, and to enumerate the edges out of a given vertex in a consistent way. We explain the difficulties.

Given the Jacobian of a genus 2 curve, the following technique can be used to compute all isogenies from this Jacobian such that the kernel is a maximal isotropic subgroup of the 2-torsion. Take 6 points $\mathscr{P}$ in the complex plane that lie on a conic C; they define a unique double cover $X_1 \to C \cong \mathbb{P}^1$ ramified exactly along $\mathscr{P}$. Divide the points of $\mathscr{P}$ into 3 pairs, $\{P_1, P_1'\}, \{P_2, P_2'\}, \{P_3, P_3'\}$ (there are 15 ways to do that). Let L_i be the line through P_i, P_i' and let t_1, t_2, t_3 the three points where two of the lines L_1, L_2, L_3 intersect. Through each point t_i there are two lines that are tangent to C and intersect it in a pair of points $\{Q_i, Q_i'\}$ and that altogether, usually, provide us with 6 points $\mathscr{Q}$ lying on C and so with another double cover $X_2 \to C$ ramified now along the points of $\mathscr{Q}$. One can prove that this corresponds to an isogeny $f \colon \operatorname{Jac}(X_1) \to \operatorname{Jac}(X_2)$ whose kernel is a $(2,2)$ maximal isotropic subgroup of $\operatorname{Jac}(X_1)[2]$ (of which there are 15). And in fact, one gets all such subgroups by a suitable partition of $\mathscr{P}$. Here one uses that by means of the Weierstrass points one can write all 2-torsion points on $\operatorname{Jac}(X_1)$; in fact the differences $\alpha - \beta$ with $\alpha, \beta \in \mathscr{P}$ give all the 2-torsion points of $\operatorname{Jac}(X_1)$ and the kernel of f is then given by $\{0\} \cup \{(P_i + P_i') - (P_j + P_j') : 1 \leq i < j \leq 3\}$. Moreover, the isogeny defined by the partition $\{Q_i, Q_i'\}_{i=1}^3$ of $\mathscr{Q}$ is exactly the dual isogeny $f^\vee \colon \operatorname{Jac}(X_2) \to \operatorname{Jac}(X_1)$. See [BM, DL].

This works rather well and has been implemented for walks in $\mathscr{G}_S^2(p, 2)$ [CDS, FT]. When we want to adapt this to the case of superspecial abelian surfaces with real multiplication by $\mathcal{O}_L$, certain problems arise; we think they are attractive and hope that an interested reader would resolve them. Assume for example that 2 is split in $\mathcal{O}_L$. (As a concrete example we may take the superspecial curve $X : y^2 = x^5 + 1$, in characteristic $p \equiv 4 \pmod 5$, that has real multiplication by $L = \mathbb{Q}(\sqrt{5})$. Then, by [Gor1], $\operatorname{Jac}(X)$ is a ssAV with RM.) The group $\operatorname{Jac}(X_1)[2]$ is a free module of rank 2 over $\mathcal{O}_L/2\mathcal{O}_L \cong \mathbb{F}_2 e_1 \oplus \mathbb{F}_2 e_2$. An analysis of the interaction of the pairing with the $\mathcal{O}_L$-structure, shows that there are nine $\mathcal{O}_L$-invariant maximal isotropic subgroup of $\operatorname{Jac}(X_1)[2]$ (a choice of line in every isotypic $\mathcal{O}_L$-submodule of $\operatorname{Jac}(X_1)[2]$). But how to determine which of the fifteen isotropic subgroups are the $\mathcal{O}_L$-invariant ones? This necessitates understanding the action of $\frac{1+\sqrt{5}}{2}$ on $\operatorname{Jac}(X_1)[2]$, or as a correspondence on the conic $C \cong \mathbb{P}^1$. Another problem that arises is how to determine the action $\frac{1+\sqrt{5}}{2}$ on $\operatorname{Jac}(X_2)[2]$. Nonetheless, for certain totally real

fields, for example for $\mathbb{Q}(\sqrt{5})$ there is a lot known. For the state of the art see [CM, CFM] and the references therein.

If such an implementation is possible, then by analogy with the elliptic curve case, one would expect the security of such hash functions to rest on the difficulty of computing endomorphism rings of ssAVs with RM. Likewise, one would expect there to be an effective method using closed paths in $\mathscr{G}_L(p, \ell)$ to generate finite index subrings of $\mathrm{End}(\underline{A})$.

In fact, one can prove quite easily that the endomorphisms $f \in \mathrm{End}(\underline{A})$ satisfying $\deg_L(f) \in \{\ell^n : n \in \mathbb{Z}_{\geq 0}\}$ generate over $\mathcal{O}_L$ a subring of finite index of $\mathrm{End}(\underline{A})$. We prove this first for $\underline{A}$ of the form $E \otimes_{\mathbb{Z}} \mathcal{O}_L$, with the induced $\mathcal{O}_L$-structure and polarization discussed in Example 6.1.1. Since we allow $\mathcal{O}_L$-linear combinations, it is enough to prove that endomorphisms with ℓ-power degree generate a finite-index sublattice of $\mathrm{End}(E)$. While this statement follows of course from Theorem 5.2.2, it can also be easily derived from the earlier criterion of Bank et. al. (§4.1.2), and can be proven even more directly as follows.

Suppose K is an imaginary quadratic field field in which p is non-split and ℓ is split; note that this is just a congruence condition on the discriminant of K, so one can find infinitely many such fields. Let $\mathfrak{l}$ be a prime of K dividing ℓ, and let a be a positive integer such that $\mathfrak{l}^a$ is a principal ideal with generator f. Let $\tilde{E}'$ be an elliptic curve with CM by $\mathcal{O}_K$, defined over a number field M and having good reduction at some prime above p. The reduction E' by this prime then has CM by $\mathcal{O}_K$ as well, so we can view f as an element of $\mathrm{End}(E')$. Then

$$\deg(f) = N_{K/\mathbb{Q}}(f) = \ell^a,$$

and since $\mathrm{Ker}(f) \cong \mathcal{O}_K/\mathfrak{l}^a$ is cyclic of order ℓ^a we see that $f \notin \mathbb{Z}$. Using that the graph $\mathscr{G}(p, \ell)$ is connected, we can find an ℓ-power isogeny $h \colon E \to E'$, and then $\phi := h^\vee f h$ is an endomorphism of E with degree a power of ℓ and $\mathbb{Q}(\phi) \cong \mathbb{Q}(f) = K$, as ϕ has the same minimal polynomial as $\deg(h)f$. If we run a similar construction with two distinct fields K_1, K_2, we obtain endomorphisms ϕ_1 and ϕ_2 of E that generate different quadratic subfields of $\mathrm{End}(E) \otimes \mathbb{Q}$ and thus the lattice spanned by $\{1, \phi_1, \phi_2, \phi_1 \phi_2\}$ (all elements of norm a power of ℓ) has finite index in $\mathrm{End}(E)$.

We now use a similar argument to bootstrap from $B := E \otimes \mathcal{O}_L$ to any superspecial $\underline{A}$. As the graph $\mathscr{G}_L(p, \ell)$ is connected, there is a path from $\underline{A}$ to $\underline{B}$ and this defines an isogeny $h \colon \underline{A} \to \underline{B}$, well-defined up to $\mathrm{End}(\underline{B})^\times$. There is also an isogeny $\tilde{h} \colon \underline{B} \to \underline{A}$ with the property that $\tilde{h} \circ h = [\ell^r]$ for some r, obtained by selecting r so that $\mathrm{Ker}(h) \leq \underline{A}[\ell^r]$ and then taking the quotient of $\underline{B}$ by $h(\underline{A}[\ell^r])$. Consider the homomorphism of $\mathcal{O}_L$-modules

(6.6) $\mathrm{End}(\underline{B}) \to \mathrm{End}(\underline{A}), \qquad f \mapsto \tilde{h} f h.$

We claim that $\deg_L(\tilde{h} f h) = \ell^{2r} \deg_L(f)$ for all $f \in \mathrm{End}(\underline{B})$. Indeed, using λ_E to denote the principal polarization on $E \otimes \mathcal{O}_L$ and λ_A the one appearing in $\underline{A}$, the diagram

(6.7)
$$
\begin{array}{ccccccc}
A^\vee & \xleftarrow{h^\vee} & B^\vee & \xleftarrow{f^\vee} & B^\vee & \xleftarrow{\tilde{h}^\vee} & A^\vee \\
{\scriptstyle \lambda_A^{-1}}\Big\downarrow & & {\scriptstyle \lambda_B}\Big\Updownarrow{\scriptstyle \lambda_B^{-1}} & & {\scriptstyle \lambda_B}\Big\Updownarrow{\scriptstyle \lambda_B^{-1}} & & {\scriptstyle \lambda_A}\Big\uparrow \\
A & \xrightarrow{h} & B & \xrightarrow{f} & B & \xrightarrow{\tilde{h}} & A
\end{array}
$$

proves that

$$\deg_L(\tilde{h} f h) = \deg_L(\tilde{h}) \deg_L(f) \deg_L(h).$$

Note that these are elements of $\mathcal{O}_L$ and therefore commute. A similar diagram without the middle square shows

$$\deg_L(\tilde{h}) \deg_L(h) = \deg_L([\ell^r]) = \ell^{2r}.$$

Thus we find $\deg_L(\tilde{h} f h) = \ell^{2r} \deg_L(f)$ as desired, so the image of (6.6) is simply a dilation of $\mathrm{End}(\underline{B})$. We showed above that $\mathrm{End}(\underline{B})$ has a finite index sub-lattice generated by elements with

$\deg_L(f)$ a power of ℓ, and the image of this sub-lattice is finite index in $\text{End}(\underline{A})$ and generated by elements with $\deg_L(f)$ a power of ℓ.

Once again we find ourselves led to the question of whether a lattice can be generated by particular sets of elements, only in this case the lattices are now $\mathcal{O}_L$-modules. For the rest of this article we turn our attention to this problem. The key result we are building towards is the following local-global principle for lattice cosets over totally real fields; see §6.3.1 for the definitions of any terms that have not yet been defined. The proof is completed in §6.5.3, and then we give applications in §6.6.

Theorem 6.2.1. *Let $n \geq 4$ be an even integer, let Λ be an integral $\mathcal{O}_L$-lattice of rank n with positive-definite quadratic form Q, and let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_L$. Let $t \in \Lambda \setminus \mathfrak{p}\Lambda$ and set $M := t + \mathfrak{p}\Lambda$. For any $\mu \in \mathcal{O}_L$ coprime to $\text{disc}\,\Lambda$ with $N_{L/\mathbb{Q}}(\mu)$ sufficiently large, if for all places v of L there exists $x \in M_v := t + (\mathfrak{p}\Lambda \otimes_{\mathcal{O}_L} \mathcal{O}_{L,v})$ with $Q(x) = \mu$, then there exists $x \in M$ with $Q(x) = \mu$.*

Remark 6.2.2. The restriction that n be even can almost certainly be removed, but justifying this would add significantly to the length of the paper and is not necessary for our main application. See Remark 6.5.7 for more details.

This theorem (with the n even restriction removed) is a consequence of Chan–Oh's [CO, Theorem 4.9]. Indeed, note that if Q is anisotropic on V_v then v divides $\text{disc}\,Q$ [Han1, Remark 3.8.1], and so if μ is coprime to $\text{disc}\,\Lambda$ and is represented by M_v then it must be primitively represented by M_v. However, the proof of [CO, Theorem 4.9] for the case $n = 4$ depends critically on an unpublished result. With considerable effort, Theorem 6.2.1 can also be deduced from [Shi5].

6.3. **Lattices and cosets over totally real fields.** As a general reference to quadratic forms we use [Ome], especially Part IV, although our terminology differs at times; see also [Han1, Han2]. Let L be a totally real field of degree g over $\mathbb{Q}$ and $\mathcal{O}_L$ its ring of integers. Let $\sigma_1, \ldots, \sigma_g \colon L \to \mathbb{R}$ denote its infinite places. For each finite place v of L, we let L_v denote the completion at v and $\mathcal{O}_{L,v}$ its valuation ring, i.e. the completion of $\mathcal{O}_L$ at v.

6.3.1. *Quadratic spaces and lattices.* By a **quadratic space**, we mean an n-dimensional vector space V over L, for some integer $n > 0$, together with an L-bilinear symmetric form:

$$(\cdot, \cdot) \colon V \times V \to L.$$

To such a bilinear form we associate the quadratic form $Q(x) = (x, x)$ and so we have the relations

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

One can therefore also define a quadratic space by means of the quadratic form Q, which is simply a function $Q \colon V \to L$ such that the form $(\cdot, \cdot)$ thus defined is L-bilinear. We shall exclusively consider **positive-definite** quadratic forms Q, meaning that under any embedding $\sigma_i \colon L \to \mathbb{R}$, the induced real-valued quadratic form is positive-definite. Therefore $Q : (V \setminus \{0\}) \to L$ takes values in the set L^+ comprising $x \in L$ such that $\sigma_i(x) > 0$ for all embeddings σ_i.

By a **quadratic module** (or a **lattice**) Λ over L we mean a finite rank $n \geq 1$ projective $\mathcal{O}_L$-module, together with a function

$$Q \colon \Lambda \to L,$$

such that the associated function $(\cdot, \cdot)$ is a bilinear form. Then $V = \Lambda \otimes_{\mathcal{O}_L} L$ is a quadratic space. Conversely, a finitely generated $\mathcal{O}_L$-submodule Λ of a quadratic space V, that contains a basis of V, is a lattice. If Q takes values in $\mathcal{O}_L$, we refer to Λ as an **integral lattice**. By a procedure similar to the one for $\mathbb{Z}$-lattices, one can associate to Λ a discriminant $\mathfrak{d}_\Lambda$, which is a fractional ideal of L.

We will also define local and adelic versions of these structures. Let $\mathbb{A}$ denote the adeles of L. That is,

$$\mathbb{A} = \prod_{v}{}' L_v,$$

which is the restricted product over all places of L (for each element, all but finitely many components are in the valuation ring $\mathcal{O}_{L,v}$). Denote also $\mathbb{A}_f$ the restricted product over all finite places, and $\mathbb{A}_\infty \cong L \otimes_{\mathbb{Q}} \mathbb{R}$ the product over the infinite places. Thus, $\mathbb{A} = \mathbb{A}_f \times \mathbb{A}_\infty$. Let O_Q denote the orthogonal group of the quadratic form Q: that is, the L-linear automorphisms of V that preserve Q. This is an algebraic group over L.

Given a finite place v of L and a lattice Λ, we let $\Lambda_v := \Lambda \otimes_{\mathcal{O}_L} \mathcal{O}_{L,v}$, which is a quadratic module in the quadratic space $V_v := V \otimes_L L_v$ with quadratic form $Q \colon \Lambda_v \to L_v$. For finite places v, we write $O_Q(\mathcal{O}_v)$ to denote elements of $O_Q(L_v)$ that stabilize Λ_v.

We can define adelic structures as restricted products of these local structures. First we have $\Lambda_{\mathbb{A}_f} = \prod_{v < \infty} \Lambda_v$. We define $V_{\mathbb{A}}$ and $V_{\mathbb{A}_f}$ as the restricted product of V_v over all (resp. all finite) places v, where all but finitely many components lie in Λ_v. Similarly, we define $O_Q(\mathbb{A})$ and $O_Q(\mathbb{A}_f)$ as the restricted product of $O_Q(L_v)$ over all (resp. all finite) places v, where all but finitely many components lie in $O_Q(\mathcal{O}_v)$.

6.3.2. *Lattice cosets and their genera.* See [Shi5, KK, CO]. Given an $\mathcal{O}_L$-lattice Λ in V and a vector $t \in V$, we say that $M = \Lambda + t$ is a **lattice coset** in V with modulus Λ. The **conductor** $\mathfrak{c}$ of M is defined to be the ideal of elements $c \in \mathcal{O}_L$ satisfying $cM \subseteq \Lambda$ (or equivalently $ct \in \Lambda$). Note that M is a lattice if and only if $t \in \Lambda$ if and only if $\mathfrak{c} = \mathcal{O}_L$.

