Rational Points on Atkin-Lehner Quotients of Shimura Curves

A thesis presented

by

Pete L. Clark

to

The department of mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of mathematics.

Harvard University
Cambridge, Massachusetts
April, 2003

Barry Mazur Pete L. Clark

Pete L. Clark

Rational Points on Atkin-Lehner Quotients of Shimura Curves

Abstract

We study three families of Atkin-Lehner quotients of quaternionic Shimura curves: X^{D+}, $X_0^{D+}(N)$, and $X_1^{D+}(N)$, which serve as moduli spaces of abelian surfaces with potential quaternionic multiplication (PQM) and level N structure. The arithmetic geometry of these curves is similar to, but even richer than, that of the classical modular curves. Two important differences are the existence of a nontrivial obstruction to an abelian surface being defined over its field of moduli and the lack of cusps, due to which there may fail to be any points rational over a given field. We study the existence of points on these curves rational over both local and global fields, and consider applications to the existence of PQM surfaces over Q.

Contents

Introduction

In this thesis we study local and global points on certain Atkin-Lehner quotients of quaternionic Shimura curves, with and without level structure: we consider the curves $X^{D+},\ X_0^{D+}(N),\ X_1^{D+}(N)$, where the + indicates a quotient by the Atkin-Lehner involution w_D. It has long been known that, without the passage to the Atkin-Lehner quotient, none of these curves have R-points, much less Q-points. On the other hand, work of [Jordan I] and [Rotger II-IV] shows that these plus quotient Shimura curves are, at least morally, moduli spaces of abelian surfaces with potentially quaternionic multiplication (PQM), i.e., principally polarized abelian surfaces A/K which admit QM over the algebraic closure $\overline{K}$. In particular it makes sense to ask about the existence of PQMs A/Q as a function of the quaternionic discriminant D.

A PQM A/Q is an interesting object: in terms of its ℓ-adic Galois representations it looks like an elliptic curve without complex multiplication (hence the terminology "false elliptic curve" coined by Serre for these objects); for example, it has ordinary reduction on a density one set of primes. On the other hand it has, like a CM elliptic curve, everywhere potentially good reduction, as well as an explicitly given (although finite) set of places of guaranteed supersingular reduction. With a suitable condition on D (that it be "nontwisting") each PQM A/Q is of GL_2-type, i.e., predicted by the generalized Taniyama-Shimura conjecture to be modular. In short, a PQM surface is the "next closest thing" to an elliptic curve in several different senses, so is a natural object to study.

Having begun by emphasizing the analogies between PQM abelian surfaces and their moduli spaces $X_\bullet^{D+}(N)$ and elliptic curves and their moduli spaces $X_\bullet(N)$, we should now point out that there are some important differences between them that make the Diophantine geometry of Shimura curves (even) more interesting than that of the classical modular curves. First and foremost, Shimura curves do not have cusps. (Indeed, the fact that the moduli space of quaternionic abelian surfaces does not need to be compactified is equivalent to the potential good reduction cited above – there are no "generalized" quaternionic surfaces.) This makes the explicit theory of Shimura curves a serious computational challenge (e.g., we do not know an algorithm for computing equations of X^D). It also opens the possibility of ruling out the existence of global points by local means, the prototypical result in this direction being the analysis in [Jordan-Livné I]

of $X^D(Q_p)$; they show that there is always a prime p dividing D such that $X^D(Q_p) = \emptyset$. In our study of the plus quotient curves, we find that as we add level structure we begin acquiring local obstructions to the existence of rational points. On the other hand, basic finiteness conjectures concerning endomorphism rings of abelian varieties defined over Q imply that the loci of points on Shimura curves rational over number fields should be much more restricted, so we expect to find a very large supply of plus quotient Shimura curves violating the Hasse principle – i.e., having points rational over every completion of a number field but no points rational over the number field itself.

Here are the main results of this thesis:

- We explicitly determine the set of quaternionic discriminants D such that there exist infinitely many geometrically nonisomorphic $\mathcal{O}_D$-PQM abelian surfaces A/Q (Main Theorem 1).
- We show that the curve X^{D+} has points over every completion of Q (Main Theorem 2).
- We give a simple necessary and sufficient condition for the locus $X_0^{D+}(N)(Q_p)$ to be empty when p is a prime dividing D (Main Theorem 3).
- We show that for fixed D and imaginary quadratic K there is an absolute bound on primes $N \equiv 1$ modulo 4 such that there exists $P \in X_0^{D+}(N)(Q)$ whose preimage in $X_0^D(N)$ splits over K (Main Theorem 4).
- We show that there are infinitely many Shimura curves of the form $X_0^D(N)$ which violate the Hasse principle over suitable quadratic fields (Main Theorem 5).
- We show that rational torsion can be uniformly and effectively bounded on abelian varieties with potentially good reduction over *local* fields (Main Theorem 6).
- We give a short list of possible orders of the rational torsion subgroup of an abelian surface A/Q with everywhere potentially good reduction (Section 5.2) and a shorter list for PQM surfaces A/Q (Section 5.3).

The organization of the thesis

The study of Shimura curves as arithmetic objects requires background knowledge in a number of different areas (it is fair to describe it as the entire story of classical modular curves, plus a bit more), but there is no one reference for this background. I have made a (somewhat quixotic, I'm afraid) attempt to remedy this with a chapter containing detailed treatment of the following topics: the Brauer group and quaternion algebras over fields; the integral theory of orders and ideals in quaternion algebras over local and global fields; Shimura curves over C via uniformization by cocompact arithmetic Fuchsian groups; Shimura

curves as coarse moduli spaces for a moduli problem that can be formulated over C; over R; over Q; over $Z[1/ND]$; over $Z[1/D]$ and finally over Z. We do not claim any of the results (with the mild exception of Proposition 82, which is not used in the remainder of the thesis) as our own. Indeed, although we have tried to give an independent and coherent presentation of the material, some portions are essentially copied from the original references (our debt to [Vignéras], [Jordan I], [Ogg I] and [Buzzard] is especially clear).

It will be apparent soon enough that the length of this introductory chapter is comparable to that of the entire rest of the thesis, and we urge the reader who approaches this thesis with some knowledge of Shimura curves to start with Chapter 1 and refer to the background chapter as needed. Having said this, I wish this chapter were yet longer and more detailed; there are important topics missing, not least of which is an account of the result $J_0^D(N) \sim_Q J_0^{D-new}(DN)$, which for most of the number-theoretic community is the reason Shimura curves are studied. It was simply not possible, for reasons of both length and authorial knowledge, to give a reasonable account of this (it would involve a discussion of the Jacquet-Langlands correspondence between automorphic forms on GL_2 and automorphic forms on $B^\times$) in the present thesis. It has been suggested by my thesis adviser that this background material could form part of a "glorious monograph" on the arithmetic of Shimura curves – suffice it to say that there is some glory missing at present.

Chapter 1 is also foundational but contains material much more specialized to the topic at hand: we explore the notion of a *potentially* quaternionic abelian surface and explain why we treat the Atkin-Lehner quotients $X_\bullet^{D+}(N)$ as moduli spaces for such objects. In fact it is not literally true that these curves are coarse moduli spaces for the moduli problem of principally polarized abelian surfaces admitting geometric QM; the relationship between the locus of $\mathcal{O}_D$-QM abelian surfaces inside the full moduli space $\mathcal{A}_2$ and certain Atkin-Lehner quotients of X^D was investigated in [Jordan I] but completed (and corrected) by Victor Rotger, cf. [Rotger II-IV].

In Chapter 2 we classify those discriminants D such that there are infinitely many (geometrically distinct) $\mathcal{O}_D$-PQM abelian surfaces defined over Q. For this we need a certain Atkin-Lehner quotient of X^D to have infinitely many rational points, and it follows that a necessary condition on D is that it be sufficiently small so that these curves have genus zero or one. Using the genus formulae from Chapter 0 it is a straightforward matter to compute this set of D. However, there is an additional wrinkle which does not arise in the theory of elliptic curves: since the moduli space X^{D+} is coarse, a Q-rational point on this curve does not correspond canonically to a PQM abelian surface defined over Q. Moreover (unlike the elliptic modular case) it need not correspond to any PQM surface A/Q at all; in other words, there is a nontrivial obstruction to a PQM abelian surface being definable over its field of moduli. The interesting part of the proof is to show that this obstruction vanishes for infinitely many points on

the moduli space. The main idea is to show that being sufficiently p-adically close to a rationally defined CM point is enough to make the obstruction, which can be computed locally, vanish at p.

In Chapter 3 we study the Shimura curves X^{D+} and $X_0^{D+}(N)$ over p-adic fields. What we are in fact able to do is understand the Galois action on the supersingular points modulo p. In the case of p a prime of good reduction and $N = 1$, we find that that there is always an F_p-rational supersingular point, hence by Hensel's lemma $X^{D+}(Q_p)$ is nonempty. To tell the truth, this is exactly what happens in the (bad reduction) case of p dividing D as well, since modulo such a p *every* point is supersingular. But to get there we need to use the Cerednik-Drinfeld uniformization, which reduces the study of the special fibre to a "combinatorial" analysis of a certain finite graph. We also give a criterion for the nonemptiness of $X_0^D(N)(Q_N)$ in the case that N is prime.

In Chapter 4 we study rational points on $X_0^D(N)$ with values in number fields. Our main result is an adaptation of the methods of Mazur's [RI] to our context. Morally the proof is easier in our case than the classical one – due to the absence of cusps, we get the potential good reduction for free – but there are some additional complications due to the possible existence of points which cannot be defined over their field of moduli. The end of the proof again exploits the fact that modulo primes dividing D we necessarily have supersingular reduction. We also put together the local analysis from Chapter 3 with another global nonexistence theorem – which follows immediately from the "largeness" of the adelic Galois representation on a QM surface – to deduce infinitely many Hasse principle violations for $X_0^D(N)$ over certain imaginary quadratic fields. The final section of the chapter contains a technical result on the scarcity of Q-rationally defined cyclic subgroups of a PQM abelian surface A/Q.

In Chapter 5 we explore possible orders of rational torsion subgroups of PQM abelian surfaces A/Q. It turns out that we can get reasonable bounds in the more general context of abelian surfaces A/Q with everywhere potentially good reduction. Indeed one can even uniformly bound the rational torsion for an abelian variety defined over a *local* field in the case of potentially good reduction. The analysis of cyclic subgroups from Chapter 4 is used to show that the rational torsion which is prime to D is especially restricted.

Acknowledgements

I feel greatly honored to be able to add, in some small way, to a subject founded by so many mathematicians whose work I so deeply admire; among them are Barry Mazur, Bruce Jordan, Ron Livné and of course Goro Shimura, the richness of whose ideas seems undilutable by time.

I thank Noam Elkies and Ken Ribet for suggestions directly relevant to the

material appearing in this thesis. I thank David Savitt for helping me – quite a while ago now – with some results of Tate-Honda theory that appear in the appendix. I thank William Stein for helping me calculate a fistful of quaternionic modular forms (I regret that none of these calculations appear in this thesis) and for generally being so free with his technical and mathematical insights.

I am grateful to Victor Rotger for making his own as yet unpublished thesis work available to me.

The graduate students in the Harvard math department have been without exception intelligent and friendly, and it has been a pleasure to learn from them and with them over the years.

Thanks to Kara Kyung-wha Byun for providing comfort in the sad days of the new milennium. Her kindness will never be forgotten.