Given a finite place v of L and a lattice coset M with modulus Λ, we let

$$M_v := \Lambda_v + t \subseteq V \otimes_L L_v;$$

such M_v will be called a **local lattice coset**. Note that $M_v = \Lambda_v$ at all finite places v coprime to the condcutor of M. The lattice coset M is uniquely determined by the collection of local lattice cosets M_v. Conversely, suppose we are given a collection of local lattice cosets $\Lambda'_v + t'_v$ for each finite v, where Λ'_v is a lattice in V_v and $t'_v \in V_v$. If $\Lambda'_v = \Lambda_v$ for almost all v, and there exists a positive integer N such that $Nt'_v \in \Lambda'_v$ for all v, then there is a unique $\mathcal{O}_L$-lattice coset $\Lambda' + t$ in V whose completions are equal to the $\Lambda'_v + t'_v$. Thus, for any $\gamma = (\gamma_v)_v \in O_Q(\mathbb{A}_f)$ there is a lattice coset in V, denoted γM, such that $(\gamma M)_v = \gamma_v M_v$ for all v.

Given another lattice coset $M' = \Lambda' + t'$, where Λ' is another $\mathcal{O}_L$-lattice in the same quadratic space V, we say that M is **isomorphic** (or **isometric**) to M' if $\gamma M = M'$ for some $\gamma \in O_Q(L)$. Note that this is equivalent to the conditions $\gamma \Lambda = \Lambda'$ and $\gamma t - t' \in \Lambda'$. For a finite place v we say M and M' are **locally isomorphic** at v if $\gamma_v M_v = M'_v$ for some $\gamma_v \in O_Q(L_v)$. We say that M' is in the **genus** $\mathfrak{g} = \mathfrak{g}(M)$ of M if the two lattice cosets are locally isomorphic at every finite place, or equivalently if $\gamma M = M'$ for some $\gamma \in O_Q(\mathbb{A}_f)$. We let $[\mathfrak{g}] = [\mathfrak{g}(M)]$ denote the set of isomorphism classes of lattice cosets within the genus. This is a finite set: see [Ome, 103:4 Theorem] for lattices, and the result for general lattice coset follows from this by a simple argument given below. We define the **genus number** $h(\mathfrak{g})$ to be the cardinality of $[\mathfrak{g}]$.

Let M be a lattice coset with modulus Λ, and $\gamma = (\gamma_v)_v \in O_Q(\mathbb{A})$. Recall that, by the definition of $O_Q(\mathbb{A})$, γ_v preserves Λ_v for all but finitely many v. Let

$$K^M_{\mathbb{A}} := \{\gamma \in O_Q(\mathbb{A}) : \gamma M = M\}$$

denote the **adelic stabilzer** of M. Then we can write $K^M_{\mathbb{A}} = \prod_v K^M_v$, where for v infinite K_v is the compact group $O_Q(L_v)$, and for v finite K_v is the compact open group $\{\gamma_v \in O_Q(L_v) : \gamma_v M_v = M_v\}$. Thus, $K^M_{\mathbb{A}}$ is a compact open subgroup of $O_Q(\mathbb{A})$, and there is a bijection

$$O_Q(\mathbb{A})/K^M_{\mathbb{A}} \longleftrightarrow \mathfrak{g}(M),$$
$$\gamma \mapsto \gamma M.$$

This induces a bijection

(6.8)
$$O_Q(L) \backslash O_Q(\mathbb{A}) / K_{\mathbb{A}}^M \longleftrightarrow [\mathfrak{g}(\Lambda)].$$

For $M = \Lambda + t$ of conductor $\mathfrak{c}$, observe that any automorphism of Λ_v that acts as the identity on the finite group $\Lambda_v / \mathfrak{c}\Lambda_v$ will stabilize M_v. Thus

$$\{\gamma \in K_{\mathbb{A}}^\Lambda : \gamma_v \equiv Id_v \pmod{\mathfrak{c}\Lambda_v} \text{ for all finite } v\} \subseteq K_{\mathbb{A}}^M \subseteq K_{\mathbb{A}}^\Lambda,$$

and since $\mathfrak{c}\Lambda_v = \Lambda_v$ for all but finitely many places, the left-hand group is finite index in $K_{\mathbb{A}}^\Lambda$. Hence the index of $K_{\mathbb{A}}^M$ in $K_{\mathbb{A}}^\Lambda$ is finite, showing that finiteness of the genus number of M follows from finiteness of the genus number of the lattice Λ. Note that one may also obtain finiteness of the genus number $h(\mathfrak{g}(M))$ directly from Equation (6.8) using general properties of reductive groups [PR, Theorems 5.1 & 8.1].

6.3.3. *Adelic double cosets.* For a given genus $\mathfrak{g} = \mathfrak{g}(M)$, let $\mathscr{B} = \{\gamma_1, \ldots, \gamma_{h(\mathfrak{g})}\} \subset \mathcal{O}_Q(\mathbb{A})$ be a finite set such that the lattice cosets $\{M_i := \gamma_i M : 1 \le i \le h(\mathfrak{g})\}$ are a set of representatives for the isomorphism classes of lattice cosets in $\mathfrak{g}$. We let $\gamma_1 = Id$ so that $M_1 = M$, and without loss of generality we can take the archimedean components of each γ_i to be the identity. The elements γ_i give us an explicit parametrization of the double cosets in Equation (6.8):

(6.9)
$$O_Q(\mathbb{A}) = \bigsqcup_{i=1}^{h(\mathfrak{g})} O_Q(L)\gamma_i K_{\mathbb{A}}^M.$$

These double cosets are not necessarily all the same size. More precisely, we consider the image of each double coset $O_Q(L)\gamma_i K_{\mathbb{A}}^M$ in the compact quotient $O_Q(L)\backslash O_Q(\mathbb{A})$; this is the same as the image of $\gamma_i K_{\mathbb{A}}^M$. Two elements $h, h' \in \gamma_i K_{\mathbb{A}}^M$ have the same image in $O_Q(L)\backslash O_Q(\mathbb{A})$ if and only if $h'h^{-1}$ is in the finite group $O_Q(L) \cap \gamma_i K_{\mathbb{A}}^M \gamma_i^{-1}$, which is the stabilizer $\mathrm{Aut}(M_i) \le O_Q(L)$ of M_i. Thus, letting

$$w_i := \#(O_Q(L) \cap \gamma_i K_{\mathbb{A}}^M \gamma_i^{-1}) = \sharp \mathrm{Aut}(M_i)$$

denote the cardinality of this group, we see that the map from $\gamma_i K_{\mathbb{A}}^M$ to $O_Q(L)\backslash O_Q(\mathbb{A})$ is w_i-to-one.

Consider a Haar measure on $O_Q(\mathbb{A})$ normalized so that the compact open subgroup $K_{\mathbb{A}}^M$ has measure 1. For any $\gamma M \in \mathfrak{g}(M)$ the adelic stabilizer of γM is $\gamma K_{\mathbb{A}}^M \gamma^{-1}$, which also has measure 1 becaues $O_Q(\mathbb{A})$ is unimodular (this follows for example from [Bor, Lemma 5.5] because $O_Q(L_v)$ is compact at all infinite v). Thus this choice of Haar measure depends only on the genus $\mathfrak{g}$.

The left coset $\gamma_i K_{\mathbb{A}}^M$ also has measure 1, so the induced measure of its image in $O_Q(L)\backslash O_Q(\mathbb{A})$ is $\frac{1}{w_i}$. As a consequence, we can compute the volume of $O_Q(L)\backslash O_Q(\mathbb{A})$ with respect to the induced Haar measure to be

(6.10)
$$m(\mathfrak{g}) := \mathrm{Vol}_{\mathfrak{g}}(O_Q(L)\backslash O_Q(\mathbb{A})) = \sum_{i=1}^{h(\mathfrak{g})} \frac{1}{w_i},$$

where we write $\mathrm{Vol}_{\mathfrak{g}}$ to emphasize that the choice of measure depends on $\mathfrak{g}$. We call $m(\mathfrak{g})$ the **mass** of the genus $\mathfrak{g}$. Thus any fundamental domain for the action of $O_Q(L)$ on $O_Q(\mathbb{A})$ has volume $m(\mathfrak{g})$ with respect to this measure, and the intersection of the double coset $O_Q(L)\gamma_i K_{\mathbb{A}}^M$ with any such fundamental domain has measure $\frac{1}{w_i}$.

6.4. **Hilbert modular forms and adelization.** In this section, we discuss some general facts about Hilbert modular forms and their adelic versions. For motivation, we begin in §6.4.1 by introducing the Hilbert modular forms that will be used for our main results and their relation to the theory of lattice cosets just discussed. The rest of the section is devoted to the general theory. We return to our specific focus on the theta functions associated to lattice cosets in §6.5.

6.4.1. *Inhomogeneous theta functions of totally real lattice cosets.* See especially [Shi5]. The setup is nearly identical to the homogeneous case (that is, the case M is a lattice); for this case see for example [Eic2, KK, Wal1], though mind the different normalizations across these references.

Let $\mathfrak{H}$ denote the complex upper half plane. As above we let L be a totally real field of degree g over $\mathbb{Q}$, and Λ a rank n quadratic module over $\mathcal{O}_L$. To a lattice coset M with modulus Λ we can associate a **theta function** $\theta_M : \mathfrak{H}^g \to \mathbb{C}$, defined for $\tau := (\tau_1, \dots, \tau_g) \in \mathfrak{H}^g$ by

$$(6.11) \qquad \theta_M(\tau) := \sum_{x \in M} e^{2\pi i \operatorname{Tr} Q(x) \cdot \tau}$$

where we write $\operatorname{Tr} \mu \cdot \tau := \sigma_1(\mu)\tau_1 + \cdots + \sigma_g(\mu)\tau_g$ for $\mu \in L$. We define the **representation numbers** for the lattice coset M and $\mu \in L$ by

$$r_M(\mu) := \sharp\{x \in M : Q(x) = \mu\}.$$

Note that $r_M(0)$ is 1 if $M = \Lambda$ is a lattice and 0 otherwise. Recalling that Q is positive-definite, so $Q(x) \in L^+$ for all nonzero $x \in V$, we can rewrite the theta function as

$$(6.12) \qquad \theta_M(\tau) = r_M(0) + \sum_{\mu \in L^+} r_M(\mu) e^{2\pi i \operatorname{Tr} \mu \cdot \tau}.$$

Although the summation ranges over L^+, the set of values $Q(M)$ acquired by Q on M is in fact contained in some fractional ideal in L. A fundamental fact is that θ_M is a **Hilbert modular form** in the sense we shall define below. It has weight $n/2$, and one has a precise description of the level and character; see [Wal1, Theorem 3.7] for the case $M = \Lambda$. Shimura [Shi4, Theorem A3.8] deals with a general case, but a more accessible reference might be [Gar, Theorem p.154] though the level it gives may not be optimal.

Example 6.4.1. Let L be the lattice $R \otimes \mathcal{O}_L$, where R is a maximal order in the quaternion algebra $B_{p,\infty}$. We discussed this above in Example 6.1.1 and on page 37. The quadratic form is simply the norm and its discriminant is $p^2 \mathcal{O}_L$. We choose an integer $N \geq 1$ and a vector $t \in R \otimes \mathcal{O}_L$. We let M be the lattice coset $t + NR \otimes \mathcal{O}_L$. If we define the ideal $\mathfrak{f} = 4pN\mathfrak{d}_L$, then $\mathfrak{f}$ satisfies the five conditions appearing in [Gar, p. 153]: (i) holds because $\mathfrak{f} \subseteq 4\mathfrak{d}_L$, (iii) amounts to $\mathfrak{f} \subseteq N\mathcal{O}_L$, (v) holds because $\mathfrak{f} \subseteq p\mathcal{O}_L$, and (ii) and (iv) hold vacuously. Then [Gar, Theorem, p. 154] gives that Θ_M is a Hilbert modular form of parallel weight 2 and level $\Gamma(\mathfrak{f}) = \{\gamma \in \mathrm{SL}_2(\mathcal{O}_L) : \gamma \equiv I_2 \pmod{\mathfrak{f}}\}$, a priori with a character, but the comments in loc. cit. following the Theorem show that this character is trivial in our example.

For a genus $\mathfrak{g} = \mathfrak{g}(M)$, we define the **theta function of the genus**,

$$(6.13) \qquad \theta_\mathfrak{g}(\tau) = \frac{1}{m(\mathfrak{g})} \sum_{i=1}^{h(\mathfrak{g})} \frac{1}{w_i} \theta_{M_i}(\tau)$$

where w_i is the order of the stabilizer of M_i in $O_Q(L)$, and the mass $m(\mathfrak{g})$ of $\mathfrak{g}$ is as defined in Eq. (6.10). Its Fourier coefficients are averaged representation numbers $\frac{1}{m(\mathfrak{g})} \sum_{i=1}^{h(\mathfrak{g})} \frac{1}{w_i} r_{M_i}(\mu)$, with each θ_{M_i} weighted by the measure of the double coset $O_Q(L)\gamma_i K_{\mathbb{A}}^M$ in $O_Q(L) \backslash O_Q(\mathbb{A})$. The key points that we will need are that:

(1) $\Theta_\mathfrak{g}$ is an *Eisenstein series* on $\mathfrak{H}^g$ of weight $n/2$ for a suitable discrete subgroup of $\mathrm{SL}_2(L)$;
(2) For every i, $\theta_{M_i} - \theta_\mathfrak{g}$ is a *cusp form*. That is, $\theta_\mathfrak{g}$ is the Eisenstein component of *every* θ_{M_i}.

So if we can show that the Fourier coefficients of the Eisenstein series grow faster than the corresponding coefficients of the cusp forms, we can use this to show that $r_M(\mu) > 0$ for certain values of μ. We will explain these points in §6.5.2 and §6.5.1 respectively, after we have given an adelic description of the theta functions involved.

6.4.2. *Adelic and classical modular functions.* Following [Gar, Gel, KL], we discuss the connection between classical Hilbert modular forms and adelic Hilbert automorphic forms. We restrict to the case of parallel weight $(w, \ldots, w)$ as this suffices for the applications in this paper, and we additionally assume that the totally real field L is not $\mathbb{Q}$. Recall that $\sigma_1, \ldots, \sigma_g$ denote the embeddings $L \to \mathbb{R}$.

On the classical side, define an **automorphy factor**

$$j(\alpha, \tau) := \prod_{i=1}^{g} (c_i \tau_i + d_i), \qquad \tau = (\tau_1, \ldots, \tau_g) \in \mathfrak{H}^g, \qquad \alpha = \left(\begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \right)_{i=1,\ldots,g} \in \mathrm{SL}_2(\mathbb{R})^g,$$

which satisfies

$$j(\alpha_1 \alpha_2, \tau) = j(\alpha_1, \alpha_2 \tau) \cdot j(\alpha_2, \tau).$$

Let $i_\infty \colon \mathrm{SL}_2(L) \hookrightarrow \mathrm{SL}_2(\mathbb{R})^g$ be the embedding induced by the g embeddings $\sigma_i \colon L \to \mathbb{R}$. For $\gamma \in \mathrm{SL}_2(L)$ we also write $j(\gamma, \tau)$ to denote $j(i_\infty(\gamma), \tau)$. Using this, we define the **slash operator** $|_w \gamma$ for $w \in \mathbb{Z}$ and $\gamma \in \mathrm{SL}_2(L)$ as an operator on continuous functions $f \colon \mathfrak{H}^g \to \mathbb{C}$ by

$$(f|_w \gamma)(\tau) := j(\gamma, \tau)^{-w} f(\gamma \tau).$$

This defines a right action of $\mathrm{SL}_2(L)$ on such functions, namely $((f|_w \gamma_1)|_w \gamma_2) = f|_w (\gamma_1 \gamma_2)$. Let Γ be a congruence subgroup of $\mathrm{SL}_2(L)$. We say a continuous function $f \colon \mathfrak{H}^g \to \mathbb{C}$ is **modular** of weight w and level Γ if $f|_w \gamma = f$ for all $\gamma \in \Gamma$. If f is modular of weight w and level Γ then $f|_w \gamma$ is modular of weight w and level $\gamma^{-1} \Gamma \gamma$. If in addition f is holomorphic then we call f a **Hilbert modular form** and then all $f|_w \gamma$ will be Hilbert modular forms as well.

On the adelic side, we consider complex-valued functions on $\mathrm{SL}_2(\mathbb{A})$. We may view $\mathrm{SL}_2(\mathbb{R})^g$ as a subgroup of $\mathrm{SL}_2(\mathbb{A})$ by sending $\alpha = (\alpha_1, \ldots, \alpha_g)$ to the element that is I_2 at each finite place and α_i at the infinite place corresponding to σ_i, an element we shall also denote $\epsilon_\infty(\alpha)$, if confusion may arise. Within this subgroup we have the compact subgroup $\mathrm{SO}_2(\mathbb{R})^g$, and we can write an element of this group as

$$k(\theta) = \left(\begin{pmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{pmatrix} \right)_{i=1,\ldots,g}, \qquad \theta = (\theta_1, \ldots, \theta_g) \in \mathbb{R}^g.$$

Let $\mathbb{K}$ be a compact open subgroup of $\mathrm{SL}_2(\mathbb{A}_f)$. We say $F \colon \mathrm{SL}_2(\mathbb{A}) \to \mathbb{C}$ is an **adelic automorphic function** of weight w and level $\mathbb{K}$ if it satisfies

$$F(\gamma \alpha k(\theta) r) = e^{iw(\theta_1 + \cdots + \theta_g)} F(\alpha)$$

for $\gamma \in \mathrm{SL}_2(L)$ (here embedded diagonally in $\mathrm{SL}_2(\mathbb{A})$), $\alpha \in \mathrm{SL}_2(\mathbb{R})^g$, $k(\theta) \in \mathrm{SO}_2(\mathbb{R})^g$, and $r \in \mathbb{K}$; in words, F is invariant under left multiplication by $\mathrm{SL}_2(L)$, invariant under right multiplication by $\mathbb{K}$, and equivariant under right multiplication by $\mathrm{SO}_2(\mathbb{R})^g$ with character $k(\theta) \mapsto e^{iw \sum \theta_i} = j(k(\theta), i)^{-w}$.

We now establish a one-to-one correspondence between modular functions $f \colon \mathfrak{H}^g \to \mathbb{C}$ and adelic automorphic functions. First, given a modular function f of weight w and level Γ, we describe how to elevate it to an adelic automorphic function

$$f^\sharp \colon \mathrm{SL}_2(\mathbb{A}) \to \mathbb{C}$$

of weight w and an appropriate choice of level $\mathbb{K}$. We call $f^\sharp$ the **adelization** of f.

We first define $f^\sharp$ on the subgroup $\mathrm{SL}_2(\mathbb{R})^g$ of $\mathrm{SL}_2(\mathbb{A})$ by

$$f^\sharp(\alpha) = j(\alpha, i)^{-w} f(\alpha(i)), \qquad \alpha \in \mathrm{SL}_2(\mathbb{R})^g.$$

The modular property of f translates into left-Γ-invariance of $f^\sharp$, and we have the relation

$$f^\sharp(\gamma \cdot \alpha \cdot k(\theta)) = e^{iw(\theta_1 + \cdots + \theta_g)} f^\sharp(\alpha), \qquad \gamma \in \Gamma, \quad \alpha \in \mathrm{SL}_2(\mathbb{R})^g, \quad k(\theta) \in \mathrm{SO}_2(\mathbb{R})^g.$$

A consequence of strong approximation for $\mathrm{SL}_2(L)$ [Gar, Proposition, p. 88] implies that the inclusion $\epsilon_\infty\colon \mathrm{SL}_2(\mathbb{R})^g \to \mathrm{SL}_2(\mathbb{A})$ induces a homeomorphism

$$\Gamma\backslash\mathrm{SL}_2(\mathbb{R})^g \xrightarrow{\cong} \mathrm{SL}_2(L)\backslash\mathrm{SL}_2(\mathbb{A})/\mathbb{K},$$
$$\Gamma\alpha \mapsto \mathrm{SL}_2(L)\alpha\mathbb{K},$$

(more pedantically, $\Gamma\alpha \mapsto \mathrm{SL}_2(L)\epsilon_\infty(\alpha)\mathbb{K}$) where $\mathbb{K} := \prod_{v<\infty}\Gamma_v$ for Γ_v the v-adic completion of Γ; this is equal to $\mathrm{SL}_2(\mathcal{O}_{L_v})$ at all but finitely many places v. Thus, any left-Γ-invariant function φ on $\mathrm{SL}_2(\mathbb{R})^g$ extends uniquely to a function on $\mathrm{SL}_2(\mathbb{A})$ that is left-$\mathrm{SL}_2(L)$-invariant and right-$\mathbb{K}$-invariant. Explicitly, we can write any $t \in \mathrm{SL}_2(\mathbb{A})$ as $t = \gamma\alpha k$ for some $\gamma \in \mathrm{SL}_2(L)$, $\alpha \in \mathrm{SL}_2(\mathbb{R})^g$, and $k \in \mathbb{K}$, and then the assignment $\varphi(t) := \varphi(\alpha)$ gives a well-defined extension of φ to all of $\mathrm{SL}_2(\mathbb{A})$. Thus we see that $f^\sharp$, a priori defined only on $\mathrm{SL}_2(\mathbb{R})^g$, extends uniquely to an adelic automorphic function $f^\sharp$ on $\mathrm{SL}_2(\mathbb{A})$.

Conversely, given an adelic automorphic function F of some level $\mathbb{K}$ and weight w, we produce a modular function

$$F^\flat\colon \mathfrak{H}^g \to \mathbb{C}$$

of level $\Gamma := \mathrm{SL}_2(L) \cap (\mathbb{K} \times \mathrm{SL}_2(\mathbb{R})^g)$ and weight w. We call $F^\flat$ the **unadelization** of F. To compute the value of $F^\flat$ at $\tau = x + iy = (x_1 + iy_1, \ldots, x_g + iy_g) \in \mathfrak{H}^g$, we define the element

$$(6.14) \qquad\qquad \alpha_\tau \in \mathrm{SL}_2(\mathbb{R})^g \subset \mathrm{SL}_2(\mathbb{A}),$$

which equals $\begin{pmatrix} y_i^{1/2} & x_i y_i^{-1/2} \\ 0 & y_i^{-1/2} \end{pmatrix}$ at the infinite place corresponding to the embedding σ_i, and is I_2 at every finite place. Note that $\alpha_{x+iy} = \alpha_{x+i}\alpha_{iy}$ and that $\alpha_\tau i = \tau$. Then we define

$$(6.15) \qquad\qquad F^\flat(\tau) := j(\alpha_\tau, i)^w F(\alpha_\tau) = \left(\prod_{i=1}^g y_i\right)^{-w/2} F(\alpha_\tau).$$

It is straightforward to check that $(F^\flat)^\sharp = F$: note that every element of $\mathrm{SL}_2(\mathbb{A})$ is contained in $\mathrm{SL}_2(L) \cdot \alpha_\tau \cdot \mathrm{SO}_2(\mathbb{R})^g \cdot \mathbb{K}$ for some $\tau \in \mathfrak{H}^g$, so one merely needs to check that the functions agree on all elements of the form α_τ and satisfy the same transformation properties with respect to $\mathrm{SL}_2(L)$, $\mathbb{K}$, and $\mathrm{SO}_2(\mathbb{R})^g$. Conversely, if f is a modular function, then $(f^\sharp)^\flat = f$. This establishes a one-to-one correspondence between modular continuous functions and adelic automorphic functions.

To characterize those F for which $F^\flat$ is holomorphic is a rather delicate issue and we refer the reader to [Gar]. For the particular choices of F we use, holomorphicity of $F^\flat$ will be evident from the Fourier expansions that we obtain.

We now address how adelization interacts with the slash operator. It is easy to check from the definitions that for $\gamma \in \mathrm{SL}_2(L) \xhookrightarrow{i_\infty} \mathrm{SL}_2(\mathbb{R})^g \subset \mathrm{SL}_2(\mathbb{A})$ one has

$$(6.16) \qquad\qquad (f|_w\gamma)^\sharp(\alpha) = f^\sharp(i_\infty(\gamma)\alpha), \qquad \gamma \in \mathrm{SL}_2(L),\ \alpha \in \mathrm{SL}_2(\mathbb{R})^g.$$

Indeed,

$$f^\sharp(i_\infty(\gamma)\alpha) = j(i_\infty(\gamma)\alpha, i)^{-w}f(\gamma\alpha(i)) = j(\alpha, i)^{-w}j(\gamma, \alpha(i))^{-w}f(\gamma\alpha(i))$$
$$= j(\alpha, i)^{-w}(f|_w\gamma)(\alpha(i)) = (f|_w\gamma)^\sharp(\alpha).$$

6.4.3. *Adelic and classical Fourier coefficients.* One of the key features of (holomorphic) Hilbert modular forms is that they have a Fourier expansion. If $\Gamma \subset \mathrm{SL}_2(L)$ is a congruence subgroup, then there exists a fractional $\mathcal{O}_L$-ideal $\mathfrak{a}$ in L such that

$$U_\mathfrak{a} := \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathfrak{a} \right\} \subset \Gamma.$$

Given any two such ideals $\mathfrak{a}, \mathfrak{b}$ we have $U_{\mathfrak{a}+\mathfrak{b}} = U_{\mathfrak{a}}U_{\mathfrak{b}} \subset \Gamma$, and therefore there exists one such $\mathfrak{a}$ that contains all the others; this fractional $\mathcal{O}_L$-ideal $\mathfrak{a}$ is called the **translation ideal** of the cusp $i\infty$ relative to Γ.

If f is a Hilbert modular form of level Γ and $\mathfrak{a}$ is the translation ideal of the cusp $i\infty$ relative to Γ, then f is invariant under $U_{\mathfrak{a}}$, and thus f has a classical Fourier expansion at the cusp $i\infty = (i\infty, \dots, i\infty)$,

$$f(z) = \sum_{\mu \in L} c_\mu(f) e^{2\pi i \operatorname{Tr} \mu \cdot z},$$

for uniquely determined complex numbers $c_\mu(f)$. These complex numbers, the **classical Fourier coefficients** of f at the cusp $i\infty$, are defined by

$$c_\mu(f) = \operatorname{vol}((L \otimes \mathbb{R})/\mathfrak{a})^{-1} \int_{x \in (L \otimes \mathbb{R})/\mathfrak{a}} f(x + iy) e^{-2\pi i \operatorname{Tr} \mu \cdot (x+iy)} \, d_\infty x \qquad \text{for } y \in \mathbb{R}^g_{>0};$$

when f is holomorphic this integral does not depend on y [Gar, §1.4]. Here $d_\infty x$ is the standard measure on $L \otimes \mathbb{R} \simeq \mathbb{R}^g$ such that $\operatorname{vol}(\mathbb{R}^g/\mathbb{Z}^g) = 1$, and the volume of $(L \otimes \mathbb{R})/\mathfrak{a}$ is computed with respect to this measure. Let $\mathfrak{a}^\vee$ the $\mathbb{Z}$-dual of $\mathfrak{a}$ relative to the trace form; that is, $\mathfrak{a}^\vee = (\mathfrak{d}_L \mathfrak{a})^{-1}$. Then we have $c_\mu(f) = 0$ for all $\mu \notin \mathfrak{a}^\vee$, and so the Fourier expansion could instead be written as a sum over $\mathfrak{a}^\vee$. In fact, as $L \neq \mathbb{Q}$, we have additionally that $c_\mu = 0$ unless $\mu = 0$ or $\sigma_i(\mu) > 0$ for all $i = 1, \dots, g$ (the **Koecher principle**).

The q-expansion of f at other cusps can be defined as the q-expansion of $f|_w \gamma$ at the cusp $i\infty$ for $\gamma \in \operatorname{SL}_2(L)$. Since $f|_w \gamma$ has level $\gamma^{-1} \Gamma \gamma$, these Fourier coefficients are computed using the translation ideal of the cusp $i\infty$ associated to $\gamma^{-1} \Gamma \gamma$. (This gives a q-expansion of f at the cusp $\gamma^{-1} i\infty$, but note that even if $\gamma_1^{-1} i\infty = \gamma_2^{-1} i\infty$, the q-expansions of $f|_w \gamma_1$ and of $f|_w \gamma_2$ may not agree, as the elements γ_1, γ_2 may parametrize the neighborhood of the cusp in different ways.) In particular, for every $\mu \in L$ we obtain a function

$$c_\mu(\cdot, f) \colon \operatorname{SL}_2(L) \to \mathbb{C},$$

defined for $\gamma \in \operatorname{SL}_2(L)$ as the coefficient $c_\mu(f|_w \gamma)$ in the Fourier expansion of $f|_w \gamma$ at the cusp $i\infty$. (Note in particular that $c_\mu(f) = c_\mu(I_2, f)$.) We say a Hilbert modular form f is a **cusp form** if $c_0(\gamma, f) = 0$ for all $\gamma \in \operatorname{SL}_2(L)$.

On the adelic side, we first define an **adelic exponential** function

$$e_{\mathbb{A}} \colon \mathbb{A} \to \mathbb{C}^\times,$$

which is a product of local exponential functions, $e_{\mathbb{A}}((x_v)_v) = \prod_v e_v(x_v)$. For v infinite and $x_v \in \mathbb{R}$ we let $e_v(x_v) = e^{2\pi i x_v}$. For v finite and $x_v \in L_v$, let p be the rational prime that v lies over, so that $\operatorname{Tr}_{L_v/\mathbb{Q}_p}(x_v) \in \mathbb{Q}_p$. The map $\mathbb{Z}[\frac{1}{p}] \to \mathbb{Q}_p$ induces an isomorphism $\mathbb{Z}[\frac{1}{p}]/\mathbb{Z} \to \mathbb{Q}_p/\mathbb{Z}_p$; thus there exists $z_v \in \mathbb{Z}[\frac{1}{p}]$, unique up to translation by $\mathbb{Z}$, satisfying $\operatorname{Tr}_{L_v/\mathbb{Q}_p}(x_v) \in z_v + \mathbb{Z}_p$. We define $e_v(x_v) = e^{-2\pi i z_v}$. The resulting adelic function $e_{\mathbb{A}}$ evaluates to 1 on L and so defines a nontrivial character on the compact abelian group $\mathbb{A}/L$. From its definition we can show that $e_{\mathbb{A}}(\widehat{\mathfrak{d}_L^{-1}}) = 1$, where for a fractional ideal $\mathfrak{a}$ of L,

$$\hat{\mathfrak{a}} = \prod_{v < \infty} \mathfrak{a} \otimes \mathcal{O}_{L,v}$$

denotes the adelization of $\mathfrak{a}$.

Let F be a function on $\operatorname{SL}_2(\mathbb{A})$ that is left-$\operatorname{SL}_2(L)$-invariant. For $\mu \in L$, define a function $\mathcal{W}_\mu \colon \operatorname{SL}_2(\mathbb{A}) \to \mathbb{C}$ – the μth **adelic Fourier coefficient** of F – by

$$\mathcal{W}_\mu(\alpha) = \int_{\mathbb{A}/L} F(u(x)\alpha) e_{\mathbb{A}}(\mu x)^{-1} d_{\mathbb{A}} x,$$

where $u(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ and the measure $d_{\mathbb{A}}x$ is normalized as to give $\mathbb{A}/L$ volume 1. Note that if F is right-$\mathbb{K}$-invariant then so is the function $\mathcal{W}_\mu$.

One has the following properties of adelic Fourier expansion ([Gar, p. 74]):

(1) We have $\mathcal{W}_\mu(u(x)\alpha) = e_{\mathbb{A}}(\mu x)\mathcal{W}_\mu(\alpha)$, $x \in \mathbb{A}$.
(2) As L^2 functions, $F(\alpha) = \sum_{\mu \in L} \mathcal{W}_\mu(\alpha)$.
(3) For $M = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathrm{SL}_2(L)$ one has $\mathcal{W}_\mu(\alpha) = \mathcal{W}_{ad^{-1}\mu}(M\alpha)$.

The function $\mathcal{W}_0 : \mathrm{SL}_2(\mathbb{A}) \to \mathbb{C}$ is called the **constant term** of F. An adelic automorphic function F is called **cuspidal** if its constant term $\mathcal{W}_0$ is 0 almost everywhere.

Given a classical holomorphic Hilbert modular form f, we compare the adelic Fourier coefficients of $F = f^\sharp$, that are functions $\mathcal{W}_\mu(\cdot, f^\sharp) \colon \mathrm{SL}_2(\mathbb{A}) \to \mathbb{C}$, with the Fourier coefficients of f that are functions $c_\mu(\cdot, f) \colon \mathrm{SL}_2(L) \to \mathbb{C}$. Let $\mathbb{K}$ be the level of $f^\sharp$. Given $\gamma \in \mathrm{SL}_2(L)$, we write $\gamma_\mathfrak{f} \in \mathrm{SL}_2(\mathbb{A})$ for the element that equals γ at all finite places and I_2 at all infinite places (that is, the image of γ under the embedding of $\mathrm{SL}_2(L)$ into $\mathrm{SL}_2(\mathbb{A}_\mathfrak{f})$, where $\mathrm{SL}_2(\mathbb{A}_\mathfrak{f})$ is viewed as a subgroup of $\mathrm{SL}_2(\mathbb{A})$).