I am indebted to my thesis adviser, Barry Mazur, for more things than I can list here, but most recently for a careful reading of an early, ugly draft of a certain lengthy mathematical document.

My father would have been proud of this thesis beyond all moderation. My mother has long made and continues to make me proud of her strength in the face of all the challenges life has to offer. I thank them both with all of my heart.

Chapter 0

Background

0.1 Quaternion algebras as central simple algebras

In this section we review a portion of the theory of quaternion algebras which may be viewed as a special case of the theory of central simple algebras.

0.1.1 Central simple algebras and the Brauer group

Let F be a field. A *quaternion algebra* B/F is a four-dimensional central simple algebra (CSA) over F. Recall that a central simple algebra A/F is a finite-dimensional associative F-algebra with unit in which there are no two-sided ideals different from 0 and A, and such that the center of A is precisely $F = F.1$.

As a starting point, recall Wedderburn's theorem that every central simple algebra A/F is isomorphic to a matrix algebra over a division algebra. In this way, to any $A \in CSA(F)$ we associate the corresponding division algebra $D_A \in CSA(F)$. If $A, B \in CSA(F)$, we write $A \sim B$ if $D_A \cong D_B$ and say they are *similar*. One knows that if $A, B \in CSA(F)$, then $A \otimes B \in CSA(F)$, and if $A_1 \sim A_2, B_1 \sim B_2$, then $A_1 \otimes B_1 \sim A_2 \otimes B_2$. It follows that upon passage to similarity classes, the tensor product induces a composition law on the set of division algebras central over F. Indeed this is a group law: the identity element is the class of F itself, and the inverse to $[A]$ is $[A^{opp}]$ (opposite algebra), via the natural isomorphism $A \otimes A^{opp} \to End(A) = M_{[A:F]}(F)$ given by $(a \otimes b)(c) := acb$ (Since the left-hand side is a simple algebra, the map is an injective; by a dimension count it is an isomorphism.) We have just constructed the Brauer group of F, denoted $Br(F)$. The first thing to observe about it is:

Proposition 1 *The Brauer group of an algebraically closed field is trivial.*

Proof: Suppose that F is an algebraically closed field and D/F is a central division algebra. For $x \in D$, let $P(t)$ be the minimal polynomial of $x\cdot$ acting on D.

Since D has no zero divisors, $P(t) \in F[t]$ is irreducible. But F is algebraically closed, so P is linear and $x\cdot$ coincides with multiplication by an element of F.

Base change and splitting fields: Since every element of the Brauer group of F is represented by a unique division algebra, one might wonder why we bother with the generality of CSA's at all. One reason is that a CSA is a more robust notion than a division algebra: it is faithfully preserved under basechange. Namely, if E/F is a field extension and $A \in CSA(F)$, then $A \otimes_F E \in CSA(E)$ [Pierce, Prop. 12.4b]; and conversely, if A/F is any algebra such that $A \otimes_F E \in CSA(E)$ then $A \in CSA(F)$. This is not the case for division algebras, since we lose them by tensoring up to any algebraically closed field. This motivates the notion of a *splitting field*: if $A \in CSA(F)$, a splitting field for A is a field extension E/F such that $A \otimes E \cong M_n(E)$. For a given extension E/F, the classes of elements of $CSA(F)$ split by E form a subgroup of $Br(F)$ which we denote $Br(E/F)$; notice that by our proposition, if $\overline{F}$ is an algebraic closure of F, then $Br(F) = Br(\overline{F}/F)$. To see that this is a fruitful concept, notice that the fact that any $A \in CSA(F)$ has a splitting field implies that $\dim_F A$ is a perfect square: $\dim_F A = \dim_{\overline{F}} A \otimes \overline{F} = \dim_{\overline{F}} M_n(\overline{F}) = n^2$. If $A \in CSA(F)$, we define the *index* $Ind(A)$ of A to be $\sqrt{\dim_F D_A}$.

The study of splitting fields of a division algebra D is closely related to the study of subfields of D. A field E, $F \leq E \leq D$ is *maximal* if it is not properly contained in any other subfield of D.

Theorem 2 *A field extension E/F such that $[E : F] = Ind(D)$ splits D if and only if E can be embedded in D as an F-subalgebra.*

See [Pierce], Chapter 13.

Proposition 3 *A subfield $F \leq E \leq D$ is maximal (among commutative subfields of D) if and only if $[E : F] = Ind(D)$.*

Proposition 4 *If $F \leq E \leq D$ is maximal among separable field extensions of F, E is then a maximal subfield of D.*

One can also find the proofs of the last two propositions in any introductory text on associative algebras (e.g. [Pierce]). On the other hand, one finds in [Grothendieck] a different and more thematic approach: the idea is to define an analogous notion to maximal subfield which is stable under base extension and in so doing reduce to the study of matrix algebras. Let us sketch this briefly: if $A \in CSA(F)$, we consider instead of subfields the *étale subalgebras* $L \leq A$ (recall that an étale algebra over a field is just a finite product of separable field extensions.) In this context the key result is

Theorem 5 *Let $A \in CSA(k)$ be of rank r^2. Let L be a subalgebra of A. The following are equivalent:*
a) L is étale of rank r.
b) L is étale and equal to its own centralizer in A.

c) L is a maximal étale subalgebra of A.
d) There exists an isomorphism $\phi : A \otimes \overline{k} \to M_r(\overline{k})$ whose restriction to $L \otimes \overline{k}$ has image equal to the diagonal matrices in $M_r(\overline{k})$.

That d) implies a) implies b) implies c) is easy. The key is to show that a maximal étale subalgebra is self-centralizing, and this in turn quickly reduces to showing that if A/k is a central simple algebra which is not just k, then it contains a nontrivial étale k-algebra. But here we can do something slick: if k is finite, we will see later that the Brauer group of k is trivial, so the result that we want is obvious. So assuming that k is infinite, consider the reduced characteristic polynomial of a *variable* element x of A. Then, for x in a Zariski-open subset of A viewed as an affine space over k (i.e., away from the discriminant hypersurface), this polynomial will have distinct roots; hence for a sufficiently general element x, $k[x]$ gives a maximal étale subalgebra.

Taking the Galois closure of any separable splitting field, we immediately obtain the

Corollary 6 *Every $A \in CSA(F)$ is split by a finite Galois extension E/F, so that*

$$Br(F) = \lim_{E/F} Br(E/F).$$

Remark: We do *not* claim that D contains E as a subfield – nor could we. More on this later.

Taken together, these results could suggest to the reader who is familiar with Galois cohomology but new to the Brauer group that there ought to be a cohomological interpretation of $Br(F)$. For this we recall:

Principle 7 *(First principle of Galois descent): If X/F is an object over a field F, the collection of twisted forms*
$\mathcal{T}(X/F) := \{$objects X'/F such that $X'/\overline{F} \cong X/\overline{F}\}$
is isomorphic, as pointed set, to the cohomology set $H^1(G_F, Aut(X/\overline{F}))$.

For information on this principle – especially for a list of what kinds of "objects" for which it is valid – see [CL] and [CG]. By the results we have collected about splitting fields, it follows that the twisted forms of $M_n(F)/F$ are precisely the n^2-dimensional central simple algebras over F, so:

Corollary 8 *There is a natural bijection of pointed sets*

$$\{n^2 - \dim A \in CSA(F)\} \longrightarrow H^1(G_F, Aut(M_n(\overline{F}))\}.$$

One knows the automorphism group of a central simple algebra:

Theorem 9 *(Noether-Skolem) Let $A \in CSA(F)$, B a simple F-algebra. Any two F-algebra homomorphisms $B \hookrightarrow A$ are conjugate by an element of $A^\times$.*

From this we deduce immediately $Aut(M_n(\overline{F})) = PGL_n(\overline{F})$. Now, applying nonabelian Galois cohomology to the short exact sequence

$$1 \longrightarrow G_m \longrightarrow GL_n \longrightarrow PGL_n \longrightarrow 1$$

and recalling $H^1(F, GL_n) = 0$, we deduce a map $\Delta_n : H^1(F, PGL_n) \hookrightarrow H^2(F, G_m)$. Let E/F be a finite Galois extension, and denote by $A_n(E/F)$ the set of central simple algebras A/F such that $A \otimes E \cong M_n(E)$ and $A(E/F)$ the set of all *classes of* CSA's split by E. Composing Δ_n with our first-principle bijection we get an injective map $\delta_n : A_n(E/F) \to H^2(G_{E/F}, E^\times)$. From [CL] we find the following

Lemma 10 *If $n = [E : F]$, $\delta_n : A_n(E/F) \to H^2(G_{E/F}, E^\times)$ is surjective.*

Finally, we conclude

Proposition 11 *: The induced map $\delta : Br(E/F) \to H^2(G_{E/F}, E^\times)$ is a bijection of pointed sets. Hence also $\delta : Br(F) \to H^2(G_F, G_m)$ is an isomorphism of pointed sets.*

Thus we have an interpretation of $Br(F)$ in terms of Galois cohomology.

Define the cohomological index of a class $\eta \in H^2(G_F, G_m)$ to be the greatest common divisor of all the degrees of splitting fields for η (i.e., of field extensions E/F, not necessarily Galois, such that $\mathrm{Res}^{G_F}_{G_E}(\eta) = 0$.

Proposition 12 *The cohomogical index of a class η coincides with the index of the division algebra D associated to η.*

Proof: Let $i := \sqrt{dim_F D}$ be the index of D. By Proposition 3, D possesses a subfield $F \leq E \leq D$ with $[E : F] = i$ and which splits D. It follows that the cohomological index divides i. For the converse, let E/F be a splitting field for D such that $[E : F] = k$; by Lemma 10, there exists a CSA A/F of dimension k^2 such that $A \sim D$. But this implies that $A \cong M_n(D)$, so that $k = ni$ and i divides k.

In view of Proposition 11, we may unambiguously refer to the index of a Brauer group element, and this invariant may be interpreted both in terms of division algebras and by means of Galois cohomology.

0.1.2 Severi-Brauer varieties

One advantage of this Galois-descent method of identifying $Br(F)$ with $H^2(G_F, G_m)$ is that it provides a bijection from n^2-dimensional central simple algebras over F to the set of twisted forms of *any* object X/F such that $Aut(X/\overline{F}) = PGL_n$. As a key example, $Aut(P^n/\overline{F}) = PGL_{n+1}(\overline{F})$; it follows that $H^1(E/F, PGL_{n+1}(E))$ classifies algebraic varieties that are E/F-twisted forms of P^n_F: by definition, these are Severi-Brauer varieties.

The case $n = 2$ is already interesting: it gives a bijection of pointed sets
{smooth genus zero curves over F} $\longrightarrow$ {quaternion algebras over F}
– thus in order to classify conics over a nonalgebraically closed field we need to understand the Brauer group of the field. We will revisit Severi-Brauer conics from a more explicit point of view later in this section.

0.1.3 Crossed product algebras

The preceding identification of $Br(F)$ with $H^2(F, G_m)$ used $H^1(F, PGL_n)$ as an intermediary, a situation which we just exploited in order to make a connection with Severi-Brauer varieties. On the other hand, there is a more classical approach linking these first two objects directly: in this we will be able to in particular say that they are isomorphic as abelian groups (in fact, one can also see this via the nonabelian cohomology route; see [CL]). What we need is the notion of a crossed product algebra, which historically was a major motivation for Galois cohomology.