Proposition 6.4.2. *Let f be a Hilbert modular form. Let $\gamma \in \mathrm{SL}_2(L)$, $\alpha \in \mathrm{SL}_2(\mathbb{R})^g$, and $r \in \mathbb{K}$. Then for all $\mu \in L$ we have*

$$\mathcal{W}_\mu(\gamma_\mathfrak{f}^{-1}\alpha r, f^\sharp) = j(\alpha, i)^{-w} e^{2\pi i \, \mathrm{Tr}\, \mu \cdot \alpha(i)} c_\mu(\gamma, f).$$

Before the proof, we mention the special case

(6.17) $$c_\mu(I_2, f) = e^{2\pi \, \mathrm{Tr}\, \mu} \mathcal{W}_\mu(I_2, f^\sharp),$$

as well as the following corollary. Recall that by strong approximation, every element of $\mathrm{SL}_2(\mathbb{A})$ can be written in the form $\gamma_\mathfrak{f}^{-1}\alpha r$ as in the statement of the proposition, meaning that every adelic Fourier coefficient of $f^\sharp$ can be written in terms of some classical Fourier coefficient. We thus obtain the following.

Corollary 6.4.3. *A classical Hilbert modular form f is a cusp form if and only if its adelization $f^\sharp$ is a cuspidal automorphic function.*

Proof of Proposition 6.4.2. We follow [KL, Proposition 12.3], which proves this result for $L = \mathbb{Q}$. Let Γ be the level of f, so that $\gamma^{-1}\Gamma\gamma$ is the level of $f|_w\gamma$, and let $\mathfrak{a}$ be the translation ideal of the cusp $i\infty$ relative to $\gamma^{-1}\Gamma\gamma$. Then $U_\mathfrak{a}$ is contained $\gamma^{-1}\Gamma\gamma$ and so $\gamma U_\mathfrak{a}\gamma^{-1} \subset \Gamma$. Taking the product of v-adic completions for all finite v we obtain $\gamma_\mathfrak{f} U_{\hat{\mathfrak{a}}}\gamma_\mathfrak{f}^{-1} \subset \mathbb{K}$. So for any $x \in \mathbb{A}$ and $\xi \in \hat{\mathfrak{a}}$ we have

$$f^\sharp(u(x)u(\xi)\gamma_\mathfrak{f}^{-1}\alpha) = f^\sharp(u(x)\gamma_\mathfrak{f}^{-1}\gamma_\mathfrak{f}u(\xi)\gamma_\mathfrak{f}^{-1}\alpha) = f^\sharp(u(x)\gamma_\mathfrak{f}^{-1}\alpha),$$

since $\gamma_\mathfrak{f}u(\xi)\gamma_\mathfrak{f}^{-1} \in \mathbb{K}$ commutes with $\alpha \in \mathrm{SL}_2(\mathbb{R})^g$ and $f^\sharp$ is right-$\mathbb{K}$-invariant.

By strong approximation we have $\mathbb{A} = L((L \otimes \mathbb{R}) \times \hat{\mathfrak{a}})$, and so $\mathbb{A}/L \simeq ((L \otimes \mathbb{R}) \times \hat{\mathfrak{a}})/\mathfrak{a}$. From this we obtain an isomorphism of measure spaces (but not of groups),

$$\mathbb{A}/L \simeq ((L \otimes \mathbb{R})/\mathfrak{a}) \times \hat{\mathfrak{a}};$$

see [KL, Theorem 7.40]. Thus if we let $d_\mathfrak{f}\xi$ denotes the Haar measure on $\mathbb{A}_\mathfrak{f}$ with the property that $d_\infty \times d_\mathfrak{f} = d_{\mathbb{A}}$, we can write

$$\mathcal{W}_\mu(\gamma_\mathfrak{f}^{-1}\alpha r) = \int_{\mathbb{A}/L} f^\sharp(u(x)\gamma_\mathfrak{f}^{-1}\alpha r)e_{\mathbb{A}}(\mu x)^{-1} \, d_{\mathbb{A}}x$$

$$= \int_{(L \otimes \mathbb{R})/\mathfrak{a}} \int_{\hat{\mathfrak{a}}} f^\sharp(u(x)u(\xi)\gamma_\mathfrak{f}^{-1}\alpha)e_{\mathbb{A}}(\mu\xi)^{-1}e_{\mathbb{A}}(\mu x)^{-1} \, d_\mathfrak{f}\xi \, d_\infty x$$

$$= \int_{(L \otimes \mathbb{R})/\mathfrak{a}} f^\sharp(u(x)\gamma_\mathfrak{f}^{-1}\alpha)e_{\mathbb{A}}(\mu x)^{-1} \, d_\infty x \cdot \int_{\hat{\mathfrak{a}}} e_{\mathbb{A}}(\mu\xi)^{-1} \, d_\mathfrak{f}\xi.$$

If $\mu \notin \mathfrak{a}^\vee$ then the integral over $\hat{\mathfrak{a}}$ is 0; recalling that we have $c_\mu(\gamma, f) = 0$ for $\mu \notin \mathfrak{a}^\vee$, we obtain the desired identity $0 = 0$. So we now consider the case $\mu \in \mathfrak{a}^\vee$. In this case we have $\mu \xi \in \widehat{\mathfrak{d}_L}^{-1}$ for all $\xi \in \hat{\mathfrak{a}}$, so that $e_{\mathbb{A}}(\mu \xi) = 1$. We can conclude that

$$\mathcal{W}_\mu(\gamma_{\mathfrak{f}}^{-1} \alpha r) = \mathrm{vol}(\hat{\mathfrak{a}}) \int_{(L \otimes \mathbb{R})/\mathfrak{a}} f^\sharp(\gamma_{\mathfrak{f}}^{-1} u(x)\alpha) e_{\mathbb{A}}(\mu x)^{-1} \, d_\infty x.$$

Here we also used the fact that $\gamma_{\mathfrak{f}}$ and $u(x)$ commute as they are supported on disjoint sets of places. Writing $z := \alpha(i)$ we have by definition of adelization

$$\begin{aligned}
f^\sharp(\gamma_{\mathfrak{f}}^{-1} u(x)\alpha) &= f^\sharp(\gamma^{-1} i_\infty(\gamma) u(x)\alpha) \\
&= j(i_\infty(\gamma) u(x)\alpha, i)^{-w} f(i_\infty(\gamma) u(x) z) \\
&= j(\alpha, i)^{-w} j(u(x), z)^{-w} j(i_\infty(\gamma), z + x)^{-w} f(i_\infty(\gamma)(z + x)) \\
&= j(\alpha, i)^{-w} (f|_w \gamma)(z + x).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathcal{W}_\mu(\gamma_{\mathfrak{f}}^{-1} \alpha r) &= j(\alpha, i)^{-w} \mathrm{vol}(\hat{\mathfrak{a}}) \int_{(L \otimes \mathbb{R})/\mathfrak{a}} (f|_w \gamma)(z + x) e^{-2\pi i \, \mathrm{Tr} \, \mu \cdot x} \, d_\infty x \\
&= j(\alpha, i)^{-w} e^{2\pi i \, \mathrm{Tr} \, \mu \cdot z} \mathrm{vol}(\hat{\mathfrak{a}}) \int_{(L \otimes \mathbb{R})/\mathfrak{a}} (f|_w \gamma)(z + x) e^{-2\pi i \, \mathrm{Tr} \, \mu \cdot (z+x)} \, d_\infty x \\
&= j(\alpha, i)^{-w} e^{2\pi i \, \mathrm{Tr} \, \mu \cdot z} \mathrm{vol}(\hat{\mathfrak{a}}) \mathrm{vol}((L \otimes \mathbb{R})/\mathfrak{a}) c_\mu(\gamma, f).
\end{aligned}$$

Finally note that

$$\mathrm{vol}(\hat{\mathfrak{a}}) \mathrm{vol}((L \otimes \mathbb{R})/\mathfrak{a}) = \int_{(L \otimes \mathbb{R})/\mathfrak{a}} \int_{\hat{\mathfrak{a}}} d_{\mathfrak{f}} \xi \, d_\infty x = \int_{\mathbb{A}/L} d_{\mathbb{A}} = 1.$$

$\square$

We remark that for a complete theory, and in particular for introducing Hecke operators into the picture, it is necessary to work with $\mathrm{GL}_2(\mathbb{A})^+$ and that introduces complications; see [Gar]. For our applications this will not be necessary. We also remark that working with non-parallel weight does not introduce substantial difficulties, but we avoid that as we will have no need for that and loc. cit. restricts to parallel weight.

6.4.4. *The Deligne bound.* To obtain an asymptotic on the growth of Fourier coefficients, we will need to have an upper bound on the growth of Fourier coefficients of a cuspidal Hilbert modular form of weight $w \geq 0$. Recall that for $L = \mathbb{Q}$, and a cusp form $f = \sum_{n \in \mathbb{Q}} a_n q^n$, $q = e^{2\pi i n z}$ on some congruence subgroup of $\mathrm{SL}_2(\mathbb{Q})$, one has the **trivial bound (Hecke bound)**

$$|a_n| \leq \Omega \cdot n^{w/2},$$

where Ω is a constant depending only on Γ and w. For a general modular form (not necessarily a cusp form) one has

$$|a_n| \leq \Omega \cdot n^{w-1}$$

for some possibly larger constant Ω. Note that for modular forms of weight 2 the two bounds coincide. However, for cusp forms one has the much deeper **Deligne bound**, for every $\epsilon > 0$

$$|a_n| \leq O(n^{(w-1)/2+\epsilon}).$$

Similary, for Hilbert modular cusp forms $f(z) = \sum_{\mu \in L} c_\mu e^{2\pi i \, \mathrm{Tr} \, \mu \cdot z}$ of weight w and level given by a congruence subgroup Γ, we have the trivial bound [Gar, p. 24]

$$|c_\mu| \leq \Omega \cdot N_{L/\mathbb{Q}}(\mu)^{w/2}.$$

However, we will need the stronger Deligne bound. Unfortunately, this bound is proven [Bla, BrL, Liv] (see also [Sha]) in the language of automorphic representations and so some translation of concepts is needed. Due to limitation of space, we will omit this[6] and use the following bound.

Theorem 6.4.4. *For a Hilbert modular cusp form $f(z) = \sum_{\mu \in L} c_\mu e^{2\pi i \operatorname{Tr} \mu \cdot z}$ of weight w and level Γ, a congruence subgroup, for every $\epsilon > 0$, we have the bound*

$$|c_\mu| \le \Omega \cdot N_{L/\mathbb{Q}}(\mu)^{(w-1)/2+\epsilon}.$$

Remark 6.4.5. In fact, we will only use this bound for $w \ge 2$ and it would suffice for us that $\epsilon < 1/2$. In particular, the case $\epsilon = 1/5$ proven by Shahidi will suffice.

6.4.5. *The theta correspondence.* In this section and the following, we discuss a recipe for turning functions on the adelic group $O_Q(\mathbb{A})$ into classical Hilbert modular forms. We will then identify those functions on $O_Q(\mathbb{A})$ that give rise to inhomogeneous theta series, Eisenstein series, and cusp forms.

The first ingredient in this recipe is the Weil representation, a special case of which we consider here. We write as before $\mathbb{A}$ for the adeles of L. Let $\mathscr{S}(V_\mathbb{A})$ denote the space of Schwartz functions on the group $V_\mathbb{A}$.[7] The orthogonal group $O(\mathbb{A})$ acts on $\mathscr{S}(V_\mathbb{A})$ by change of variable, $(\beta \cdot \phi)(x) = \phi(\beta^{-1}x)$.

The **Weil representation** of $\operatorname{SL}_2(\mathbb{A})$ is a representation acting on $\mathscr{S}(V_\mathbb{A})$,

$$\rho \colon \operatorname{SL}_2(\mathbb{A}) \to \operatorname{GL}(\mathscr{S}(V_\mathbb{A})).$$

The construction of the Weil representation ρ is involved, but once granted its existence, it is uniquely determined by the specifying it on matrices that generate $\operatorname{SL}_2(\mathbb{A})$:

(6.18) $\left(\rho\left(\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)\right)\phi\right)(v) = (a, (-1)^{n/2}\det(Q))_L |a|_\mathbb{A}^{n/2}\phi(av),$

(6.19) $\left(\rho\left(\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)\right)\phi\right)(v) = e_\mathbb{A}(xQ(v))\phi(v),$

              $\left(\rho\left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\right)\phi\right)(v) = \hat{\phi}(-v),$

where $(\cdot, \cdot)_L$ denotes the Hilbert symbol (the product of local Hilbert symbols $(\cdot, \cdot)_{L_v}$), $|a|_\mathbb{A}$ is the adelic absolute value (product of all local absolute values), $e_\mathbb{A}$ is the adelic exponential defined in §6.4.3, and $\hat{\phi}$ denotes an adelic Fourier transform (we will not need its explicit description in this article; see [Han2] for details). The Weil representation extends multiplicatively to a representation of $\operatorname{SL}_2(\mathbb{A}) \times O_Q(\mathbb{A})$ on $\mathscr{S}(V_\mathbb{A})$, where $\beta \in O_Q(\mathbb{A})$ acts by change of variable $(\rho(\beta)\phi)(x) = \phi(\beta^{-1}x)$ as above; a crucial fact we make use of is that the actions of $\operatorname{SL}_2(\mathbb{A})$ and $O_Q(\mathbb{A})$ commute. See [Han2, p. 156][8], [Pra], and [Gar] for a more complete treatment.

Using the Weil representation we can define the adelic **theta kernel**, which is a function

$$\theta \colon \mathscr{S}(V_\mathbb{A}) \times \operatorname{SL}_2(\mathbb{A}) \times O_Q(\mathbb{A}) \to \mathbb{C},$$

taking a triple (ϕ, α, β) to

$$\theta_\phi(\alpha, \beta) = \sum_{s \in V} (\rho(\alpha, \beta)\phi)(s) = \sum_{s \in V} (\rho(\alpha)\phi)(\beta^{-1}s).$$

---

[6]In a nutshell, one deduces the general case from the case of cuspidal eigenforms, and those are treated through the associated L-series. The associated L series have an Euler product and the control over the factors comes from constructing the associated global Galois representation in a way that provides information on the eigenvalues of Frobenii; typically, one finds the representations in this case in cohomology of algebraic varieties.

[7]We will also introduce Schwartz functions on $\operatorname{SL}_2(\mathbb{A})$ and $O_Q(\mathbb{A})$. We do not give full definitions of these functions; for our purposes, it is enough that the reader trusts that the Schwartz functions form a space of functions on which all the manipulations we do are well defined. See [Bum, §3.1] for details.

[8]Note a minor inaccuracy in the reference: ρ is not defined modulo $\operatorname{Sp}_2(F)$ as is written there, but the action of $\operatorname{Sp}_2(F)$ on the theta kernel introduced below is indeed trivial.

An important fact about the theta kernel is that it is invariant under left multiplication by $\mathrm{SL}_2(L)$: if $\gamma \in \mathrm{SL}_2(L)$, then $\theta_\phi(\gamma\alpha, \beta) = \theta_\phi(\alpha, \beta)$. Using the theta kernel one defines the **theta correspondence** Θ that takes a Schwartz function Ψ on $O_Q(\mathbb{A})$ to a Schwartz function $\Theta(\Psi)$ on $\mathrm{SL}_2(\mathbb{A})$ defined by

$$\Theta(\Psi)(\alpha) := \int_{O_Q(L)\backslash O_Q(\mathbb{A})} \Psi(\beta)\theta_\phi(\alpha, \beta)d\beta, \qquad \alpha \in \mathrm{SL}_2(\mathbb{A}),$$

where $d\beta$ is a Haar measure on $O_Q(\mathbb{A})$.

6.5. **Theta functions via the theta correspondence.** Fix a lattice coset $M = \Lambda + t$ in V. Let

$$\phi := \prod_v \phi_v$$

be the Schwartz function on $V_\mathbb{A}$ such that ϕ_v is the characteristic function of M_v at each finite place v, and $\phi_\sigma(x) = e^{-2\pi\sigma(Q(x))}$ at infinite places $\sigma \colon L \to \mathbb{R}$. Note that for $\beta \in O_Q(\mathbb{A})$ we have $\beta \cdot \phi_M = \phi_{\beta M}$. In addition, we fix the Haar measure $d\beta$ on $O_Q(\mathbb{A})$ so that the adelic stabilizer $K_\mathbb{A}^M$ of M has measure 1.