Let E/F be a finite Galois extension with $[E : F] = n$, and let A/F be an n^2- dimensional CSA containing E as a (necessarily maximal) subfield. From this data we shall construct a cocycle $\Phi \in Z^2(E/F, E^\times)$, as follows: by Noether-Skolem, each $\sigma \in G_{E/F}$ can be represented as conjugation by some $u_\sigma \in A^\times$: for all $e \in E$, $e^\sigma = u_\sigma^{-1} e u_\sigma$. One can easily check that $\{u_\sigma \mid \sigma \in G_{E/F}\}$ gives an E-basis for A. Moreover, setting $\Phi(\sigma, \tau) := (u_{\sigma\tau})^{-1} u_\sigma u_\tau \in E^\times$, we find that Φ satisfies the cocycle condition. Finally, we may take $u_1 = 1$ and then $\Phi(\sigma, 1) = \Phi(1, \sigma) = 1$ ($\sigma \in G_{E/F}$) – such a Φ is said to be normalized.

Conversely, given the data of a Galois extension E/F of degree n and a cocycle $\Phi \in Z^2(G_{E/F}, E^\times)$, we can construct a CSA A as follows: take

$$A := \bigoplus_{\sigma \in G} u_\sigma E$$

as E-vector space, where the u_σ are formal symbols. Define a product $\mu : A \times A \to A$ via

$$\mu(\sum_{\sigma \in G} u_\sigma c_\sigma, \sum_{\tau \in G} u_\tau d_\tau) := \sum_{\sigma, \tau} u_{\sigma\tau} \Phi(\sigma, \tau) c_\sigma^\tau d_\tau.$$

Proposition 13 *The algebra A constructed above is central simple over F, with maximal subfield E. Moreover the set u_σ represents the Galois action on E as in the first construction, and $u_{\sigma\tau}^{-1} u_\sigma u_\tau = \Phi(\sigma, \tau)$.*

This algebra $A = (E, G_{E/F}, \Phi)$ is called the crossed product algebra of E and G relative to Φ. We have the

Theorem 14 *If E/F is a finite Galois extension, then the mapping $[\Phi] \mapsto [(E, G_{E/F}, \Phi)]$ gives a group isomorphism $H^2(G_{E/F}, E^\times) \to Br(E/F)$.*

Remark: The preceding construction may seem to be the end of the story, but closer inspection reveals a subtlety: given a Galois extension E/F, it is the case that cohomologous cocycles $\Phi, \Psi \in Z^2$ yield isomorphic crossed product algebras, not merely similar ones. Thus, given an element $\eta \in Br(F)$ and a Galois splitting field E/F, the theorem constructs a unique representative of η (fair enough: E must be a maximal subfield, which determines the dimension of A). The question is: given η, can we always choose E such that the associated crossed product algebra is a division algebra? A moment's thought shows this to be equivalent to the question we raised earlier: does every division algebra D/F of index i have a Galois splitting field E/F of degree i? The answer is no; [Amitsur] constructed counterxamples. It remains an important unsolved problem to characterize which division algebras are crossed product algebras (on the other hand, we will see that division algebras over local and global fields are well-behaved enough so that this phenomenon does not arise).

0.1.4 The period-index problem (an advertisement)

We have seen that the index of a Brauer group element is an invariant which measures the size of the associated division algebra. Another measure of the size of $\eta \in Br(F)$ is simply its order in the Brauer group – we choose to call this its *period* (it is the period of the sequence $[D], [D \otimes D], [D^{\otimes 3}], \ldots$) It is not hard to see that the period divides the index – so that in particular $Br(F)$ is a torsion abelian group – and that the period and the index of a class share the same prime factors. In this level of generality, there is no more to say: given integers $a|b$ with the same prime divisors, there exists a field F and a division algebra D/F with period a and index b: indeed, one can take F to be a rational function field $C(t_1, \ldots, t_n)$ (the choice of n depends upon the discrepancy between the period and the index). We note in passing that this is just one example of a period-index problem in arithmetic geometry: if A/F is any commutative algebraic group, then it is interesting to ask about the relations between the period and the index (defined cohomologically as above) of a cocycle $\eta \in H^k(F, A)$. For a discussion of these ideas I heartily recommend the paper [Clark].

0.1.5 Finally, quaternion algebras

As we have said, a quaternion algebra B/F is a four-dimensional central simple algebra. We naturally distinguish between two kinds of quaternion algebras over F: the matrix algebra $M_2(F)$, which we call *split*, representing the trivial element of $Br(F)$, and a division algebra of index 2, *nonsplit*. Notice that quaternion algebras are characterized among Brauer group elements by having a quadratic splitting field (it follows that a nonsplit quaternion algebra has period two in the Brauer group, but the converse need not be true). The existence of a quadratic splitting field allows us to study quaternion algebras more explicitly than general CSA's.

Proposition 15 *Let F be a field whose characteristic is not 2. Let $a, b \in F^\times$.*

Then the F-algebra $(\frac{a,b}{F})$ generated by elements i, j and subject to the relations $i^2 = a, j^2 = b, ij = -ji$ is a quaternion algebra over F.

Proof: It wouldn't be too painful to prove this from scratch, but let's try to do something a little more insightful: I claim that $(\frac{a,b}{F}) \otimes F(\sqrt{a}) \cong M_2(F(\sqrt{a}))$. If so, $(\frac{a,b}{F})$ is a twisted form of $M_2(F)$, i.e., a quaternion algebra. Indeed, consider matrices

$$I = \begin{bmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{bmatrix}, \quad J = \begin{bmatrix} 0 & b \\ 1 & 0 \end{bmatrix};$$

they satisfy $I^2 = a, J^2 = b, IJ = -IJ$, so that they generate a subalgebra of $M_2(F(\sqrt{a}))$ isomorphic to $(\frac{a,b}{F}) \otimes F(\sqrt{a})$. But indeed I, I^2, J, IJ are $F(\sqrt{a})$-linearly independent matrices, so they generate the entire matrix algebra. Since the condition of being a quaternion algebra is faithfully preserved under base change, we're done.

Proposition 16 *When $char(F) \neq 2$, every quaternion algebra B/F is of the form $(\frac{a,b}{F})$ for suitable $a, b \in F^\times$.*

Proof: Choose a maximal subfield E of B, so that E/F is a quadratic extension (notice that such a subfield is given as the subalgebra generated by any nonscalar element of B). We may find $i \in E$ such that $i^2 \in F^\times$, say $i^2 = a$. By Noether-Skolem, the unique nontrivial automoprhism σ of E/F is represented as conjugation by some $u_\sigma \in B^\times : (e \in E)$, $e^\sigma = u_\sigma^{-1} e u_\sigma$. Since $\sigma^2(e) = u_\sigma^{-2} e u_\sigma^2 = e$, $u_\sigma^2 \in Z_B(E) = E$. Clearly $u_\sigma \notin E$, so $F \subseteq F[u_\sigma^2] \cap E \subseteq F[u_\sigma] \cap E = F$. Therefore $u_\sigma^2 \in F$, say $u_\sigma^2 = b$. By construction $u_\sigma^{-1} i u_\sigma = \sigma(i) = -i$, so we're done.

Remark: In characteristic 2, something similar can be done but with a slightly more complicated set of defining relations. We will not meet quaternion algebras over fields of positive characteristic in this thesis, so we refer the reader to [Vignéras].

Reduced trace, reduced norm, main involution: Let B/F be a CSA of dimension n^2. If $B = M_n(F)$, we have familiar maps $t : M_N(F) \to F$, $det : M_N(F) \to F$. For general B, by tensoring up to a splitting field (say $\overline{F}$), we get $B \hookrightarrow B \otimes \overline{F} \cong M_n(\overline{F})$, and via this embedding we can define maps $t : B \to \overline{F}, n : B \to \overline{F}$ called the reduced trace and the reduced norm (coming from the determinant) respectively. Moreover, these maps land in F: indeed, for $\sigma \in G_F$, twisting by σ gives another representation $r_\sigma : B \hookrightarrow B \otimes_\sigma \overline{F} \cong M_n(\overline{F})$ such that e.g. $n^\sigma = n_\sigma$. On the other hand, $M_n(\overline{F})$ has a unique n-dimensional representation up to isomorphism, so the twisted representation is conjugate to the original representation, whence $n = n_\sigma = n^\sigma$. Similarly for the trace (and indeed, for the entire characteristic polynomial).

Returning to the case of quaternion algebras, we also define the main involution $b \mapsto \overline{b} = t(b) - b$. If E/F is a quadratic subfield of B, then one can check

easily that the main involution stabilizes E and induces the unique nontrivial automorphism of E/F.

We can make all of this explicit: let $B = (\frac{a,b}{F})$. Then an arbitrary element of B may be written as $u = x.1 + y.i + z.j + w.ij$, where $x, y, z, w \in F$. Via the splitting representation exhibited in Proposition 14 above, we find $t(u) = 2x$, $n(u) = x^2 - ay^2 - bz^2 + abw^2$, and $\overline{u} = x.1 - y.i - z.j - w.ij$.

Observe that the norm is a quadratic form in the coefficients x, y, z, w. Now, being a quaternion algebra, B is either a division algebra or $M_2(F)$, and it will be the latter if and only if it has nonzero nilpotent elements, i.e., if and only if there exists $0 \neq u \in B$ such that $t(u) = n(u) = 0$. We get then that B is split if and only if the conic $ay^2 + bz^2 - abw^2 = 0$ has a nontrivial zero. Touching up the form of the equation a bit, we get:

Proposition 17 *A quaternion algebra $B = (\frac{a,b}{F})$ (char(F) $\neq 2$) is split if and only if the conic*

$$C/F : aX^2 + bY^2 - Z^2 = 0$$

is F-isomorphic to P^1.

We have in fact rediscovered Sever-Brauer conics in an explicit form:

Proposition 18 *The correspondence $(\frac{a,b}{F}) \mapsto C/F : aX^2 + bY^2 - Z^2 = 0$ gives the bijection between quaternion algebras over F and Severi-Brauer conics – i.e., smooth genus zero curves.*

Proof: Since the conic is constructed in terms of intrinsic properties of the quaternion algebra (the elements of norm zero on the trace zero subspace), it is clear that the mapping is well-defined at the level of quaternion algebras. Since any conic can be put in the exhibited form, the mapping is surjective. It remains to be seen that the conic determines the quaternion algebra up to isomorphism, i.e., that a quaternion algebra can be recovered from the norm form on its trace zero subspace. For this, note that $\langle h, k\rangle := tr(h\overline{k})$ is the associated bilinear form. Restricted to the trace zero subspace (say B_0 of B) it is given simply by $\langle h, k\rangle = -(hk + kh)$. It follows from this that two elements of B_0 anticommute if and only if they are orthogonal for the bilinear form. Now let $f : (B_0, \langle\rangle) \to (B_0', \langle\rangle)$ be an isometry of quadratic spaces, where $B = (\frac{a,b}{F})$. Then $i, j \in B_0$ and $-2f(i)^2 = \langle f(i), f(i)\rangle = \langle i, i\rangle = -2i^2 = -2a$, so that $f(i)^2 = a$; similarly, $f(j)^2 = b$. Also i, j anticommute, so they are orthogonal, so $f(i), f(j)$ are orthogonal, so they anticommute. It follows that $B' \cong (\frac{a,b}{F}) \cong B$.

0.1.6 Brauer group of a local or global field

Br(R): Indeed $H^2(R, G_m) = H^2(Z/2Z, C^\times) = \hat{H}^0(Z, 2Z, C^\times) = R^\times / N_{R^\times}^{C^\times}(C^\times) = Z/2Z$. The nontrival element is therefore a division quaternion algebra, and it is none other than Hamilton's quaternions $H := (\frac{-1,-1}{R})$.

Now let F be a non-Archimedean local field, i.e., a field complete with respect to a discrete valuation and with finite residue field. The computation of $Br(F)$ is one of the main steps in local class field theory; we content ourselves to recall the main results in a form which make the division algebras involved explicit. Write F_n/F for the unique degree n unramified extension of F, and let π be a uniformizer of F.

Returning momentarily to the case of an arbitrary field F, let E/F be a cyclic Galois extension of degree n; write $G_{E/F} = \langle\sigma\rangle$, and let $A \in CSA(F)$ contain E as a maximal subfield. Recall that we exploited the Noether-Skolem theorem to build a cocycle in $Z^2(G_{E/F}, G_m)$ representing A. Our assumption that $G_{E/F}$ is cyclic allows us to put the cocycle in an especially nice form: choose $u \in A^\times$ such that $\sigma(e) = u^{-1}eu$. Then for $1 \leq j < n$, we have $\sigma^j(e) = u^{-j}eu^j$, i.e., we get a cyclic E-basis $\{1, u, \ldots, u^{n-1}\}$ for A. Now $u^n \in Z_A(E) = E$; since $A = \bigoplus_{0 \leq j < n} u^j E$, we get $u^n \in Z_A(A) = F$. Put $a = u^n$. The corresponding cocycle Φ is just $\Phi(\sigma^i, \sigma^j) = 1$ if $i + j < n$, and a if $i + j \geq n$. We abbreviate $(E, \sigma, a) := (E, G_{E/F}, \Phi)$. We say that A is a cyclic algebra, and we may equally well view it as $E_\sigma[u]/(u^n - a)$ with the understanding that E acts σ-semilinearly: $eu^i = u^i\sigma^i(e)$.

Coming back to our local field F, the most important fact is that every element of $Br(F)$ is split by an unramified extension [CL]; this implies that every element of $Br(F)$ is represented by a cyclic algebra. We can exhibit an obvious family of cyclic algebras: $\mathrm{B}_{k/n} := (F_n, \sigma_n, \pi^k)$, where σ_n is the Frobenius of F_n/F. Now we have the

Theorem 19

a) The map $Q/Z \to Br(F)$, $\frac{k}{n} + Z \mapsto [B_{k/n}]$, is an isomorphism of groups.
b) Let $(k, n) = 1$. Then the index of $B_{k/n}$ is n. In particular the period and index coincide, and $B_{k/n}/F$, being an n^2-dimensional CSA of index n, is a division algebra.

In fact, the second statement follows from the first, since $n = \mathrm{period}(B_{k/n})$ divides $Ind(B_{k/n}) \leq \sqrt{\dim_F B_{k/n}} = n$. In particular, every division algebra over a local field is a crossed product algebra. Note also that since the period equals the index, quaternion algebras over F correspond to $Br(F)[2] = 1/2Z/Z$, so that there exists a unique nonsplit quaternion algebra over any local field. There is just one more thing to know about quaternion algebras over local fields:

Proposition 20 *Let B/F be a quaternion algebra and E/F any quadratic field extension. Then E splits B.*

Proof: Looking at the above construction of the unique division quaternion algebra $B_{\frac{1}{2}}/F$, we see that it is certainly split by the unramified quadratic extension F_2/F. Otherwise E/F is ramified and we can choose the uniformizer π to be the square of an element ρ of E, in which case the algebra $E_\sigma[u]/(u^2 - \pi)$ has

$u - \rho$ as a zero-divisor.

Global fields: Let F be a global field and Σ_F its set of places (including Archimedean places, if any). Let D/F be a division algebra. For each $v \in \Sigma_F$, the basechange $D \mapsto D_v := D \otimes_F F_v$ gives us a map $Br(F) \to Br(F_v)$. Now we can state another big

Theorem 21

a) We have an exact sequence $0 \to Br(F) \to \bigoplus_{v \in \Sigma_F} Br(F_v) \stackrel{\Sigma}{\to} Q/Z \to 0$. *Here, since each* $Br(F_v)$ *is either* $0, 1/2Z/Z$ *or* Q/Z, *there is a natural map* Σ *to* Q/Z, *"adding up the invariants."*
b) For any $A \in CSA(F)$, A *is a cyclic algebra (in particular a crossed product algebra), and the period equals the index.*

We remark that the first part is a cornerstone of global class field theory and the proof can be found in many places. For b), see [Pierce, Chapter 18].

Again we find that quaternion algebras over F correspond to 2-torsion elements in the Brauer group. Taking $F = Q$, we like to view $Br(Q)[2]$ as the subspace of the F_2- vector space on the set of of prime numbers together with ∞ given by formal sums with an even number of nonzero entries. The place ∞ plays a distinguished role in the theory (as we shall see in the next chapter): to prepare for this we say a quaternion algebra over Q is *indefinite* if it is split at ∞ and definite if it is ramified at ∞, i.e., $B \otimes R \cong H$. (To see why this terminology is used, consider the associated quadratic space $(B_0, \langle\rangle)$.) The *discriminant* of B is by definition the product over the finite ramified primes (we will see later that it is a discriminant in the sense of geometry of lattices). For the remainder of the thesis we shall reserve the letter D to denote the discriminant of a rational quaternion algebra. Clearly there exists a unique rational quaternion algebra with any given squarefree discriminant D, which will be indefinite or definite according to whether D has an even or odd number of prime factors.

Finally, we record a simple but indispensable criterion for a quaternion algebra over a global field to be split by a quadratic field extension:

Proposition 22 *(Hasse's criterion) Let* B/F *be a quaternion algebra over a global field. Let* K/F *be a quadratic field extension. Then* K *embeds in* B *as* F*-algebra if and only if for all* $v \in \Sigma_F$, $K_v = K \otimes_F F_v$ *embeds in* B_v.

Proof: Certainly the existence of a global embedding $K \hookrightarrow B$ implies, by tensoring up to F_v, the existence of all the local embeddings. Conversely, assume that for all $v \in \Sigma_F$, the quadratic F_v-algebra K_v embeds in B_v. We must show that K is a splitting field for B, or equivalently, that for all $w \in \Sigma_K$, $[B \otimes_F K_w] = 0$. There are two cases to consider: if w/v is split, then $K_v \cong F_v \oplus F_v$ has nontrivial idmpotents, so the assumption that it can be embedded in B_v implies that B_v is a matrix algebra, i.e., B_v was already split, and a fortiori $B \otimes K_w$

must be as well. If w/v is inert or ramified, then K_w/F_v is a quadratic extension of local fields, so by Proposition 19, $B \otimes_F K_w = B_v \otimes_{F_v} K_w$ is split.

When $F = Q$, this simplifies to:

Corollary 23 *Let B/Q be a rational quaternion algebra, and let K/Q be a quadratic field. If K is real and B is definite, K does not embed in B. Otherwise, K embeds in B if and only if for every prime p dividing the discriminant of B, p is nonsplit in K.*

0.2 Orders and ideals in quaternion algebras

In this section we summarize the basic theory of orders ideals in a quaternion algebra (we concentrate on the case where F is a local or global field). The most important results, namely class number formulas and formulas counting the number of embeddings of a quadratic order, are due to Eichler. The canonical modern presentation of this material is [Vignéras].

0.2.1 Basic theory of orders and ideals

Let H/K be a quaternion algebra, and let R be a Dedekind ring with quotient field K, considered fixed for the following discussion (imagine $R = \mathcal{O}_K$ when K is local/global). Viewing H as K-vector space, we have the notion of a (complete) R-lattice $L \leq H$, i.e., a finitely generated R-submodule such that $L \otimes_R K = H$. An element $x \in H$ is said to be integral (with respect to R) if $R[x]$ is a finitely generated R-module (i.e., the same definition as in the commutative case). It is not hard to see that an equivalent characterization of integrality is $t(x), n(x) \in R$ (use the reduced characteristic polynomial).

What *is* different from the commutative case, and in some sense makes the non-commutative theory of orders and ideals correspondingly more involved, is that the set of integral elements of H need not form a ring: indeed in the algebra $M_2(Q)$, the two elements

$$A = \begin{bmatrix} \frac{1}{2} & -3 \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & \frac{1}{5} \\ 5 & 0 \end{bmatrix}$$

are both integral, but neither $A + B$ nor AB is integral.

An *ideal* of H is just an R-sublattice (analogous to a fractional ideal in the commutative case). An *order* $\mathcal{O} \leq H$ is an ideal which is a subring. Equivalently, $\mathcal{O}$ is a ring of integral elements generating H over K. A maximal order is indeed an order which is not properly contained in any other order. An Eichler order is the intersection of two maximal orders. I claim that any integral element

$x \in H$ lies in a maximal order. Indeed, we may assume that x is not in R (otherwise it lies in every maximal order), so that $K(x) = L$ is a quadratic extension in which $R[x] = \mathcal{O}'$ is an R-order in the commutative sense. So $x = a + \sqrt{c}$, with $a, c \in R$, and clearly the same quadratic R-order is generated by $\sqrt{c}$, so it is enough to construct a maximal order containing $x := \sqrt{c}$. Now let y be an integral Noether-Skolem element: $y^{-1}xy = \overline{x}$ (the choice of such a y is unique up to a scalar from K, and of course some multiple of any given element is integral). From the theory of bases of quaternion algebras recalled in the previous section, we see that R submodule generated by x and y is $R + Rx + Ry + Rxy$; in particular it is finitely generated, so we have constructed an R-order. It need not be maximal, so we must also show that every order is contained in a maximal order. One can see this by applying Zorn's Lemma to a chain of orders: the union is a subring consisting of integral elements, hence gives an upper bound for the chain. However, it seems inelegant to apply appeal to Zorn's lemma in such a "finite" situation, so a little later we will give an alternate proof of the existence of maximal orders using the discriminant.

If $I \leq H$ is an ideal, we define the associated left and right orders: $\mathcal{O}_l(I) := \{h \in H \mid hI \subseteq I\}, \mathcal{O}_r(I) := \{h \in H \mid Ih \subseteq I\}$. We say an ideal is integral if it is contained in its left and right orders (i.e., if $II \subseteq I$) – this clarifies the terminology, since an integral ideal really is a left $\mathcal{O}_l(I)$-ideal and a right $\mathcal{O}_r(I)$-ideal. We also say an ideal is two-sided if $\mathcal{O}_l(I) = \mathcal{O}_r(I)$. An ideal is principal if $I = \mathcal{O}_l h = h'\mathcal{O}_d$. If I, J are ideals, their product IJ (defined in the expected way as the collection of finite sums of $i.j$) is an ideal. Since we have a Dedekind ring, we should be able to invert ideals: define $I^{-1} := \{h \in H \mid IhI \subseteq I\}$. We have by definition $II^{-1} \subseteq \mathcal{O}_l(I)$, $I^{-1}I \subseteq \mathcal{O}_r(I)$, with equality if I is principal. Indeed, we even have equality if I is *locally principal*; since this is the case for all ideals when K is a local or global field (as we'll see), for our purposes we will always have equality. Moreover, we also have $\mathcal{O}_l(IJ) = \mathcal{O}_l(I), \mathcal{O}_r(IJ) = \mathcal{O}_r(J)$ for locally principal ideals, so similarly we may make use of these identities in cases of interest.