Recall that $\gamma_1 = 1, \gamma_2, \ldots, \gamma_{v(\mathfrak{g})} \in \mathrm{SL}_2(\mathbb{A})$ are elements supported at the finite places such that $M_i := \gamma_i M$ form a set of representatives for the genus of M. For each $1 \leq i \leq h(\mathfrak{g})$, let Ψ_i be the function on $O_Q(\mathbb{A})$ such that $\Psi_i(\beta) = 1$ if $\beta M \simeq M_i$ and $\Psi_i(\beta) = 0$ otherwise; thus Ψ_i is the characteristic function of the double coset $O_Q(L)\gamma_i K_\mathbb{A}^M$ in $O_Q(\mathbb{A})$ as in Equation (6.9). Consequently, the sum of all the Ψ_i is the constant function $\mathbb{1}$ on $O_Q(\mathbb{A})$.

Lemma 6.5.1. *We have* $\Theta(\Psi_i)^\flat = \frac{1}{w_i}\theta_{M_i}$ *and* $\Theta(\mathbb{1})^\flat = m(\mathfrak{g})\theta_\mathfrak{g}$.

(Here θ_{M_i} and $\theta_\mathfrak{g}$ are as defined in Eqs. (6.11) and (6.13), respectively.)

Proof. We follow [Han2, Section 4.5], which performs this computation in the case M is a lattice. We decompose $O_Q(L)\backslash O_Q(\mathbb{A})$ into double cosets via Equation (6.9). If $\beta \in O_Q(L)\gamma_j K_\mathbb{A}^M$, then we have $\Psi_i(\beta) = 1$ if $i = j$ and $\Psi_i(\beta) = 0$ otherwise, so

$$\Theta(\Psi_i)(\alpha) = \int_{O_Q(L)\backslash O_Q(\mathbb{A})} \Psi_i(\beta)\theta_\phi(\alpha, \beta)\, d\beta$$

$$= \int_{O_Q(L)\backslash(O_Q(L)\gamma_i K_\mathbb{A}^M)} \theta_\phi(\alpha, \beta)\, d\beta.$$

Recalling that the map from $\gamma_i K_\mathbb{A}^M$ to its image in $O_Q(L)\backslash O_Q(\mathbb{A})$ is w_i-to-one,

$$\Theta(\Psi_i)(\alpha) = \frac{1}{w_i} \int_{\gamma_i K_\mathbb{A}^M} \theta_\phi(\alpha, \beta)\, d\beta$$

$$= \frac{1}{w_i} \int_{K_\mathbb{A}^M} \theta_\phi(\alpha, \gamma_i \beta)\, d\beta$$

$$= \frac{1}{w_i} \int_{K_\mathbb{A}^M} \sum_{s \in V} (\rho(\alpha)\phi)(\beta^{-1}\gamma_i^{-1}s)\, d\beta.$$

Taking $\tau \in \mathfrak{H}^g$, we evaluate $\Theta(\Psi_i)$ at the element $\alpha_\tau \in \mathrm{SL}_2(\mathbb{A})$ as defined in §6.4.2: this is $\begin{pmatrix} 1 & x_i \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_i^{1/2} & 0 \\ 0 & y_i^{-1/2} \end{pmatrix}$ at the infinite place σ_i, and the identity at all finite places. We can explicitly compute the action of $\rho(\alpha_\tau)$ on ϕ using Equations (6.18) and (6.19). Note in particular that the adelic Hilbert symbol, absolute value, and exponential are each a product over all places, as is ϕ, so we can compute the action at each place independently. At all finite places v we have $(\rho(\alpha_\tau)\phi)_v = \phi_v$, and since $\beta \in K_\mathbb{A}^M$ we find that β_v stabilizes M_v. Hence $\phi_v(\beta_v^{-1}\gamma_{i,v}^{-1}s)$ is 1 if

$s \in \gamma_{i,v} M_v$ and 0 otherwise. We conclude that

$$\prod_{v<\infty}((\rho(\alpha_\tau)\phi)(\beta^{-1}\gamma_i^{-1}s))_v = \begin{cases} 1, & s \in \gamma_i M = M_i, \\ 0 & \text{otherwise.} \end{cases}$$

Now consider the infinite place v corresponding to the embedding σ_i. Recall that we chose γ_i to be the identity at infinite places, and since $y_i^{1/2} > 0$ the local Hilbert symbol evaluates to 1. We therefore have

$$((\rho(\alpha_\tau)\phi)(\beta^{-1}\gamma_i^{-1}s))_v = y_i^{n/4}e^{2\pi i x_i \sigma_i(Q(\beta_v^{-1}s))}\phi_v(y_i^{1/2}\beta_v^{-1}s)$$
$$= y_i^{n/4}e^{2\pi i(x_i+iy_i)\sigma_i(Q(s))},$$

using the fact that $\beta_v \in O_Q(\mathbb{R})$. We can conclude that

$$\Theta(\Psi_i)^\flat(\tau) = j(\alpha_\tau, i)^{n/2}\Theta(\Psi_i)(\alpha_\tau)$$
$$= \left(\prod_{i=1}^g y_i^{-1/2}\right)^{n/2}\left(\frac{1}{w_i}\int_{K_\mathbb{A}^M}\sum_{s\in M_i}\left(\prod_{i=1}^g y_i^{n/4}e^{2\pi i(x_i+iy_i)\sigma_i(Q(s))}\right)d\beta\right)$$
$$= \frac{1}{w_i}\sum_{s\in M_i}e^{2\pi i\sum_{i=1}^g(x_i+iy_i)\sigma_i(Q(s))}\left(\int_{K_\mathbb{A}^M}d\beta\right)$$
$$= \frac{1}{w_i}\theta_{M_i}(\tau).$$

The result for $\Theta(\mathbb{1})^\flat$ follows from the fact that $\mathbb{1}$ is a sum over all Ψ_i and the theta correspondence and unadelization are both linear.   □

6.5.1. *The cuspidal part.* Recall that the integral of Ψ_i on $O_Q(L)\backslash O_Q(\mathbb{A})$ is $\frac{1}{w_i}$, and the integral of $\mathbb{1}$ on $O_Q(L)\backslash O_Q(\mathbb{A})$ is $m(\mathfrak{g})$. Thus we can write

$$\Psi_i = \frac{1}{w_i m(\mathfrak{g})}\mathbb{1} + \Psi_i^{(\text{cusp})}$$

for some function $\Psi_i^{(\text{cusp})}$ with integral 0 on $O_Q(L)\backslash O_Q(\mathbb{A})$. We saw above what Ψ_i and $\mathbb{1}$ map to under the theta correspondence and unadelization, so we now consider the suggestively named $\Psi_i^{(\text{cusp})}$.

Lemma 6.5.2. *If Ψ has integral 0 on $O_Q(L)\backslash O_Q(\mathbb{A})$ then $\Theta(\Psi)^\flat$ is a cusp form.*

Proof. By Corollary 6.4.3, it suffices to prove that $\Theta(\Psi)$ is a cusp form on $SL_2(\mathbb{A})$. We compute

$$\mathcal{W}_0(\alpha) = \int_{\mathbb{A}/L}\Theta(\Psi)(u(x)\alpha)\,dx$$
$$= \int_{\mathbb{A}/L}\int_{O_Q(L)\backslash O_Q(\mathbb{A})}\Psi(\beta)\left(\sum_{v\in V}(\rho(u(x)\alpha)\phi)(\beta^{-1}s)\right)d\beta\,dx.$$

Since $\mathbb{A}/L$ and $O_Q(L)\backslash O_Q(\mathbb{A})$ are both compact and ϕ is a Schwartz function, we can interchange the integrals and sum:

$$\mathcal{W}_0(\alpha) = \int_{O_Q(L)\backslash O_Q(\mathbb{A})}\Psi(\beta)\sum_{s\in V}\left(\int_{\mathbb{A}/L}(\rho(u(x)\alpha)\phi)(\beta^{-1}s)\,dx\right)d\beta.$$

We now examine the inner integral with α, β, s fixed. Using the fact that ρ is a representation, we can apply Equation (6.19) to obtain

$$\int_{\mathbb{A}/L} (\rho(u(x)\alpha)\phi)(\beta^{-1}s)\, dx = \int_{\mathbb{A}/L} (\rho(u(x)) \cdot (\rho(\alpha)\phi))(\beta^{-1}s)\, dx$$

$$= (\rho(\alpha)\phi)(\beta^{-1}s) \int_{\mathbb{A}/L} e_{\mathbb{A}}(xQ(s))\, dx$$

(Recall that $\beta \in O_Q(\mathbb{A})$ so $Q(\beta_v s) = Q(s)$ for all places v). For $s \neq 0$, we have $Q(s) \in L^\times$, and so $x \mapsto e_{\mathbb{A}}(xQ(s))$ is a nontrivial character on $\mathbb{A}/L$; thus this integral equals 0. Thus only the $s = 0$ summand survives, and we have

$$\mathcal{W}_0(\alpha) = \int_{O_Q(L)\backslash O_Q(\mathbb{A})} \Psi(\beta)(\rho(\alpha)\phi)(0)\, d\beta.$$

Now $(\rho(\alpha)\phi)(0)$ is constant in β, so $\mathcal{W}_0(\alpha)$ is some constant times the integral of $\Psi(\beta)$ over $O_Q(L)\backslash O_Q(\mathbb{A})$, which is 0 by assumption. Therefore $\Theta(\Psi)$ is a cusp form on $\mathrm{SL}_2(\mathbb{A})$. $\qquad\square$

Remark 6.5.3. In [Ral] Rallis gives a criterion under which the theta correspondence takes cusp forms to cusp forms in a much more general setting.

Combining Lemmas 6.5.1 and 6.5.2 we obtain the following (proven previoulsy in [Shi5, Lemma 2.2(4)]; see also [Wal2] for the case $M = \Lambda$).

Proposition 6.5.4. *For all $1 \leq i \leq h(\mathfrak{g})$, $\theta_{M_i} - \theta_{\mathfrak{g}}$ is a cusp form.*

We can therefore bound the Fourier coefficients of $\theta_{M_i} - \theta_{\mathfrak{g}}$ using Theorem 6.4.4. If we can prove a sufficiently large *lower bound* for the Fourier coefficients of $\theta_{\mathfrak{g}}$, we will obtain a *nontrivial lower bound* for the Fourier coefficients of θ_{M_i} for all i.

6.5.2. *The Eisenstein part.* The key idea of this section is that the Fourier coefficients of the theta function of a genus, $\theta_{\mathfrak{g}}$, can be written as a product of local densities. A version of this claim is expressed in [Han1, §5] in the case that M is a lattice. A product formula of this kind can be done by combining the **Siegel-Weil formula** — a special case of which says that the theta function of a genus (6.13) is equal to an appropriately defined Eisenstein series — with a **product formula** for the Fourier coefficients of an Eisenstein series. Both the Siegel-Weil formula and the product formula are originally due to Siegel [Sie] but we shall use more general versions of these results below. These steps were carried out by Shimura in order to get an explicit formula for the Fourier coefficients of the theta function of a lattice coset genus, [Shi5, Theorem 1.5]. For convenience and ease of exposition we will want to use a slightly different form for the coefficients.

Let Λ be an *integral $\mathcal{O}_L$-lattice*. Breaking with the notation of the previous sections, we will assume throughout this section that M is a lattice coset *contained in Λ*; that is, for some sublattice $\Lambda' \subset \Lambda$ and some $t \in \Lambda$, we have $M = t + \Lambda'$.

Fix an isomorphism $V \simeq L^n$. For each finite place v of L, let $\ell_v : V_v \to \mathbb{C}$ denote the characteristic function of the lattice coset M_v. Using this setup, we can define an **Eisenstein series** $E_M(x, s)$ for $x \in \mathbb{A}$ and $s \in \mathbb{C}$ as in [Shi3, (2.3.4)]. This Eisenstein series has a Fourier expansion

(6.20) $$E_M(x, s) = \sum_{\mu \in L} c_\mu(E; s) e_{\mathbb{A}}(\mu x),$$

and by the Siegel–Weil formula, we have

$$E_M(x, 0) = \Theta(\mathbb{1})(u(x))$$

for all $n \geq 4$. When M is a lattice this is identical to [Shi3, (5.3.1)]. For $n \geq 5$ this uses the version of the Siegel–Weil formula proven by Weil [Wei, Théorème 5]. For $n = 4$ this uses the version of the Siegel–Weil formula proven by Kudla and Rallis [KR], though it does not follow immediately as they use a different definition of Eisenstein series from Shimura; see [Shi3, §5.3]

for how to deduce this identity from [KR] (the argument is given for lattices but holds equally well for lattice cosets). We can therefore compute the Fourier coefficients of $\Theta(\mathbb{1})$ using (6.20):

$$\mathcal{W}_\mu(I_2, \Theta(\mathbb{1})) = \int_{\mathbb{A}/L} \left(\sum_{\mu \in L} c_\mu(E_M; 0) e_\mathbb{A}(\mu x) \right) e_\mathbb{A}(\mu x)^{-1} d_\mathbb{A} x = c_\mu(E_M; 0),$$

so by Lemma 6.5.1 and (6.17),

(6.21)
$$c_\mu(\theta_{\mathfrak{g}(M)}) = \frac{e^{2\pi \operatorname{Tr} \mu}}{m(\mathfrak{g}(M))} c_\mu(E_M; 0).$$

The Fourier coefficients $c_\mu(E_M; s)$ of $E_M(x, s)$ can each be expressed as an infinite product of local factors

(6.22)
$$c_\mu(E_M; s) = d_L^{1/2} \prod_v c_{\mu,v}(E_M; s)$$

as in [Shi3, p. 71], where d_L denotes the discriminant of $L/\mathbb{Q}$ as above.

For v a finite place of L, $c_{\mu,v}(E_M; s)$ is a **local density**, measuring the v-adic volume of the level set $Q^{-1}(\mu)$ inside the lattice coset M_v. To be more precise, let $\mathfrak{p}$ denote the maximal ideal of $\mathcal{O}_{L,v}$, and $q := [\mathcal{O}_{L,v} : \mathfrak{p}]$ the residue degree. For $\mu \in L$ and $k \geq 0$, set

(6.23)
$$A_{\mu,k}(M_v) := \{x \in M_v/\mathfrak{p}^k \Lambda_v : Q(x) \equiv \mu \pmod{\mathfrak{p}^k}\}.$$

Recall that the Eisenstein series E_M was defined using a choice of isomorphism $V \simeq L^n$; this induces a measure on $V_v \simeq L_v^n$ normalized so that $\mathcal{O}_{L,v}^n$ has measure 1.

Lemma 6.5.5. *For all sufficiently large k, $c_{\mu,v}(E_M; 0) = \tau_v \cdot \sharp A_{\mu,k}(M)/q^{k(n-1)}$, where τ_v is the volume of Λ_v according to the measure defined above.*

Proof. This is proven in the case $M = \Lambda$ in [Shi3, Lemma 2.5]; we will merely point out the changes needed for our context. The fact that $\sharp A_{\mu,k}(M)/q^{k(n-1)}$ stabilizes for sufficiently large k follows from the proof of [Shi2, Proposition 14.3]: the proposition is stated for lattices, but the proof proceeds by working modulo a sufficiently large power of $\mathfrak{p}$ and considering one congruence class at a time (it is essentially an argument by Hensel lifting), and therefore holds even if we impose congruence conditions such as those defining the coset M_v. Let d denote the limiting value of $\sharp A_{\mu,k}(M)/q^{k(n-1)}$ as $k \to \infty$.

As in [Shi3] we can compute

$$(1 - q^{-s})^{-1} |\beta|_v^{-s} c_{\mu,v}(E_M; s) = \sum_{k=0}^\infty q^{-ks} \int_{M_v} \int_{\mathfrak{p}^{-k-e}} e_v(\delta_v^{-1} \sigma(Q(x) - \mu)) \, d\sigma \, dx$$

where $\beta \in L_v^\times$ is a parameter used to define the Eisenstein series [Shi3, (2.2.3)], δ_v is a generator of the different ideal of $L_v/\mathbb{Q}_p$, e is the v-adic valuation of β/δ, e_v is the v-adic exponential defined in §6.4.3, and dx is the Haar measure on $V_v \simeq L_v^n$ under which $\mathcal{O}_{L,v}^n$ has measure 1. Observe that $e_v(\delta_v^{-1} x)$ is identically 1 on $\mathcal{O}_{L,v}$ but is a nontrivial character on $\mathfrak{p}^{-\ell}$ for any $\ell > 0$, so the inner integral over $\mathfrak{p}^{-k-e}$ is q^{k+e} if $Q(x) \equiv \mu \pmod{\mathfrak{p}^{k+e}}$ and 0 otherwise. Setting

$$W_k := \{x \in M_v : Q(x) \equiv \mu \pmod{\mathfrak{p}^k}\},$$

we can therefore write

$$(1 - q^{-s})^{-1} |\beta|_v^{-s} c_{\mu,v}(E_M; s) = \sum_{k=0}^\infty q^{k+e-ks} \operatorname{vol}(W_{e+k}).$$

The equality $\operatorname{vol}(W_k)/\operatorname{vol}(\Lambda_v) = \sharp A_{\mu,k}(M_v)/q^{kn}$ holds for sufficiently large k, namely for all k such that the modulus of M contains $\mathfrak{p}^k \Lambda$ (so that M is a union of cosets of $\mathfrak{p}^k \Lambda$). Because of this,

and the fact that $\sharp A_{\mu,k}(M_v)/q^{k(n-1)} = d$ for sufficiently large k, we have $d \cdot \tau_v = q^k \operatorname{vol}(W_k)$ for all but finitely many k, and so

$$(1 - q^{-s})^{-1}|\beta|_v^{-s}c_{\mu,v}(E_M; s) - d\tau_v \sum_{k=0}^{\infty} q^{-ks}$$

is a polynomial in q^{-s}. Multiplying by $1 - q^{-s}$ and evaluating at $s = 0$ we obtain the desired equality $c_{\mu,v}(E_M; 0) = d \cdot \tau_v$. $\qquad\square$

In light of (6.21), our next goal is to find a lower bound for $c_\mu(E_M; 0)$ as a function of μ; if we can show the growth rate is sufficiently large (in particular, fast enough that it overtakes the growth rate of any cusp form), then we will be able to conclude that the μth Fourier coefficient of the theta function of M is nonzero. Obtaining an explicit formula for this growth rate requires a fairly involved computation. Since this has already been done in [Han1] in the case M is a lattice, we will avoid this computation by relating the Fourier coefficients coming from lattice cosets to the Fourier coefficients coming from lattices.

All the above setup holds for the lattice Λ as well: that is, there is an Eisenstein series E_Λ for which

$$c_\mu(\theta_{\mathfrak{g}(\Lambda)}) = \frac{e^{2\pi \operatorname{Tr}\mu}}{m(\mathfrak{g}(\Lambda))}c_\mu(E_\Lambda; 0) \qquad \text{and} \qquad c_\mu(E_\Lambda; s) = d_L^{1/2}\prod_v c_{\mu,v}(E_\Lambda; s).$$

For finite places v, defining

(6.24) $$A_{\mu,k}(\Lambda_v) := \{x \in \Lambda_v/\mathfrak{p}^k\Lambda_v : Q(x) \equiv \mu \pmod{\mathfrak{p}^k}\}$$

we have $c_{\mu,v}(E_\Lambda; 0) = \tau_v(\sharp A_{\mu,k}(\Lambda))/q^{k(n-1)}$.

For further simplicity we now restrict our attention to cosets with modulus $\mathfrak{p}\Lambda$ for a single prime $\mathfrak{p}$ of $\mathcal{O}_L$.

Lemma 6.5.6. *Let Λ be an $\mathcal{O}_L$-integral lattice of rank $n \geq 4$, and let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_L$ with ideal norm q. Consider the lattice coset $M := t + \mathfrak{p}\Lambda$ for some $t \in \Lambda \setminus \mathfrak{p}\Lambda$ (so M has modulus $\mathfrak{p}\Lambda$). Suppose $\mu \in \mathcal{O}_L$ is coprime to $\operatorname{disc}\Lambda$, and that μ is locally represented at v by M. Then*

$$c_\mu(\theta_{\mathfrak{g}(M)}) > \frac{m(\mathfrak{g}(\Lambda))}{2q^{\lambda n}m(\mathfrak{g}(M))}c_\mu(\theta_{\mathfrak{g}(\Lambda)}),$$

where $\lambda = 2\operatorname{ord}_{\mathfrak{p}}(2) + 1$.

Proof. Let E_M and E_Λ denote the Eisenstein series corresponding to M and Λ respectively. Let v denote the place corresponding to $\mathfrak{p}$. At all places $v' \neq v$ (including infinite places) we have $M_{v'} = \Lambda_{v'}$ and so $c_{\mu,v'}(E_M; s) = c_{\mu,v'}(E_\Lambda; s)$. By (6.22) and its analogue for Λ,

$$c_\mu(E_M; 0)c_{\mu,v}(E_\Lambda; 0) = c_\mu(E_\Lambda; 0)c_{\mu,v}(E_M; 0).$$

By (6.21) and its analogue for Λ we can then write

$$c_\mu(\theta_{\mathfrak{g}(M)})c_{\mu,v}(E_\Lambda; 0) = \frac{m(\mathfrak{g}(\Lambda))}{m(\mathfrak{g}(M))}c_\mu(\theta_{\mathfrak{g}(\Lambda)})c_{\mu,v}(E_M; 0).$$

So it suffices to show $c_{\mu,v}(E_\Lambda; 0) \neq 0$ and $c_{\mu,v}(E_M; 0)/c_{\mu,v}(E_\Lambda; 0) > \frac{1}{2}q^{-\lambda n}$.

First consider the case $\mu \notin \mathfrak{p}$. We say $x \in \Lambda_v$ is "good" (cf. [Han1, Remark 3.1.1]) if there exists $y \in \Lambda_v$ with $(x, y) \notin \mathfrak{p}$, where $(\cdot, \cdot)$ is the inner product associated to Q as defined in §2.1. If $x \in \Lambda_v$ satisfies $Q(x) \equiv \mu$, then x is good as we can take $y = x$. Thus $A_{\mu,k}(M_v)$ and $A_{\mu,k}(\Lambda_v)$ both consist only of good elements. We can therefore use Hensel lifting to show that for any $\ell \geq 0$, the number of solutions mod $\mathfrak{p}^{\lambda+\ell}\Lambda_v$ to $Q(x) \equiv \mu \pmod{\mathfrak{p}^{\lambda+\ell}}$ is $q^{\ell(n-1)}$ times the number of solutions mod $\mathfrak{p}^\lambda\Lambda_v$ to $Q(x) \equiv \mu \pmod{\mathfrak{p}^\lambda}$ (see for instance [Han1, Lemma 3.2]). We can conclude that the equality in Lemma 6.5.5 holds already for $k = \lambda$, as does the

corresponding equality for $c_{\mu,v}(E_\Lambda;0)$. By assumption there exists $z \in M$ with $Q(z) = \mu$, so $\sharp A_{\mu,\lambda}(\Lambda_v) \geq \sharp A_{\mu,\lambda}(M_v) \geq 1$. From this we can conclude that $c_{\mu,v}(E_\Lambda;0) \neq 0$ and

$$\frac{c_{\mu,v}(E_M;0)}{c_{\mu,v}(E_\Lambda;0)} = \frac{\sharp A_{\mu,\lambda}(M_v)}{\sharp A_{\mu,\lambda}(\Lambda_v)} \geq \frac{1}{\sharp \Lambda_v / \mathfrak{p}^\lambda \Lambda_v} = \frac{1}{q^{\lambda n}} > \frac{1}{2q^{\lambda n}}.$$

Now suppose instead that $\mu \in \mathfrak{p}$. Then the assumption that μ is coprime to disc Λ implies that $\mathfrak{p} \nmid \operatorname{disc} \Lambda$. In particular, this implies that the quadratic form Q is non degenerate on $\Lambda_v / \mathfrak{p}\Lambda_v$, so that every element of $\Lambda_v \setminus \mathfrak{p}\Lambda_v$ (and in particular every element of M) is good. As above, we can conclude that

$$c_{\mu,v}(E_M;0) = \tau_v \frac{\sharp A_{\mu,\lambda}(M_v)}{q^{\lambda(n-1)}} \geq \frac{\tau_v}{q^{\lambda(n-1)}}$$

and that $c_{\mu,v}(E_\Lambda;0) \neq 0$. To compute an upper bound for $c_{\mu,v}(E_\Lambda;0)$ will take a bit more work, as there may be elements $x \in \mathfrak{p}\Lambda_v$ with $Q(x) = \mu$. These are not good solutions, and therefore the solutions mod $\mathfrak{p}^\lambda \Lambda_v$ may have more than the expected number of lifts to $\mathfrak{p}^{\lambda+\ell}\Lambda_v$.

To resolve this, we will partition the solutions to $Q(x) = \mu$ based on their $\mathfrak{p}$-adic valuation. Let k denote the $\mathfrak{p}$-adic valuation of μ. Every $x \in \Lambda_v$ with $Q(x) \equiv \mu \pmod{\mathfrak{p}^{k+1}}$ has valuation at most $k/2$, because if $x \in \mathfrak{p}^r$ for $r > k/2$, then $Q(x) \in \mathfrak{p}^{k+1}$ while $\mu \notin \mathfrak{p}^{k+1}$. For each $0 \leq r \leq k/2$, the set of $x \in \mathfrak{p}^r \Lambda_v \setminus \mathfrak{p}^{r+1}\Lambda_v$ with $Q(x) \equiv \mu \pmod{\mathfrak{p}^{2r+\lambda}}$ is in bijection (by dividing by a power of a uniformizer π of $\mathfrak{p}$) with the set of $x \in \Lambda_v \setminus \mathfrak{p}\Lambda_v$ with $Q(x) \equiv \mu/\pi^{2r} \pmod{\mathfrak{p}^\lambda}$. This latter set consists only of good elements. Thus for all $\ell \geq k$,

$$\sharp A_{\mu,\ell+\lambda}(\Lambda_v) = \#\{x \in \Lambda_v / \mathfrak{p}^{\ell+\lambda}\Lambda_v : Q(x) \equiv \mu \pmod{\mathfrak{p}^{\ell+\lambda}}\}$$

$$= \sum_{r=0}^{k/2} \#\{x \in (\mathfrak{p}^r \Lambda_v \setminus \mathfrak{p}^{r+1}\Lambda_v)/\mathfrak{p}^{\ell+\lambda}\Lambda_v : Q(x) \equiv \mu \pmod{\mathfrak{p}^{\ell+\lambda}}\}$$

$$= \sum_{r=0}^{k/2} \#\{x \in (\Lambda_v \setminus \mathfrak{p}\Lambda_v)/\mathfrak{p}^{\ell-r+\lambda}\Lambda_v : Q(x) \equiv \mu/\pi^{2r} \pmod{\mathfrak{p}^{\ell-2r+\lambda}}\}$$

$$\leq \sum_{r=0}^{k/2} q^{rn} \#\{x \in (\Lambda_v \setminus \mathfrak{p}\Lambda_v)/\mathfrak{p}^{\ell-2r+\lambda}\Lambda_v : Q(x) \equiv \mu/\pi^{2r} \pmod{\mathfrak{p}^{\ell-2r+\lambda}}\}$$

$$= \sum_{r=0}^{k/2} q^{rn} q^{(\ell-2r)(n-1)} \#\{x \in (\Lambda_v \setminus \mathfrak{p}\Lambda_v)/\mathfrak{p}^\lambda \Lambda_v : Q(x) \equiv \mu/\pi^{2r} \pmod{\mathfrak{p}^\lambda}\}.$$

The second last line follows because each coset of $\mathfrak{p}^{\ell-r+\lambda}\Lambda_v$ contains q^{rn} cosets of $\mathfrak{p}^{\ell-2r+\lambda}\Lambda_v$, and the last line follows by Hensel lifting of good solutions as above. Since the set in the last line is a subset of $\Lambda_v / \mathfrak{p}^\lambda \Lambda_v$ which has $q^{\lambda n}$ elements, we have

$$\sharp A_{\mu,\ell+\lambda}(\Lambda_v) \leq \sum_{r=0}^{k/2} q^{rn+(\ell-2r)(n-1)+\lambda n} < q^{\ell(n-1)+\lambda n} \sum_{r=0}^\infty q^{r(2-n)} = \frac{q^{(\ell+\lambda)(n-1)+\lambda}}{1 - q^{2-n}},$$

so that by Lemma 6.5.5,

$$c_{\mu,v}(E_\Lambda;0) < \frac{\tau_v q^\lambda}{1 - q^{2-n}} < 2q^\lambda \tau_v.$$

Therefore $c_{\mu,v}(E_M;0)/c_{\mu,v}(E_\Lambda;0) > \frac{1}{2} q^{-\lambda n}$ as desired. $\square$

6.5.3. *Proof of Theorem 6.2.1.* We first recall the theorem statement. We have $n \geq 4$ even, and Λ, $\mathfrak{p}$, and M as in the statement of Proposition 6.5.6. Assuming $\mu \in \mathcal{O}_L$ is coprime to disc Λ, $N_{L/\mathbb{Q}}(\mu)$ is sufficiently large, and μ and is everywhere locally represented by M, we wish to show that μ is (globally) represented by M. We use the notation $f(\mu) \gg g(\mu)$ to mean that there exists a positive constant c (not depending on μ, though it may depend on $L, \Lambda, M, \varepsilon, \mathfrak{p}$, etc.) such that for all μ with $N_{L/\mathbb{Q}}(\mu)$ sufficiently large we have $f(\mu) > cg(\mu)$.

We begin by bounding the Fourier coefficients of $\theta_{\mathfrak{g}(\Lambda)}$ from below; this was done by Hanke [Han1]. If $n \geq 5$ then by [Han1, Theorem 5.7(c)] we have

$$c_\mu(\theta_{\mathfrak{g}(\Lambda)}) \gg N_{L/\mathbb{Q}}(\mu)^{(n-2)/2}.$$

Now assume $n = 4$. The assumption that μ is coprime to disc Λ implies that Q is isotropic on V_v at all places v dividing μ [Han1, Remark 3.8.1]. So by [Han1, Theorem 5.7(b)] we have

$$c_\mu(\theta_{\mathfrak{g}(\Lambda)}) \gg N_{L/\mathbb{Q}}(\mu) \prod_{\mathfrak{p}|\mu,\, \chi(\mathfrak{p})=-1} \frac{N_{L/\mathbb{Q}}(\mathfrak{p}) - 1}{N_{L/\mathbb{Q}}(\mathfrak{p}) + 1}$$

where χ is the Hecke character on L defined by $\chi(\mathfrak{p}) := \left(\frac{\mathrm{disc}\,\Lambda_{\mathfrak{p}}}{\mathfrak{p}}\right)$. Noting that $\frac{x-1}{x+1} \geq \left(1 - \frac{1}{x}\right)^2$ for all $x \geq 1$, we can bound

$$\prod_{\mathfrak{p}|\mu,\, \chi(\mathfrak{p})=-1} \frac{N_{L/\mathbb{Q}}(\mathfrak{p}) - 1}{N_{L/\mathbb{Q}}(\mathfrak{p}) + 1} \geq \prod_{\mathfrak{p}:N_{L/\mathbb{Q}}(\mathfrak{p})\leq N_{L/\mathbb{Q}}(\mu)} \left(1 - \frac{1}{N_{L/\mathbb{Q}}(\mathfrak{p})}\right)^2 \gg \log(N_{L/\mathbb{Q}}(\mu))^{-2}$$

by Merten's third theorem for number fields [Leb, Theorem 5]. In particular, for all $n \geq 4$ and all $\varepsilon > 0$ we have

$$c_\mu(\theta_{\mathfrak{g}(\Lambda)}) \gg N_{L/\mathbb{Q}}(\mu)^{(n-2)/2-\varepsilon}.$$

By Lemma 6.5.6, the ratio of $c_\mu(\theta_{\mathfrak{g}(M)})$ to $c_\mu(\theta_{\mathfrak{g}(\Lambda)})$ is bounded below by a positive constant (not depending on μ), so this bound still holds if we replace $\mathfrak{g}(\Lambda)$ by $\mathfrak{g}(M)$.

Now assume n is even; then by Theorem 6.4.4, $|c_\mu(\theta_{\mathfrak{g}(M)}) - c_\mu(\theta_M)| \ll N_{L/\mathbb{Q}}(\mu)^{(n/4-1/2)-\varepsilon'}$ for all $\varepsilon' > 0$. Since $n/2 - 1 > n/4 - 1/2$ for all $n \geq 4$, we see that $c_\mu(\theta_{\mathfrak{g}(M)})$ grows faster than $|c_\mu(\theta_{\mathfrak{g}(M)}) - c_\mu(\theta_M)|$, and therefore $c_\mu(\theta_M) > 0$ for $N_{L/\mathbb{Q}}(\mu)$ sufficiently large. This concludes the proof.