Ideal classes: We say two ideals I, J are equivalent on the right (resp. on the left) if $I = Jh$, (resp. $I = hJ$) for some $h \in H$. If $\mathcal{O}$ is an order, we define the set $Pic_l(\mathcal{O})$ of left-ideal classes of $\mathcal{O}$: this is the set of ideals with *right* order $\mathcal{O}$ modulo equivalence on the left. (Note that we have to do it this way: modifying an ideal on the left does not change its right order.) We may similarly define $Pic_r(\mathcal{O})$ of right-classes of left $\mathcal{O}$-ideals. Indeed we can make do with either one of these objects, since the map $I \to I^{-1}$ induces a bijection $Pic_l(\mathcal{O}) \to Pic_r(\mathcal{O})$. Let $\mathcal{O}$, $\mathcal{O}'$ be two orders with the property that there exists an ideal I such that $\mathcal{O} = \mathcal{O}_l(I)$, $\mathcal{O}' = \mathcal{O}_r(I)$; we say $\mathcal{O}$, $\mathcal{O}'$ are *linked*. Note that any two maximal orders are linked: just take $I = \mathcal{O}.\mathcal{O}'$.

Lemma 24 *Linked orders have the same number of (left or right) ideal classes.*

Proof: We define a map from the set of left $\mathcal{O}$-ideals to the set of right $\mathcal{O}'$-ideals by $J \mapsto J^{-1}I$. The map $P \mapsto \mathrm{IP}^{-1}$ provides an inverse. Moreover, the map

preserves ideal classes, since $Jh \mapsto (Jh)^{-1}I = h^{-1}J^{-1}I$.

In view of this lemma, we may define the class number of H/K to be the number of (either left or right) ideal classes of any maximal order.

Order types: We say two orders are of the same type if one can be conjugated to the other by an element of H. We have the following technical result:

Lemma 25 *The following are equivalent:*

a) Two orders $\mathcal{O}$, $\mathcal{O}'$ are of the same type.

b) There exists a principal ideal I linking $\mathcal{O}$ and $\mathcal{O}'$.

Let us define the type number of the class of all orders linked to a given order to be the number of types of ideals in this class. The type number of H/K is defined to be the number of types of maximal orders. The previous lemma immediately gives that the type number is less than or equal to the class number. One of the main results we are going for here is the computation of the type number and the class number for quaternion algebras over local and global fields.

The discriminant: Happily, the theory of the discriminants works just as in the commutative case: indeed, let I be an ideal of H. We define $n(I)$ to be the fractional ideal of R generated by the reduced norms of the elements of I. Now the different of an order $\mathcal{O}$ is the inverse dual of $\mathcal{O}$ for the trace form: $D(\mathcal{O}) = (\mathcal{O}^\star)^{-1}$, where $\mathcal{O}^\star := \{x \in H \mid t(x\mathcal{O}) \subseteq R\}$. We define the discriminant $\Delta(\mathcal{O})$ as the norm of the different ideal. We have as in the commutative case the useful fact that if $\mathcal{O}$ is a free R-module with basis v_i, then $\Delta(\mathcal{O})^2 = R(\det(t(v_i v_j))$, as well as the fact that Δ can be computed locally on R.

Proposition 26 *Let $\mathcal{O} \leq \mathcal{O}$ be two orders. Then $\Delta(\mathcal{O}') \subseteq \Delta(\mathcal{O})$, with equality if and only if $\mathcal{O}' = \mathcal{O}$.*

This proposition gives a "more geometric" proof that every order is contained in a maximal order: R being a Dedekind ring, is Noetherian!

As examples, consider the order $M_2(R) \leq M_2(K)$. We find that the discriminant ideal is R itself, which implies that $M_2(R)$ is a maximal order. Now take $K = Q$ and $H = (\frac{-1,-1}{Q})$. The order associated to the integral basis is $\mathcal{O} := Z[1, i, j, ij]$: we find that its discriminant is $4Z$, which is not a maximal ideal of Z. Indeed $\mathcal{O}$ is not a maximal order: it is contained in $\mathcal{O}' := Z[1, i, j, \frac{1+i+j+ij}{2}]$ (and one should check that this is actually an order) of discriminant $2Z$. Notice that 2 is the discriminant of this definite rational quaternion algebra in the Brauer group sense of the previous section. We will see that indeed the discriminant of any maximal order of a quaternion algebra over Q is its Brauer group discriminant.

Optimally embedded quadratic orders: In some sense, a quaternion algebra over a field K is a bunch of quadratic extensions glued together in a non-commutative

way. Earlier, we have seen the importance of quadratic subfields (= quadratic splitting fields) of a quaternion algebra H/K. A large part of the integral theory of quaternion algebras concerns the relation of orders $\mathcal{O}$ of the quaternion algebra to orders S of a quadratic splitting field L for H/K. Let L/K be a quadratic extension splitting H, so that L embeds into H as K-algebra. There will be many such embeddings: by the Noether-Skolem theorem, any two will be conjugate by an element of $H^\times$; since L is a maximal commutative subalgebra of H, by orbit-stabilizer considerations we see that the set of K-embeddings of L into H is $B^\times/L^\times$. But now fix an order $\mathcal{O}$ and let $\iota : L \hookrightarrow H$ be an embedding. We have the notion of the associated optimally embedded quadratic order, namely $S := \iota^{-1}(\mathcal{O} \cap \iota(L))$. In other words, S is an order of L, and the embedding ι has the property that it carries S into $\mathcal{O}$ and does so for no larger order of L. Let $N = N(\mathcal{O}) \leq B^\times/K^\times$ be the subgroup of automorphisms preserving $\mathcal{O}$ (i.e., the normalizer of $\mathcal{O}$). It is immediate that the condition of an embedding $\iota : L \hookrightarrow H$ being S-optimal is stable under N. Notice that it may not be clear a priori what this group N is, but certainly it contains $\mathcal{O}^\times$ as a subgroup. In general, if $G \leq N$, we write $v_G(S, \mathcal{O})$ for the number of G-orbits of optimal embeddings $S \hookrightarrow \mathcal{O}$ (possibly infinite, in this level of generality); we write $v(S)$ when $N = \mathcal{O}^\times$ for some $\mathcal{O}$ that is understood to be fixed. When K is a global field, we will see that $v(S)$ is indeed finite, and give a product formula involving terms from the local places and a "global contribution" – the class number of S.

0.2.2 Local fields

In this subsection K shall always be a non-Archimedean local field, and all orders and ideals are taken with respect to $R := \mathcal{O}_K$. We write π for an arbitrary, but fixed, uniformizer of K.

Split case: we give ourselves V/K a two-dimensional vector space.

Proposition 27

a) The maximal orders of $End(V)$ are the rings $End(L)$, where L is a complete R-lattice of V.

b) The ideals of these maximal orders are $Hom(L, M)$, L, M complete lattices of V.

Theorem 28

a) The maximal orders of $M_2(K)$ form a single type: all conjugate to $M_2(R)$.

b) The two-sided ideals of $M_2(R)$ form a cyclic group generated by the prime ideal $P = M_2(R)\pi$.

c) The integral left $M_2(R)$ ideals are the distinct ideals $M_2(R)\begin{bmatrix} \pi^n & r \\ 0 & \pi^m \end{bmatrix}$, where n, m are non-negative integers and r runs through a set of coset representatives of $R/\pi^m R$ in R.

Definition: Let $\mathcal{O} = End(L)$, $\mathcal{O}' = End(M)$ be two maximal orders of $End(V)$. If $x, y \in K^\times$, notice that $End(Lx) = \mathcal{O}$, $End(My) = \mathcal{O}'$, i.e., the maximal order depends only on the lattice up to homothety. Therefore, by after rescaling, the theory of elementary divisors furnishes us with a basis (f_1, f_2) of L such that $(\pi^a f_1, \pi^b f_2)$ is a basis for M. The integer $|b-a|$ is visibly independent of the scaling. We define the distance between two maximal orders $\mathcal{O}, \mathcal{O}'$ to be this quantity $|b-a|$. As an example, the distance between $M_2(R)$ and the order $\begin{bmatrix} R & \pi^{-n}R \\ \pi^n R & R \end{bmatrix}$ is n. We define an Eichler order of (local) level n to be an order obtained by intersecting two maximal orders of distance n. We have the following

Lemma 29 *Let $\mathcal{O} \leq M_2(K)$ be an order. The following are equivalent:*

a) There exists a unique pair of maximal orders $\mathcal{O}_1, \mathcal{O}_2$ such that $\mathcal{O} = \mathcal{O}_1 \cap \mathcal{O}_2$.

b) $\mathcal{O}$ is an Eichler order.

c) There exists a unique nonnegative integer n such that $\mathcal{O}$ is conjugate to $\mathcal{O}_n := \begin{bmatrix} R & R \\ \pi^n R & R \end{bmatrix}$.

In summary, for H/K the split quaternion algebra over a local field: there exist infinitely many maximal orders but they are all conjugate; the class number is one; and there is a unique type of Eichler order of any given level (in fact Eichler orders of distinct levels are not even linked, so this is the strongest possible statement along these lines.

Remark: We note in passing that the maximal orders of a split quaternion algebra are in bijection to the homogeneous space $GL_2(K)/Stab(M_2(\mathcal{O}_K)) = GL_2(K)/K^\times GL_2(\mathcal{O}_K) = PGL_2(K)/PGL_2(\mathcal{O}_K)$. Taking this as the vertex set of a graph and decreeing two vertices to be adjacent if they have distance 1, we recover the Bruhat-Tits tree associated to $PGL_2(K)$.

Definition (Eichler symbol): Let L/K be a separable quadratic field extension, π a uniformizer of K. The Eichler symbol $(\frac{S}{\pi})$ is defined as follows: if S is not the maximal order of L, then $(\frac{S}{\pi}) := 1$. Otherwise, it coincides with the Artin symbol $(\frac{L}{\pi})$ – i.e., is -1 if the extension is unramified (inert!) and 0 if the extension is ramified.

Theorem 30 *(Optimal embedding theorem, split local case) Let K be a local field, L/K be an étale quadratic algebra (i.e., a separable field extension or $K \oplus K$), S an order of L, and $\mathcal{O} \leq M_2(K)$ a maximal order. Then $v(S, \mathcal{O}) = 1$. If $\mathcal{O}'$ is instead a level π-Eichler order, then $v(S, \mathcal{O}) = 1 + (\frac{S}{\pi})$: in particular, S can be embedded in $\mathcal{O}'$ unless S is the maximal order and L/K is unramified.*

The reader is invited to consult [Vignéras, pp. 44-47] for a proof of this theorem.

Nonsplit case: Now let H/K be the (unique, up to isomorphism) division quaternion algebra over the local field K. The theory of valuations extends to this non-commutative setting to give very nice results, namely: define a map $v : H^\times \to Z$ by $v(x) := v(n(x))$ for all $x \in H^\times$, $v(0) := \infty$. Obviously v gives a group homomorphism with the property that $v'|_K = 2v$; moreover it is surjective ($v(\sqrt{\pi}) = 1$), and one easily checks that it has the property that $v(x+y) \geq inf(v(x), v(y))$, i.e., it gives a valuation on H. In particular, the set $\mathcal{O}_K$ of elements whose valuation is nonnegative is a subring of H. Moreover, since the valuation restricted to the quadratic field generated by any given noncentral element coincides with the natural (prolonged) valuation of this quadratic extension of local fields, every element of the valuation ring is integral in our sense. Conversely every element even with norm in $\mathcal{O}_K$ lies in the valuation ring; it follows that $\mathcal{O}_K$ is the ring of all integral elements, i.e., it is the unique maximal order. Every integral ideal of $\mathcal{O}_K$ is therefore twosided. Indeed, $\mathcal{O}_K$ has a unique maximal ideal P, namely the elements of positive valuation, and the complete set of ideals is $\{P^i\}_{i\geq 1}$. In particular the class number is one.