Remark 6.5.7. The theorem statement should hold for odd $n \geq 5$ as well (and potentially also for $n = 3$ if additional constraints are imposed). The main missing piece is a bound on the Fourier coefficients of the cusp form $\theta_{\mathfrak{g}} - \theta_M$. When n is odd, this cusp form is a modular form of half-integral weight. It is certainly possible to bound the Fourier coefficients of cusp forms of half-integral weight, by relating a cusp form of weight $w \in \mathbb{Z} + \frac{1}{2}$ to its "Shimura lift" of weight $2w - 1$ as described in [Shi1]. This is worked out explicitly for theta functions of lattices with $L = \mathbb{Q}$ in [Han1, §4], but to reproduce such bounds for theta functions of lattice cosets over totally real fields would take us too far afield for this paper, and is not necessary for our main application.

Remark 6.5.8. The condition that μ be coprime to disc Λ is slightly stronger than necessary, but for $n = 4$ it cannot be removed completely. In general, the size of the Fourier coefficient $c_\mu(\theta_{\mathfrak{g}(\Lambda)})$ depends not on $N_{L/\mathbb{Q}}(\mu)$, but rather on $N_{L/\mathbb{Q}}(\mu_{\mathrm{Iso}})$, where μ_{Iso} is the greatest divisor of μ that is coprime to all primes v such that Q is anisotropic on Λ_v. For $n \geq 5$ no such primes exist and so $\mu = \mu_{\mathrm{Iso}}$ always; in general, such primes always divide disc Λ, and so taking μ coprime to disc Λ ensures $\mu = \mu_{\mathrm{Iso}}$.

However, if $n = 4$, Q is anisotropic on Λ_v, and μ is divisible by a very large power of the prime at v, then $c_\mu(\theta_{\mathfrak{g}(\Lambda)})$ can be extremely small even if $N_{L/\mathbb{Q}}(\mu)$ is large; thus it is possible to have μ of arbitrarily large norm that is everywhere locally representable by M and yet $c_\mu(\theta_M) = 0$. An example of this behavior is given by Watson, who showed that for the lattice $\Lambda = \mathbb{Z}^4$ with quadratic form $w^2 + x^2 + 7y^2 + 7z^2$, and for $\mu := 3 \cdot 7^{2u}$ with any $u \geq 0$, the Fourier coefficient $c_\mu(\theta_\Lambda)$ is zero (there is no element of Λ with norm μ) even though μ is everywhere locally represented [Wat, Section 7.7]. In this example μ can be arbitrarily large, but $\mu_{\mathrm{Iso}} = 3$ since the lattice is anisotropic at 7.

6.6. **Generating lattices over totally real fields by elements of specified norm.**

Theorem 6.6.1. *Let Λ be an integral positive-definite $\mathcal{O}_L$-lattice of even rank $n \geq 4$. Let $S \subset \mathcal{O}_L$ be a set of totally positive elements coprime to $\operatorname{disc}\Lambda$ such that $\{N_{L/\mathbb{Q}}(s) : s \in S\}$ is infinite. Suppose that for all $s \in S$ and all finite places v of L, Λ_v is generated as an $\mathcal{O}_{L,v}$-module by elements of norm s. Then Λ is generated as an $\mathcal{O}_L$-module by elements with norm in S.*

Remark 6.6.2. The condition that n be even can likely be removed; see Remark 6.5.7.

Proof. Let $s_1, s_2, \ldots$ be a sequence of elements of S with $N_{L/\mathbb{Q}}(s_i) \to \infty$. Let $\mathfrak{p}$ be a prime of $\mathcal{O}_L$ with corresponding place v. For each $i = 1, 2, \ldots$, let $C_i \subset \Lambda_v$ be a generating set of Λ_v consisting of elements x with $Q(x) = s_i$. Since there are only finitely many subsets of $\Lambda_v/\mathfrak{p}\Lambda_v$ but infinitely many C_i, there must exist an infinite sequence $i_1, i_2, \ldots$ such that C_{i_j} have the same image $\overline{C} \subset \Lambda_v/\mathfrak{p}\Lambda_v \simeq \Lambda/\mathfrak{p}\Lambda$ for all j.

Fix some element $\bar{t} \in \overline{C}$, and let $t \in \Lambda$ be a preimage of $\bar{t}$ under reduction mod $\mathfrak{p}\Lambda$. Then $M = t + \mathfrak{p}\Lambda$ is a lattice coset with modulus $\mathfrak{p}\Lambda$, and M_v is the preimage of $\bar{t}$ under the map $\Lambda_v \to \Lambda_v/\mathfrak{p}\Lambda_v$. For all j, there exists an element of C_{i_j} in M_v; it has norm s_{i_j}. Further, for all finite places $w \neq v$, we have $M_w = \Lambda_w$, and Λ_w has an element of norm s_{i_j} by assumption. Thus by Theorem 6.2.1, there exists an element of M with norm s_{i_j} for some sufficiently large j. Repeating this for each $\bar{t} \in \overline{C}$, we find that there exists a set $D \subset \Lambda$ consisting of elements with norm in S such that D maps surjectively onto $\overline{C}$ under reduction mod $\mathfrak{p}\Lambda$. Since $\overline{C}$ generates $\Lambda/\mathfrak{p}\Lambda$, the lattice generated by D maps surjectively onto $\Lambda/\mathfrak{p}\Lambda$.

We have shown that for all primes $\mathfrak{p}$, the set of elements in Λ with norm in S generate a lattice that surjects onto $\Lambda/\mathfrak{p}\Lambda$. This implies that the set of elements in Λ with norm in S generate Λ. $\qquad\square$

As a consequence we obtain the desired theorem for endomorphism rings of ssAVs with RM.

Theorem 6.6.3. *Let $\underline{A}$ be a ssAV with RM by $\mathcal{O}_L$ defined over a finite field k with characteristic p. Let $\ell \neq p$ be a rational prime that is inert in L. Then $\operatorname{End}(\underline{A})$ is generated as an $\mathcal{O}_L$-module by elements with norm in $\{\ell^k : k \geq 0\}$.*

Proof. As discussed in §6.2.1, $\underline{A}$ is isogenous to $\underline{E \otimes \mathcal{O}_L}$ for some supersingular elliptic curve E/k. As explained in §6.2, this implies that the rank 4 $\mathcal{O}_L$-lattices $\operatorname{End}(\underline{E \otimes \mathcal{O}_L})$ and $\operatorname{End}(\underline{A})$ are in the same genus, so for all finite places v of L lying above a finite prime q, $\operatorname{End}(\underline{A})_v \simeq \mathcal{O}_q \otimes_{\mathbb{Z}_q} \mathcal{O}_{L,v}$. In [GLo, Section 2.3] we proved that for all $s \in S$, $\mathcal{O}_q$ has a $\mathbb{Z}$-basis consisting of elements of norm s. These same elements (tensored with $1 \in \mathcal{O}_{L,v}$) form a $\mathcal{O}_{L,v}$-basis of $\operatorname{End}(\underline{A})_v$ consisting of elements with norm $s \otimes 1^2 = s$. As every element of S is coprime to $\operatorname{disc}\operatorname{End}(\underline{A}) = p\mathcal{O}_L$, we may apply Theorem 6.6.1 to $\Lambda = \operatorname{End}(\underline{A})$. $\qquad\square$

REFERENCES

[Adj] Adj, Gora; Ahmadi, Omran; Menezes, Alfred: On isogeny graphs of supersingular elliptic curves over finite fields. Finite Fields Appl. 55 (2019), 268–283.

[ATY] Aikawa, Yusuke; Tanaka, Ryokichi; Yamauchi, Takuya: Isogeny graphs on superspecial abelian varieties: eigenvalues and connection to Bruhat-Tits buildings. Canadian Journal of Mathematics (2023). https://doi.org/10.4153/S0008414X23000676.

[Arp] Arpin, Sarah: Adding Level Structure to Supersingular Elliptic Curve Isogeny Graphs. J. Théor. Nombres Bordeaux, to appear.

[ACN+] Arpin, Sarah; Camacho-Navarro, Catalina; Lauter, Kristin; Lim, Joelle; Nelson, Kristina; Scholl, Travis; Sotáková, Jana: Adventures in supersingularland. Exp. Math. 32 (2023), no.2, 241–268.

[BG] Bachmat, E.; Goren, E. Z.: On the non-ordinary locus in Hilbert-Blumenthal surfaces. Math. Ann. 313 (1999), no.3, 475–506.

[BGJ+] Baker, Matthew H.; González-Jiménez, Enrique; González, Josep; Poonen, Bjorn: Finiteness results for modular curves of genus at least 2. Amer. J. Math. 127 (2005), no.6, 1325–1387.

[BCN+] Bank, Efrat; Camacho-Navarro, Catalina; Eisenträger, Kirsten; Morrison, Travis; Park, Jennifer: Cycles in the supersingular ℓ-isogeny graph and corresponding endomorphisms, in *Research directions in number theory–Women in Numbers IV*, 41–66. Assoc. Women Math. Ser., 19.

[BL+] Bernstein, Daniel J.; Lange, Tanja; Martindale, Chloe; Panny, Lorenz: Quantum circuits for the CSIDH: optimizing quantum evaluation of isogenies, in *Advances in cryptology—EUROCRYPT 2019. Part II*, 409–441. Lecture Notes in Computer Science, 11477, Springer.

[BIJ] Biasse, J.-F.; Iezzi, A.; Jacobson, M. J.: A note on the security of CSIDH, in *Progress in cryptology—INDOCRYPT 2018*, 153–168, Lecture Notes in Comput. Sci., 11356, Springer.

[BiL] Birkenhake, Christina; Lange, Herbert: *Complex abelian varieties.* Second edition. Grundlehren Math. Wiss., 302. Springer, 2004.

[Bla] Blasius, D.: Hilbert modular forms and the Ramanujan conjecture. https://arxiv.org/abs/math/0511007.

[BM] Bost, Jean-Benoît; Mestre, Jean-François: Moyenne arithmético-géométrique et périodes des courbes de genre 1 et 2. Gaz. Math.(1988), no.38, 36–64.

[BS] Bonnetain, X.; Schrottenloher, A.: Quantum security analysis of CSIDH, in *Advances in Cryptology—EUROCRYPT 2020. Part II*, 493–522, Lecture Notes in Comput. Sci., 12106, Springer.

[BB+] Booher, Jeremy; Bowden, Ross; Doliskani, Javad; Fouotsa, Tako Boris; Galbraith, Steven D.; Kunzweiler, Sabrina; Merz, Simon-Philipp; Petit, Christophe; Smith, Benjamin; Stange, Katherine E.; Ti, Yan Bo; Vincent, Christelle; Voloch, José Felipe; Weitkämper, Charlotte; Zobernig, Lukas: Failing to Hash Into Supersingular Isogeny Graphs. The Computer Journal, vol. 67 (8), 2024.

[Bor] Borel, Armand: Some finiteness properties of adele groups over number fields. Inst. Hautes Études Sci. Publ. Math. (1963), no.16, 5–30.

[BLS] Bröker, Reinier; Lauter, Kristin; Sutherland, Andrew V.: Modular polynomials via isogeny volcanoes. Math. Comp. 81 (2012), no.278, 1201–1231.

[BrL] Brylinski, J.-L.; Labesse, J.-P.: Cohomologie d'intersection et fonctions L de certaines variétés de Shimura. Ann. Sci. École Norm. Sup. (4) 17 (1984), no.3, 361–412.

[Bue] Buell, Duncan A.: *Binary quadratic forms.* Classical theory and modern computations. Springer-Verlag, New York, 1989.

[Bum] Bump, Daniel: *Automorphic forms and representations.* Cambridge Stud. Adv. Math., 55 Cambridge University Press, Cambridge, 1997.

[CD] Castryck, Wouter; Decru, Thomas: An efficient key recovery attack on SIDH, in *Advances in cryptology–EUROCRYPT 2023*. Part V, 423–447. Lecture Notes in Comput. Sci., 14008.

[CDS] Castryck, Wouter; Decru, Thomas; Smith, Benjamin: Hash functions from superspecial genus-2 curves using Richelot isogenies. J. Math. Cryptol.14 (2020), no.1, 268–292.

[CL+] Castryck, Wouter; Lange, Tanja; Martindale, Chloe; Panny, Lorenz; Renes, Joost. CSIDH: an efficient post-quantum commutative group action, in *Advances in cryptology—ASIACRYPT 2018. Part III*, 395–427, Lecture Notes in Comput. Sci., 11274, Springer.

[CV] Castryck, Wouter; Vercauteren, Frederik: A polynomial time attack on instances of M-SIDH and FESTA, in *Advances in cryptology–ASIACRYPT 2023. Part VII*, 127–156. Lecture Notes in Comput. Sci., 14444.

[CO] Chan, Wai Kiu; Oh, Byeong-Kweon: Representations of integral quadratic polynomials, in *Diophantine methods, lattices, and arithmetic theory of quadratic forms*, 31–46. Contemp. Math., 587 American Mathematical Society, Providence, RI, 2013.

[CGL1] Charles, Denis X.; Goren, Eyal Z.; Lauter, Kristin E.: Cryptographic hash functions from expander graphs. J. Cryptology 22 (2009), no.1, 93–113.

[CGL2] Charles, Denis X.; Goren, Eyal Z.; Lauter, Kristin E.: Families of Ramanujan graphs and quaternion algebras, in *Groups and symmetries*, 53–80. CRM Proc. Lecture Notes, 47. American Mathematical Society, Providence, RI, 2009.

[CG] Chevyrev, Ilya; Galbraith, Steven D.: Constructing supersingular elliptic curves with a given endomorphism ring. LMS J. Comput. Math.17 (2014), 71–91.

[CK] Colò, L.; Kohel, D. R.: Orienting supersingular isogeny graphs, J. Math. Cryptol. **14** (2020), no. 1, 414–437.

[Con] Conrad, Brian: Gross-Zagier revisited, in *Heegner points and Rankin L-series*, 67–163. With an appendix by W. R. Mann. Math. Sci. Res. Inst. Publ., 49. Cambridge University Press, 2004.

[CSl] Conway, J. H.; Sloane, N. J. A.: Four-dimensional lattices with the same theta series, Internat. Math. Res. Notices **1992**, no. 4, 93–96.

[CSm] Costello, Craig; Smith, Benjamin: The supersingular isogeny problem in genus 2 and beyond, in *Post-quantum cryptography*, 151–168. Lecture Notes in Comput. Sci., 12100 Springer, Cham, 2020.

[CM]   Cowan, Alex; Martin, Kimball: Moduli for rational genus 2 curves with real multiplication for discriminant 5. https://arxiv.org/abs/2206.05752.

[CFM] Cowan, Alex; Frengley, Sam; Martin, Kimball: Generic models for genus 2 curves with real multiplication. https://arxiv.org/abs/2403.03191.

[DdF] Dartois, P.; De Feo,L.: On the security of OSIDH, in *Public-key cryptography—PKC 2022. Part I*, 52–81, Lecture Notes in Comput. Sci., 13177, Springer.

[DH+] Darmon, Henri; Harris, Michael; Rotger, Victor; Venkatesh, Akshay: The derived Hecke algebra for dihedral weight one forms. Michigan Math. J. 72 (2022), 145–207.

[Dar] Dartois, Pierrick: Fast computation of 2-isogenies in dimension 4 and cryptographic applications. https://arxiv.org/abs/2407.15492.

[DL+] Dartois, Pierrick; Leroux, Antonin; Robert, Damien; Wesolowski, Benjamin: SQIsignHD: new dimensions in cryptography, in *Advances in cryptology—EUROCRYPT 2024. Part I*, 3–32, Lecture Notes in Comput. Sci., 14651, Springer.

[dFJP] De Feo, Luca; Jao, David; Plût, Jérôme: Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. J. Math. Cryptol. 8 (2014), no.3, 209–247.

[dFK+] De Feo, Luca; Kohel, David; Leroux, Antonin; Petit, Christophe; Wesolowski, Benjamin: SQISign: compact post-quantum signatures from quaternions and isogenies, in *Advances in cryptology–ASIACRYPT 2020. Part I*, 64–93. Lecture Notes in Comput. Sci., 12491.

[dFM+] De Feo, Luca; Masson, Simon; Petit, Christophe; Sanso, Antonio: Verifiable delay functions from supersingular isogenies and pairings, in *Advances in cryptology–ASIACRYPT 2019*. Part I, 248–277. Lecture Notes in Comput. Sci., 11921.