Theorem 31 *(Optimal embedding theorem, nonsplit local case) Let K be a local field, L/K an étale quadratic algebra, H/K the (unique) division quaternion algebra, $\mathcal{O} \leq H$ the maximal order. If $S \leq L$ is a maximal order, $v(S, \mathcal{O}) = 1 - (\frac{L}{\pi})$. If S is not maximal, $v(S, \mathcal{O}) = 0$.*

Proof: Clearly we may assume L is a field, otherwise it does not embed in the nonsplit quaternion algebra H, and conversely, we saw in Section 1.1 that any quadratic local extension L does embed in H. So let $\iota : LH$ be an embedding. If S is the maximal order, then clearly ι embeds it inside H as a subring of integral elements, therefore it is contained in the unique maximal order $\mathcal{O}$. Since S is maximal, this embedding is automatically optimal; this shows $v(S, \mathcal{O}) \geq 1$. This argument simultaneously shows that if S is not maximal, no embedding into $\mathcal{O}$ is optimal. Now, if u is any uniformizer of $\mathcal{O}$ (i.e., an element of valuation 1) we have $H^\times/K^\times = \mathcal{O}^\times/K^\times \cup \mathcal{O}^\times u/K^\times$, so that $v(S, \mathcal{O})$ is either 1 or 2. In general, it is easy to see that if $S = R[u]$, $v_G(S, \mathcal{O})$ is equal to the number of G-conjugacy classes in H of the element u. It follows that $v_(S, \mathcal{O}) = 1$ if and only if there exists an element $u \in L$ with $w(u) = v(n(u)) = 1$, i.e., if and only if L/K is ramified.

0.2.3 Global fields

Now H/K is a quaternion algebra over a number field and all orders and ideals are with respect to $R := \mathcal{O}_K$. We are going to explain, briefly, how the theory of adelic points on the algebraic group $G := H^\times$, leads to the theorems on class numbers, type numbers, and embedding numbers. We are going to assume as

given the results recalled on $Br(K)[2]$ in Section 1.1 (to see how they could be proven using these adelic methods, consult [Vignéras, Ch. III].

Theorem 32 *(Norm theorem) Let K_H be the set of elements of K which are non-negative at every ramified real place of H. Then $K_H = n(H)$. In particular, if H is totally indefinite (= split at every real place of K), $n(H) = K$.*

Proof: [Vignéras, p. 80].

Theorem 33 *(Vignéras' "fundamental theorem") Let v be an infinite place of K. Then there exists a compact subset $C \subset H$ such that $H_K^\times H_v^\times C$ is dense in H.*

Proof: [Vignéras, pp. 62-63].

Theorem 34 *(Strong approximation for H^1) Let H^1/Q be the group of norm 1 quaternions. Let S be a set of places of K containing at least one Archimedean place. Write $H_S^1 = \Pi_{v \in S} H_v^1$. Then if H_S^1 is* not *compact, $H_K^1 H_S^1$ is dense in H_A^1.*

Proof: [Vignéras, p. 81].

Let us say that H/K is not totally definite (ntd) if there exists a real place of K at which H is split (en [Vignéras], c'est le Condition Eichler). We find at this point a parting of the ways: the answers to our questions about class numbers and type numbers depend very much on whether H/K is totally definite or not. Notice that in the strong approximation theorem, if H if ntd we can take S to be a split real place, H_v^1 is compact if and only if H_v is nonsplit.

Local properties of orders and ideals: fix X a complete lattice in H. Recall that there exists a bijection between the set of all complete lattices in H and sets of data $\{Y_v\}$, where for each NA place v of K, Y_v is a lattice in H_v and $H_v = X_v$ for almost every v. Indeed, we just send a lattice Y to its family of local lattices Y_v. In other words, a lattice is determined by all of its localizations, and we can get from any lattice to any other lattice by modifying finitely many localizations. We thus have the notion of a local property of ideals, i.e., one that holds for Y if and only if it holds for all Y_v. Examples of local properties of ideals and orders in H: being an order, being a maximal order, being an Eichler order, being an integral ideal, being a two-sided ideal. Let us remark that we can define the *level* of an Eichler order locally, to be the integral R-ideal which at each place v is given by the ideal $\mathcal{P}_v^k$, where $\mathcal{O}_v$ is of level π_v^k.

Discriminants: It is easy to see that if I is an ideal of H, then $n(I_v) = n(I)_v$, and if $\mathcal{O} \leq H$ is an order, $\delta(\mathcal{O}_v) = \delta(\mathcal{O})_v$, i.e., norms and discriminants can be computed locally. Whence immediately the important:

Proposition 35 *An order $\mathcal{O}$ is maximal if and only if its discriminant is equal to the discriminant of H in the sense of the Brauer group – the product over all the finite ramified places of H.*

For example, the order $Z[1, i, j, \frac{1}{2}(1 + i + j + ij)]$ in $(\frac{-1,-1}{Q})$ had discriminant $2Z$, so must indeed be a maximal order. More generally, one can use this proposition to write down a level N-Eichler order in a quaternion algebra over Q of arbitrary discriminant D (an exercise which we do not find a need for here).

Adelization: We find that the sets of orders and ideals are interested in can be represented as adelic (possibly double) coset spaces: let $\mathcal{O}$ be a level N-Eichler order in H/K. Then:

left $\mathcal{O}$-ideals correspond to $\mathcal{O}_A^\times \backslash H_A^\times$;
two-sided $\mathcal{O}$-ideals correspond to $\mathcal{O}_A^\times \backslash N(\mathcal{O}_A)$;
level N-Eichler orders correspond to $N(\mathcal{O}_A)\backslash H_A^\times$;
$Pic_r(\mathcal{O})$ corresponds to $\mathcal{O}_A^\times \backslash H_A^\times / H_K^\times$;
types of level N-Eichler orders correspond to $H_K^\times \backslash H_A^\times / N(\mathcal{O}_A)$.

In analogy to the commutative case, it thus becomes plausible that these sets will be finite and related to the class number of K (i.e., to the cardinality of $\mathcal{O}_K^\times(A)\backslash K_A/K^\times$). Here is a result which is of the highest importance for us:

Theorem 36 *For a maximal order $\mathcal{O}$ in H/K, we have $N(\mathcal{O})/K^\times\mathcal{O}^\times \cong (Z/2Z)^r$, where r is the number of ramified primes of H.*

Proof: Indeed $N(\mathcal{O})$ is the set of $x \in H$ such that $x \in N(\mathcal{O}_v)$ for all NA v. Since $N(M_2(\mathcal{O}_v)) = K_v^\times GL_2(\mathcal{O}_v)$, we find no contribution from the split places. On the other hand, if v is a ramified place, the uniqueness of the maximal order implies $N(\mathcal{O}_v) = H^\times$, so $N(\mathcal{O}_v)/K_v^\times\mathcal{O}_v^\times = H_v^\times/K_v^\times\mathcal{O}_v^\times \cong Z/2Z$.

Remark: A similar result could be worked out for Eichler orders – again $N(\mathcal{O})/K^\times\mathcal{O}^\times$ will be an elementary 2-group.

Theorem 37 *Let $\mathcal{O} \leq H$ be an order in H. Then $Pic_l(\mathcal{O})$ is finite. It follows that the number of types of Eichler orders of any given level is finite.*

Proof: This is indeed immediate from the "fundamental theorem": we are looking at the double coset space $\mathcal{O}_A^\times \backslash H_A^\times / H_K^\times$. Fixing any infinite place v, we have $\mathcal{O}_A^\times \backslash H_A^\times = \mathcal{O}_A^\times \backslash H_K^\times H_v^\times C$, for some compact C. Since v is infinite, we have $H_v^\times \subset \mathcal{O}_A^\times$ and we see that our discrete quotient is the image of a compact set.

Eichler's theorem: We assume now that H/K is not totally definite.

Definition: Let P_H be the subgroup of $Frac(K)$ of principal ideals generated by an element of K_H, i.e., which are positive at all the real ramified places of H. Let h_H be the cardinality of $Frac(K)/P_H$ – i.e., h_H lies somewhere between the class number and the narrow class number of K and coincides with the former when H is totally indefinite.

Theorem 38 *(Eichler) Let $\mathcal{O}$ be an Eichler order of the ntd quaternion algebra H/K. The reduced norm map induces a bijection $n : Pic_r(\mathcal{O}) \to Frac(K)/P_H$. In particular, the class number of H is h_H.*

Corollary 39 *The class number of an indefinite rational quaternion algebra is 1. Therefore, it is moreover the case that all maximal orders in an indefinite rational quaternion algebra are conjugate.*

Proof of Eichler's theorem: The reduced norm induces a map

$$\mathcal{O}_A^\times \backslash H_A^\times / H_K \to R_A^\times \backslash K_A^\times / K_H$$

For the surjectivity, observe that unless v is an infinite ramified place, $n(H_v^\times) = K_v^\times$, and otherwise $K_v^\times \subset R_A^\times$. For the injectivity, $H_A^1 \subset \mathcal{O}_A^\times H_K^\times$ by strong approximation.

Remarks: We do not need it here, but Vignéras proves an interesting relation between the type number and the class number for an ntd H/K. Indeed, let $\mathcal{O}$ be an Eichler order of level N, and write $h := Pic_r(\mathcal{O})$, t for the type number of $\mathcal{O}$ and h_2 for the number of classes of two-sided $\mathcal{O}$-ideals. Then $h = th_2$, and h_2 is equal to the order of the subgroup of $Frac(K)/P_H$ generated by: squares of ideals of R, prime ideals ramifying in H and prime ideals dividing the level to an odd power. We deduce:

Corollary 40 *If h_H is odd, there is a unique conjugacy class of Eichler order of any given level.*

The relevance of this corollary is that the general optimal embedding formulas for $v_G(S, \mathcal{O})$ are rather unwieldy. In fact in the remainder of the thesis we meet only quaternion algebras over Q (class number one!), but we aspire to give the general set-up in a form which makes possible the generalization to other totally real fields. It turns out that when there is a unique type of Eichler order we get nicer embedding formulas, so this seems like a fair compromise.

Theorem 41 *(Optimal embedding theorem, global ntd case) Let H/K be an ntd quaternion algebra over a number field K, and assume there exists a unique conjugacy class of Eichler order of level N; let $\mathcal{O}$ be a representative of this class. Then $v(S, \mathcal{O}) = h(S)\Pi v_p(S_p, \mathcal{O}_p)$, the product extending over all NA places; here $h(S)$ is the class number of the quadratic order S.*

Proof: [Vignéras, pp. 92-94].

Corollary 42 *Let $\mathcal{O}$ be an Eichler order of squarefree level N in a quaternion algebra of discriminant D. Then*

$$v(S, \mathcal{O}) = h(S) \prod_{p|D} (1 - (\frac{S}{p})) \prod_{q|N} (1 + (\frac{S}{p})).$$

0.3 Quaternionic Shimura varieties over C

Let F be a totally real number field, B/F a quaternion algebra, $\mathcal{O} \leq B$ an Eichler order of squarefree level $\mathcal{N}$. Let $r = [F : Q]$. We may order the real places

of F $\infty_1, \ldots \infty_r$ such that $B \otimes_{\infty_i} R \cong M_2(R)$ for $1 \leq i \leq g$ and $B \otimes_{\infty_i} R \cong H$ for $g+1 \leq i \leq r$. We say that B/F is of type $(g, r-g)$. If $g = r$ it is totally indefinite; if $g = 0$ it is totally definite.

Compiling all the real embeddings at the split places, we get a map $\phi : B \to \prod_{i=1}^{g} M_2(R)$. Write $\mathcal{O}^+$ for the group of units of $\mathcal{O}$ of totally positive reduced norm and $\mathcal{O}^1 \leq \mathcal{O}^+$ the group of units of reduced norm 1. By restriction, we get maps

$$\varphi : \mathcal{O}^+ \hookrightarrow \prod_{i=1}^{g} GL_2^+(R),$$

$$\varphi : \mathcal{O}^1 \hookrightarrow \prod_{i=1}^{g} SL_2(R)$$

.

Proposition 43

a) $\phi(\mathcal{O}^1)$ is a discrete subgroup of $SL_2(R)^g$ of finite covolume.

b) If H/F is nonsplit, $\mathcal{O}^1 \leq SL_2(R)^g$ is cocompact, and $S_{\mathcal{O}} := \mathcal{O}^1 \backslash SL_2(R)^g$ has the natural structure of a projective C-variety.

c) If $g = r$, we have the double coset interpretation

$$\mathcal{O}^1 \backslash SL_2(R)^g \cong B^\times \backslash B^\times(A) / K_\infty \widehat{\mathcal{O}^\times}$$

where $K_\infty = C^g \subset M_2(R)^g$.

Comments: Parts a) and b) are classical in a very strong sense (special cases were known to Fricke and Poincaré). A nice, rather elementary presentation of these results (when g = 1) from the perspective of arithmetic Fuchsian groups can be found in [Katok]. For the general case see [Vignéras, p.104] – the proof given there uses the notion of strong approximation for the group $B^\times$. We remark that the same methods yield a more general compactness result for the C-points of a Shimura variety associated to a reductive algebraic group G/Q (and certain additional data):

Proposition 44 *Let G/Q be a reductive group, and let S_G be any associated Shimura variety. Then $S_G(C)$ is compact if and only if the derived subgroup $[G, G]$ is anisotropic over Q.*

Once we have discussed the moduli interpretation, c) is easily to see directly – it is the sort of double-coset construction that is ubiquitous in the theory of Shimura varieties. The adelic perspective allows us a generalization: if $K \leq \widehat{\mathcal{O}^\times}$ is a compact open subgroup, we put

$$S_K := B^\times \backslash B^\times(A) / K_\infty K,$$

a quaternionic Shimura variety with level-K-structure.

There is already a certain redundancy in the objects we have introduced: starting with a maximal order, we can recover the Shimura curve associated to a level $\mathcal{N}$ Eichler order by taking $K = \widehat{\Gamma_0(\mathcal{N})}$. We will find both perspectives useful: the notion $\Gamma_0(\mathcal{N})$-level structure seems more familiar and leads directly to the moduli problem we want to study, but on the other hand we can exploit the theory of Eichler orders developed in the preceding section to study Shimura curves with no level structure and $\Gamma_0(\mathcal{N})$-level structure at the same time. In essence, the fact that we can develop the entire theory equally well at the level of an Eichler order highlights the special role played by $\Gamma_0(\mathcal{N})$-level structure – evidence for this turns up in each of the next three sections.

Remark: As a C-manifold, $S_{\mathcal{O}}$ depends a priori on the choice of Eichler order, or more precisely on its type. Recall (Corollary 40) that when the narrow class number of F is odd (so certainly when $F = Q$) there is a unique type of Eichler order of a given level, so in fact this ambiguity will not arise in the sequel. On the other hand, working with real quadratic F of class number divisible by a sufficiently large power of 2, [Vignéras] exploits this dependency on the type to exhibit arbitrary large families of Shimura curves S/F which, as Riemann surfaces, are cospectral but pairwise non-isometric.

0.3.1 Genus formulae for Shimura curves

From now on, we assume that B/F is a nonsplit quaternion algebra of type $(1, g-1)$, so that $S_{\mathcal{O}}$ is a compact Riemann surface (soon enough we will assume $F = Q$).

Let $\Gamma \leq SL_2(R)$ be a Fuchsian group of the first kind. Recall the general formula for the genus of (the compactified curve) $X_\Gamma := \Gamma\backslash\overline{\mathcal{H}}$:

$$2 - 2g(X_\Gamma) = -\frac{1}{2\pi} Vol(\Gamma\backslash\overline{\mathcal{H}}) + \sum_{q \geq 1} e_q \frac{q-1}{q} + e_\infty$$

where Vol denotes volume with respect to the standard invariant metric $\frac{dxdy}{y^2}$ and e_q indicates the number of elliptic points of order q and e_∞ the number of cusps. Notice that we need only compute the volume for one Γ in each commensurability class, since if $\Gamma' \leq \Gamma$, $Vol(\Gamma'\backslash\overline{\mathcal{H}}) = Vol(\Gamma'\backslash\overline{\mathcal{H}})[\Gamma : \Gamma']$. So let $\mathcal{O}$ be a maximal order; we record the

Proposition 45

$$Vol(\mathcal{O}^1\backslash\mathcal{H}) = \frac{1}{\pi}\zeta_F(2)\delta_F^{\frac{3}{2}}(4\pi^2)^{1-[F:Q]}\Pi_{p|D}N(p-1),$$

where ζ_F is the Dedekind zeta function, $\delta_F = \delta_{F/Q}$ is the absolute discriminant of F, and D is the discriminant of the quaternion algebra. When $F = Q$, this simplifies to

$$\frac{1}{2\pi} Vol(\mathcal{O}^1\backslash\mathcal{H}) = \frac{1}{6}\prod_{p|D}(p-1).$$

Notice that elliptic points correspond to roots of unity in B; since there could be lots of these for a quaternion algebra over an arbitrary totally real field, now is a good time to take $F = Q$ – finding genus formulae for more general Shimura curves is a subject unto itself (cf. [Sadykov] and [JLV].)

For the remainder of this thesis, B will denote a nonsplit indefinite rational quaternion algebra of discriminant D, and $\mathcal{O}$ will denote an Eichler order of (squarefree) level N. We write $X_0^D(N)$ for the Shimura curve $S_{\mathcal{O}}$; we abbreviate $X^D := X_0^D(1)$. Notice that in this case we can only have elliptic points of order 2 and 3 – indeed every nonreal root of unity $\zeta \in B$ lies in an imaginary quadratic field. More precisely, elliptic points of order 2 correspond to classes (modulo $\mathcal{O}^\times$) of optimal embeddings $Z[\sqrt{-1}] \hookrightarrow \mathcal{O}$, whereas elliptic points of order 3 correspond to classes of optimal embeddings $Z[\zeta_3] \hookrightarrow \mathcal{O}$. One of the fruits of our labors in Section 2 was formulae for these class numbers, so we get:

Proposition 46 *(Genus formula for X^D)*

$$g(X^D) = 1 + \frac{1}{12}\prod_{p|D}(p-1) - \frac{1}{4}\prod_{p|D}(1-(\frac{-1}{p})) - \frac{1}{3}\prod_{p|D}(1-(\frac{-3}{p}))$$

Using the fact that if $\mathcal{O}$ is a maximal order containing a squarefree level N-Eichler order $\mathcal{O}_N$, we have $[\mathcal{O}^1 : \mathcal{O}_N^1] = \Pi_{q|N}(q+1)$ and adjusting the class number formulas for embeddings into $\mathcal{O}_N$, we get:

Proposition 47 *(Genus formula for $X_0^D(N)$)*

$$g(X_0^D(N)) =$$

$$1+\frac{1}{12}\Pi_{p|D}(p-1)\Pi_{q|N}(q+1)-\frac{1}{4}\Pi_{p|D}(1-(\frac{-1}{p}))\Pi_{q|N}(1+(\frac{-1}{q}))-\frac{1}{3}\Pi_{p|D}(1-(\frac{-3}{p}))\Pi_{q|N}(1+(\frac{-3}{q}))$$

The Atkin-Lehner group: If $X = \Gamma\backslash\overline{\mathcal{H}}$ is a compact Riemann surface uniformized by a Fuchsian group, $N_{GL_2^+(R)}(\Gamma)/\Gamma$ acts as automorphisms on X: if $[\alpha] \in N\Gamma/\Gamma, \alpha : z \mapsto \alpha z$, $\alpha(\gamma z) = \gamma'\alpha z$, so that α acts on $\Gamma\backslash\overline{\mathcal{H}}$. If we take $\Gamma = \mathcal{O}^1$, where $\mathcal{O}$ is a level N Eichler order of B, then from Section 2 we know $N\mathcal{O}^\times/\mathcal{O}^\times \cong (Z/2Z)^{r+s}$, where $r = \#\{p \text{ such that } p|D\}$ and $s = \#\{q$ such that $q|N\}$. For $a|DN$, write γ_a for any representative, which is given by any element of reduced norm a. We call any such element an Atkin-Lehner element. Also, write $w_a : X_0^D(N) \to X_0^D(N)$ for the corresponding Shimura curve automorphism, called an Atkin-Lehner involution. In fact, for the most part we will be interested only in the γ_d for $d|D$; accordingly we write W for the subgroup generated by the γ_d for $d|D$ and call it the Atkin-Lehner group. (The terminology comes from the fact that when $D = 1$ the γ_s for $s|N$ are the classical Atkin-Lehner involutions that occur in the theory of elliptic modular curves, newforms and functional equations.) We shall see that the non-classical Atkin-Lehner involutions play an even greater role than one would expect from the classical case: indeed, throughout this thesis we shall be studying not the

Shimura curves $X_0^D(N)$ themselves, but certain Atkin-Lehner quotient curves.

Remark: We say w_D is the main Atkin-Lehner involution, and write $X_0^{D+}(N) := X_0^D(N)/w_D$. As an introduction to what makes w_D so special, notice that by Hasse's criterion, $Q(\sqrt{-D})$ splits B, so that we may take for our γ_D an element such that $\gamma_D^2 + D = 0$.

We want a genus formula for $X_0^{D+}(N)$ and indeed for $X_0^D(N)/W_H$ for any $W_H \leq W$. By Riemann-Hurwitz, this is equivalent to a formula for the number of fixed points $e_d := \#Fix(w_d)$. Such fixed-point formulae have been provided in several of the important papers written on the arithmetic geometry of Shimura curves circa 1980 – unfortunately, many of these published formulae are incorrect (a good check is to see that every Atkin-Lehner involution on a genus zero Shimura curve has precisely two fixed points!) Luckily for us, there is a quite careful treatment of this matter found in [Ogg I]; he proves the following

Proposition 48 *The number of fixed points of a nontrivial Atkin-Lehner involution w_m on $X_0^D(N)$ is given as a sum*

$$e_m = \sum_S h(S) \prod v_p(S, \mathcal{O}) = \sum_S h(S) \prod_{p|D} (1 - (\frac{S}{p})) \prod_{p|N} (1 + (\frac{S}{p}))$$

where we sum over certain imaginary quadratic orders S:
- *if $m = 2$, we sum over $S = Z[\sqrt{-1}],\ Z[\sqrt{-2}]$;*
- *if $m > 2$ and $m \equiv 1,\ 2\ (4)$, we have only $S = Z[\sqrt{-d}]$;*
- *if $m \equiv -1\ (4)$, we sum over $Z[\frac{1+\sqrt{-d}}{2}],\ Z[\sqrt{-d}]$.*

Proof: More precisely, we show that the w_m-fixed point locus is naturally in bijection with the union over sets of representatives for each of the inequivalent optimal embeddings of the quadratic orders given in the statement of the theorem. To be sure, that the locus of points P on the Shimura curve with representative $z \in \mathcal{H}$ whose stabilizer μ in $\mathcal{O}$ generates a given imaginary quadratic order R corresponds to the set of inequivalent optimal embeddings of that order R into $\mathcal{O}$ was – to put it mildly! – well-known to Shimura and can be found in many other sources. The point is to compute *which* quadratic orders intervene.