[DG]   Delfs, Christina; Galbraith, Steven D.: Computing isogenies between supersingular elliptic curves over $\mathbb{F}_p$. Des. Codes Cryptogr. 78 (2016), no.2, 425–440.

[DP]   Deligne, Pierre; Pappas, Georgios: Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant. Compositio Math. 90 (1994), no.1, 59–79.

[DH]   Diffie, Whitfield; Hellman, Martin E.: New directions in cryptography. IEEE Trans. Inform. Theory IT – 22 (1976), no.6, 644–654.

[DL]   Donagi, Ron; Livné, Ron: The arithmetic-geometric mean and isogenies for curves of higher genus. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 28 (1999), no.2, 323–339.

[EH+] Eisenträger, Kirsten; Hallgren, Sean; Lauter, Kristin; Morrison, Travis; Petit, Christophe: Supersingular isogeny graphs and endomorphism rings: reductions and solutions, in *Advances in cryptology–EUROCRYPT 2018*. Part III, 329–368. Lecture Notes in Comput. Sci., 10822.

[Eic1] Eichler, M.: The basis problem for modular forms and the traces of the Hecke operators, in *Modular functions of one variable, I* (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), pp. 75–151. Lecture Notes in Math., Vol. 320.

[Eic2] Eichler, Martin: On theta functions of real algebraic number fields. Acta Arith. 33 (1977), no.3, 269–292.

[EOY] Elkies, Noam; Ono, Ken; Yang, Tonghai: Reduction of CM elliptic curves and modular function congruences. Int. Math. Res. Not. (2005), no.44, 2695–2707.

[FT]   Flynn, E. V.; Ti, Yan Bo: Genus two isogeny cryptography, in *Post-quantum cryptography*, 286–306. Lecture Notes in Comput. Sci., 11505.

[Gar] Garrett, Paul B.: *Holomorphic Hilbert modular forms.* Wadsworth & Brooks/Cole Math. Ser., Wadsworth & Brooks/Cole Advanced Books & Software, 1990.

[vdG] van der Geer, Gerard: *Hilbert modular surfaces.* Ergeb. Math. Grenzgeb. (3), 16. Springer-Verlag, Berlin, 1988.

[Gel] Gelbart, Stephen S.: *Automorphic forms on adèle groups.* Ann. of Math. Stud., No. 83, Princeton University Press, Princeton, NJ; University of Tokyo Press, Tokyo, 1975.

[Gha] Ghantous, Wissam: Loops, Multi-Edges and Collisions in Supersingular Isogeny Graphs. Advances in Mathematics of Communications, Volume 18, Issue 4 (2024): 935–955. Doi: 10.3934/amc.2022038

[GMR] Goldwasser, Shafi; Micali, Silvio; Rivest, Ronald L.: A digital signature scheme secure against adaptive chosen-message attacks. Special issue on cryptography SIAM J. Comput. 17 (1988), no.2, 281–308.

[Gor1] Goren, Eyal Z.: On certain reduction problems concerning abelian surfaces. Manuscripta Math. 94 (1997), no.1, 33–43.

[Gor2] Goren, Eyal Z.: *Lectures on Hilbert modular varieties and modular forms.* With the assistance of Marc-Hubert Nicole CRM Monogr. Ser., 14. American Mathematical Society, Providence, RI, 2002.

[GK]   Goren, Eyal Z.; Kassaei, Payman L.: *p*-adic dynamics of Hecke operators on modular curves. J. Théor. Nombres Bordeaux 33 (2021), no.2, 387–431.

[GL1] Goren, Eyal Z.; Lauter, Kristin E.: Class invariants for quartic CM fields. Ann. Inst. Fourier (Grenoble) 57 (2007), no.2, 457–480.

[GL2] Goren, Eyal Z.; Lauter, Kristin E.: Evil primes and superspecial moduli. Int. Math. Res. Not. (2006), Art. ID 53864, 19 pp.

[GLo] Goren, Eyal Z.; Love, Jonathan: On elements of prescribed norm in maximal orders of a quaternion algebra. 29 pp. *To appear in Canad. J. Math.* https://arxiv.org/pdf/2307.16828

[GO] Goren, E. Z.; Oort, F.: Stratifications of Hilbert modular varieties. J. Algebraic Geom. 9 (2000), no.1, 111–154.

[Gro] Gross, Benedict H.: Heights and the special values of L-series, in *Number theory (Montreal, Que., 1985)*, 115–187. CMS Conf. Proc., 7, Published by the American Mathematical Society, Providence, RI, 1987.

[GZ] Gross, Benedict H.; Zagier, Don B.: On singular moduli. J. Reine Angew. Math. 355 (1985), 191–220.

[Han1] Hanke, Jonathan: Local densities and explicit bounds for representability by a quadratic form. Duke Math. J. 124 (2004), no.2, 351–388.

[Han2] Hanke, Jonathan: Quadratic forms and automorphic forms, in *Quadratic and higher degree forms*, 109–168. Dev. Math., 31. Springer, 2013.

[Jao] Jao, David et. al.: SIKE: a Supersingular Isogeny Key Encapsulation. (2017). https://sike.org/.

[Jor] Jordan, Bruce W.: Isogeny graphs of superspecial abelian varieties, in *Theory and Applications of Supersingular Curves and Supersingular Abelian Varieties, 131–144.* Research Institute for Mathematical Sciences (RIMS), Kyoto, 2022.

[JZ] Jordan, Bruce W.; Zaytman, Yevgeny: Isogeny graphs of superspecial abelian varieties and Brandt matrices. https://arxiv.org/abs/2005.09031.

[KK] Kane, B.; Kim, Daejun: Theta series of ternary quadratic lattice cosets. 28 pp. https://arxiv.org/abs/2211.03987.

[Kan] Kaneko, Masanobu: Supersingular j-invariants as singular moduli mod p. Osaka J. Math. 26 (1989), no.4, 849–855.

[KL] Knightly, Andrew; Li, Charles: *Traces of Hecke operators*. Math. Surveys Monogr., 133 American Mathematical Society, 2006.

[Koh] Kohel, David: *Endomorphism rings of elliptic curves over finite fields*. Berkeley: University of California. Ph.D. thesis, 1996.

[KL+] Kohel, David; Lauter, Kristin; Petit, Christophe; Tignol, Jean-Pierre: On the quaternion ℓ-isogeny path problem. LMS J. Comput. Math. 17 (2014), 418–432.

[KSW] Koo, K.T.-L.; Stein, W.; Wiese, G.: On the generation of the coefficient field of a newform by a single Hecke eigenvalue, J. Théor. Nr. Bordx. 20 (2) (2008) 373–384.

[KR] Kudla, S. S.; Rallis, S. J.: On the Weil-Siegel formula, J. Reine Angew. Math. **387** (1988), 1–68.

[Lan] Lang, Serge: *Elliptic functions.* With an appendix by J. Tate. Second edition. Grad. Texts in Math., 112 Springer, 1987.

[LP] Lauter, Kristin E.; Petit, Christophe: Supersingular isogeny graphs in cryptography. Surveys in combinatorics 2019, 143–165. London Math. Soc. Lecture Note Ser., 456. Cambridge University Press, 2019.

[Leb] Lebacque, P.: Generalised Mertens and Brauer-Siegel theorems, Acta Arith. **130** (2007), no. 4, 333–350.

[Liv] Livné, R.: Communication networks and Hilbert modular forms, in *Applications of algebraic geometry to coding theory, physics and computation* (Eilat, 2001), 255–270. NATO Sci. Ser. II Math. Phys. Chem., 36. Kluwer Academic Publishers, 2001.

[Lov] Love, Jonathan: *Isogeny graphs, zero-cycles, and modular forms : computations over algebraic curves and surfaces*. Ph.D. Thesis, Stanford 2021. https://searchworks.stanford.edu/view/13874834.

[LB] Love, Jonathan; Boneh, Dan: Supersingular curves with small noninteger endomorphisms, in *ANTS XIV – Proceedings of the Fourteenth Algorithmic Number Theory Symposium*, 7–22. Open Book Ser., 4.

[LPS] Lubotzky, A.; Phillips, R.; Sarnak, P.: Ramanujan graphs. Combinatorica 8 (1988), no.3, 261–277.

[MM+] Maino, Luciano; Martindale, Chloe; Panny, Lorenz; Pope, Giacomo; Wesolowski, Benjamin: A direct key recovery attack on SIDH, in *Advances in cryptology–EUROCRYPT 2023. Part V*, 448–471. Lecture Notes in Comput. Sci., 14008.

[May] Mayo, Sam: Automorphisms of the supersingular isogeny graph. Finite Fields Appl. 93 (2024), Paper No. 102314, 7 pp.

[Mes] Mestre, J.-F.: La méthode des graphes. Exemples et applications, in *Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata, 1986)*, 217–242.

[NIS1] National Institute of Standards and Technology. Post-quantum cryptography: Selected algorithms 2022. (2022). https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022.

[NIS2] National Institute of Standards and Technology. Standardization of Additional Digital Signature Schemes. (2023). https://csrc.nist.gov/Projects/pqc-dig-sig/standardization.

[Nic1] Nicole, Marc-Hubert: *Superspecial abelian varieties, theta series and the Jacquet-Langlands correspondence.* Thesis (Ph.D.)–McGill University (Canada) ProQuest LLC, Ann Arbor, MI, 2005. 145 pp.

[Nic2] Nicole, Marc-Hubert: Superspecial abelian varieties and the Eichler basis problem for Hilbert modular forms. J. Number Theory 128 (2008), no.11, 2874–2889.

[Ome] O'Meara, O. Timothy: *Introduction to quadratic forms.* Classics Math., Springer, 2000.

[Oor] Oort, Frans: Which abelian surfaces are products of elliptic curves? Math. Ann. 214 (1975), 35–47.

[PW] Page, Aurel; Wesolowski, Benjamin: The supersingular Endomorphism Ring and One Endomorphism problems are equivalent. arXiv:2309.10432, 2024.

[Piz1] Pizer, Arnold: The action of the canonical involution on modular forms of weight 2 on $\Gamma_0(M)$. Math. Ann. 226 (1977), no.2, 99–116.

[Piz2] Pizer, Arnold: An algorithm for computing modular forms on $\Gamma_0(N)$. J. Algebra 64 (1980), no.2, 340–390.

[PLQ] Petit, Christophe: Lauter, Kristin; Quisquater, Jean-Jacques: Full cryptanalysis of LPS and Morgenstern hash functions, in *Security and Cryptography for Networks (Berlin, Heidelberg) (Rafail Ostrovsky, Roberto De Prisco, and Ivan Visconti, eds.)*, Springer, 2008, pp. 263–277.

[PW] Page, Aurel; Wesolowski, Benjamin: The supersingular Endomorphism Ring and One Endomorphism problems are equivalent. https://arxiv.org/abs/2309.10432

[PR] Platonov, Vladimir; Rapinchuk, Andrei: Algebraic groups and number theory. Pure Appl. Math., 139. Academic Press, Inc., Boston, MA, 1994.

[Pra] Prasad, Dipendra: A brief survey on the theta correspondence.Number theory (Tiruchirapalli, 1996), 171–193. Contemp. Math., 210. American Mathematical Society, Providence, RI, 1998.

[Ral] Rallis, S. On the Howe duality conjecture. Compositio Math. 51 (1984), no.3, 333–399.

[Rap] Rapoport, M.: Compactifications de l'espace de modules de Hilbert-Blumenthal. Compositio Math. 36 (1978), no.3, 255–335.

[Rod] Roda, Megan: *Supersingular isogeny graphs with level n structure and path problems on ordinary isogeny graphs.* M.Sc. Thesis, McGill University, Montréal, 2019.

[Sch] Schoof, René: Elliptic curves over finite fields and the computation of square roots mod p. Math. Comp. 44 (1985), no.170, 483–494.

[Ser] Serre, J.-P.: Complex multiplication, in *Algebraic Number Theory*, J.W.S. Cassels and A. Fröhlich, eds., Academic Press, London, 1967, 292–296.

[Sha] Shahidi, Freydoon: On the Ramanujan conjecture for quasisplit groups. Asian J. Math. 8 (2004), no.4, 813– 835.

[Shi1] Shimura, G.: On the Fourier coefficients of Hilbert modular forms of half-integral weight, Duke Math. J. **71** (1993), no. 2, 501–557.

[Shi2] Shimura, G.: *Euler products and Eisenstein series*, CBMS Regional Conference Series in Mathematics, 93. Amer. Math. Soc., Providence, RI, 1997.

[Shi3] Shimura, G.: The number of representations of an integer by a quadratic form, Duke Math. J. **100** (1999), no. 1, 59–92.

[Shi4] Shimura, Goro: *Arithmeticity in the theory of automorphic forms.* Math. Surveys Monogr., 82. American Mathematical Society, Providence, RI, 2000.

[Shi5] Shimura, Goro: Inhomogeneous quadratic forms and triangular numbers. Amer. J. Math. 126 (2004), no.1, 191–214.

[Sh] Shioda, Tetsuji: Supersingular K3 surfaces, in *Algebraic geometry* (Proc. Summer Meeting, Univ. Copenhagen, Copenhagen, 1978), pp. 564–591 Lecture Notes in Math., 732. Springer, 1979.

[Sie] Siegel, C. L.: Über die analytische Theorie der quadratischen Formen, Ann. of Math. (2) **36** (1935), no. 3, 527–606.

[Sil1] Silverman, Joseph H.: *The arithmetic of elliptic curves.* Second edition. Grad. Texts in Math., 106. Springer, 2009.

[Sil2] Silverman, Joseph H.: *Advanced topics in the arithmetic of elliptic curves.* Grad. Texts in Math., 151. Springer, 1994.

[Sut1] Sutherland, Andrew V.: Isogeny volcanoes, in *ANTS X. – Proceedings of the Tenth Algorithmic Number Theory Symposium*, 507–530. Open Book Ser., 1 Mathematical Sciences Publishers, Berkeley, CA, 2013.

[Sut2] Sutherland, Andrew V.: Modular polynomials. https://math.mit.edu/~drew/ClassicalModPolys.html

[TZ] Tillich, Jean-Pierre; Zémor, Gilles: Collisions for the LPS expander graph hash function, in *Advances in cryptology–EUROCRYPT 2008*, 254–269. Lecture Notes in Comput. Sci., 4965.

[Ulm] Ulmer, Dougles L.: p-descent in characteristic p. Duke Mathematical Journal Vol. 62, Issue 2, (Mar 1991) , pgs 237-265

[Vél] Vélu, Jacques: Isogénies entre courbes elliptiques. C. R. Acad. Sci. Paris Sér. A-B 273 (1971), A238–A241.

[Vig] Vignéras, Marie-France: *Arithmétique des algèbres de quaternions.* Lecture Notes in Math., 800. Springer, 1980.

[Voi] Voight, John: *Quaternion algebras.* Grad. Texts in Math., 288. Springer, 2021.

[Wal1] Walling, Lynne H.: Hecke operators on theta series attached to lattices of arbitrary rank. Acta Arith. 54 (1990), no.3, 213–240.

[Wal2] Walling, Lynne H.: A remark on differences of theta series. J. Number Theory 48 (1994), no.2, 243–251.

[Was] Washington, Lawrence C.: *Elliptic curves. Number theory and cryptography*. Second edition. Discrete Math. Appl. (Boca Raton), Chapman & Hall/CRC, Boca Raton, FL, 2008.

[Wat] Watson, G. L.: *Integral quadratic forms*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 51, Cambridge Univ. Press, 1960.

[Wei] Weil, A.: Sur la formule de Siegel dans la théorie des groupes classiques, Acta Math. **113** (1965), 1–87.

[Wes] Wesolowski, Benjamin: The supersingular isogeny path and endomorphism ring problems are equivalent, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science–FOCS 2021*, 1100–1111.

[Yu] Yu, Chia-Fu: On the mass formula of supersingular abelian varieties with real multiplications. J. Aust. Math.
       Soc. 78 (2005), no.3, 373–392.
[Zag] Zagier, Don: On the values at negative integers of the zeta-function of a real quadratic field. Enseign. Math.
       (2) 22 (1976), no.1-2, 55–95.

DEPARTMENT OF MATHEMATICS AND STATISTICS, MCGILL UNIVERSITY, 805 SHERBROOKE ST. W., MONTREAL
H3A 2K6, QC, CANADA.
   E-mail address: `eyal.goren@mcgill.ca; jon.love@mcgill.ca`