Recall that we may take as an Atkin-Lehner element any $\mu \in \mathcal{O}$ of reduced norm m, and that any of these elements generates the two-sided ideal $I(m) = \mu\mathcal{O} = \mathcal{O}\mu$. Suppose that $P \in X_0^D(N)(C)$ is a w_m-fixed point, and take a representative z for P in the upper halfplane. By definition of a w_m-fixed point, we have $\mu z = \gamma z$, for some norm one element $\gamma \in \mathcal{O}^\times$. Since we could equally well have chosen $\gamma\mu$ as our Atkin-Lehner element, we may assume $\mu z = z$; also, be replacing $\mu \mapsto -\mu$ if necessary, we may assume $tr(\mu) \geq 0$. Observe that $Q(\mu)$ is an imaginary quadratic field – indeed, representing μ by a matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we have the equation $cz^2 + (d-a)z - b = 0$. Since z lies in the upper halfplane, we have $(d-a) < 4bc$, which is equivalent to the inequality $tr(\mu) < 4\det(\mu)$.

Consider now the image μ' of μ under the canonical involution. It is equally well a generator of the ideal $I(m)$ and moreover satisfies $\mu + \mu' = tr(\mu) \in Q$, so we find that $\mu' = \epsilon\mu$ for some $\epsilon \in \mathcal{O}^\times \cap Q(\mu)$. Usually $\epsilon = -1$ so that μ is a "pure quaternion": $\mu^2 + m = 0$; more precisely this occurs unless the ring of integers of $Q(\mu)$ has nontrivial units. This can happen exactly when $m = 2$, so that $\mu = 1 + i$ (and $\epsilon = i$) or when $m = 3$, so that $\mu = 1 - \zeta_3$ (so that $\epsilon = \zeta_3$).

Take now R to be the imaginary quadratic order $4Q(\mu) \cap \mathcal{O}$. Clearly R contains $Z[\mu]$; we claim that if this containment is proper then necessarily $m \equiv 3 \bmod 4$ and then $R = Z[(1+\mu)/2]$. Indeed, assume that $\mu^2 = -m$, and put $\alpha = \frac{a+b\mu}{2} \in \mathcal{O}$. Then $a = tr(\alpha)$ and hence lies in Z, while $-bm = tr(\alpha\mu) \in tr(I(m)) \subset mZ$, so that b also lies in Z. Since $n(\alpha) = \frac{a^2+mb^2}{4} \in Z$, we get that if $m \equiv 3 \bmod 4$ $a \equiv b \bmod 2$, and otherwise a, b are both even, establishing the claim.

Thus we have seen that the imaginary quadratic orders generated by the $\mathcal{O}$-stabilizers are exactly as in the statement of the proposition; conversely, the argument shows that given such an element there is a w_m-fixed point. Clearly inequivalently embedded quadratic orders give rise to distinct fixed points, but what about equivalent embeddings? Consider an embedding equivalent to μ, so given by $\gamma\mu\gamma^{-1}$ for $\gamma \in \mathcal{O}^\times$. If γ has norm 1, then $\gamma\mu\gamma^{-1}$ fixes $\gamma(z)$, which represents the same point P on the Shimura curve. Suppose then that γ has norm -1; then (considering complex conjugation with respect to the canonical R-structure – see the next section), we get $\gamma\mu\gamma^{-1}$ fixes $\gamma(\overline{z})$. If P happens to be a real point – i.e., if $P = \overline{P}$, then (without loss of generality) $\gamma z = \overline{z}$, and $\overline{z}$ hence also z is fixed by $\gamma\mu\gamma^{-1}$, so that $\gamma\mu\gamma^{-1} = \mu'$ (not μ, because otherwise γ would be a unit in R hence have norm 1). Conversely, suppose $\mu' = \gamma\mu\gamma^{-1}$ for some $\gamma \in \mathcal{O}^\times$, necessarily of norm -1. As above we get $\gamma(z) = \overline{z}$, or $P = \overline{P}$. In summary, we get that $P = \overline{P}$ if and only if $\mu \sim \mu'$, which establishes that the fixed points correspond to equivalence classes of optimally embedded quadratic orders. Applying Eichler's embedding theorem, we are done.

As a corollary of the proof we deduce that the fixed points of w_d are all special points (soon to be called CM points, when we introduce the modular interpretation). Let us provisionally call a point $z \in \mathcal{H}$ lying over P on a Shimura curve K-special for some (unique) imaginary quadratic field K if $Stab(z) \cap B^\times = K^\times$. Notice that the designation special is justified by the fact that the K-special points form a countable set; indeed, given a fixed order S of K, S-special points correspond to classes of optimal embeddings $S \hookrightarrow \mathcal{O}$, which we know from Section 0.2 is a finite set. Notice also that different $d|D$ give rise to points which are K_d-special for different fields K_d, so that the fixed-point sets of the various nontrivial w_d's are pairwise disjoint. This enables us to write down a genus formula.

Proposition 49 *Let $HG \cong (Z/2Z)^r$ be a group of Atkin-Lehner involutions of*

cardinality 2^s. *Then*

$$g(X_0^D(N)/H) = 1 + 2^{-s}(g(X_0^D(N) - 1) - 2^{-s-1} \sum_{w_d \in H-1} \#Fix(w_d)$$

Corollary 50 *The genus of* $X_0^D(N)/H$ *goes to infinity in the sense that for any fixed number* G*, there exist only finitely many values of* D, N, H *such that* $g(X_0^D(N)/H) \leq G$.

Proof: We may without loss of generality take $N = 1, H = G$ since any Shimura curve has a finite map to one of these Shimura curves, hence the genera of the general curves will be at least as large as those of this particular form. Looking at the genus formula for X^D, we see that $g(X^D) \sim \frac{1}{12}D$ (as D approaches infinity through squarefree values). Using the facts that the class number of $Q(\sqrt{-d}) = O(\sqrt{d})$ and that $D/2^r \gg D^{\frac{2}{3}}$ (say), the result follows easily.

0.3.2 The moduli interpretation

We will now explain how Shimura curves like X^D and $X_0^D(N)$ are coarse moduli varieties for certain moduli problems (in the category of C-schemes, for now). We begin with the $N = 1$ case. Recall we fixed $\mathcal{O} \leq B$ a maximal order. We need to introduce a piece of auxiliary data, namely a choice of $\mu \in \mathcal{O}$ such that $\mu^2 = -D$. Associated to this μ we have an involution $b^\star := \mu^{-1}\overline{b}\mu$, where we reserve $b \mapsto \overline{b}$ for the canonical involution. One checks easily that $\star$ is a positive involution.

Consider now the following moduli problem: triples (A, ι, P), where A/C is an abelian surface, $\iota : \mathcal{O} \hookrightarrow End(A)$ is a ring homomorphism, and P is a principal polarization on A. The homomorphism $\mathcal{O} \hookrightarrow End(A)$ is called a QM structure. We require ι and P to be compatible in the following sense: the (Rosati) involution induced by P on $End(A)$ must stabilize $B \leq End^0(A) = End(A) \otimes Q$ and induce the involution $\star$ on $\mathcal{O}$. An isomorphism $\phi : (A, \iota, P) \to (A', \iota', P')$ is an isomorphism of underlying abelian varieties $\phi : A \to A'$ preserving the polarization: $\phi^\star(P') = P$ and which respects the two QM structures: $\iota' = \phi \circ \iota$.

We remark that it is well-known that the Rosati involution on a polarized abelian variety is positive for the trace form [Mumford II], which explains why we have made sure to choose our auxiliary data so as to make the involution positive. Philosophically speaking, the fact that the canonical involution on an indefinite rational quaternion algebra is never positive (due to the existence of real quadratic splitting fields) – compare with the positivity of the canonical involution on a definite quaternion algebra – forces us to make a noncanonical choice and thus complicates the entire picture in a way that will become fully clear in Chapter 1.

To connect this moduli problem with our Shimura curves, we construct a uniformization map $\mathcal{H} \to \{(A, \iota, P)/C\}/\cong$:

$z \mapsto (A_z, \iota_z, P_z)$, where $A_z = C^2/\mathcal{O}\begin{bmatrix} z \\ 1 \end{bmatrix}$; here we view

$$B \hookrightarrow B \otimes R = M_2(R) \hookrightarrow B \otimes C = M_2(C).$$

The complex torus A_z has an evident $\mathcal{O}$-action; moreover, it is projective via the Riemann form

$$E_z : \mathcal{O}\begin{bmatrix} z \\ 1 \end{bmatrix} \times \mathcal{O}\begin{bmatrix} z \\ 1 \end{bmatrix} \to Q,\ E_z(x\begin{bmatrix} z \\ 1 \end{bmatrix}, y\begin{bmatrix} z \\ 1 \end{bmatrix}) := tr(\mu x \overline{y}).$$

The data of the QM-structure and the principal polarization are not independent; indeed

Proposition 51 *(Milne) Given $(B, \mathcal{O}, \mu)$ as above and $\iota : \mathcal{O} \hookrightarrow End(A)$, there exists a unique principal polarization P on A such that (A, ι, P) is a compatible QM surface.*

A comment on polarized abelian varieties with many endomorphisms: this proposition is a little curious, since one of the most important technical ideas in the theory of moduli of abelian varieties is that it is more natural to study a polarized abelian variety than a naked abelian variety. Thus it may seem strange that given a QM structure the polarization comes for free. What we are seeing is the advantage of studying abelian varieties with sufficiently many endomorphisms: say a simple abelian variety over C of dimension d has sufficiently many endomorphisms (SM) if $End^0(A)$ contains a number field K/Q of degree d, and a general abelian variety is SM if it is isogenous to a power of a simple SM variety. Indeed, make the same definitions for an arbitrary complex torus. A generic complex torus of dimension $d > 1$ does not admit a polarization (and hence is not a projective variety) but any SM complex torus is polarizable: as in the QM case there will be a canonical Riemann form constructed from $K \leq End^0(A)$. Abelian varieties with SM play an especially large role in 21st century arithmetic geometry, since the condition of SM is necessary for modularity.

Dichotomy of simple versus CM: Recall that by definition the endomorphism algebra of a QM abelian surface contains at least the quaternion algebra B, so that its dimension is at least 4. It is easy to see that this is the maximal possible dimension for a simple abelian surface defined over a field of characteristic 0 (or equivalently by the Lefschetz principle, over the complex numbers). Indeed, let A/C be an abelian surface such that $D = End^0(A)$ is a division algebra. Writing $A = C^2/\Lambda$, D acts on $\Lambda \otimes Q$ (the "rational" rational representation); every module over a division algebra is free, hence has a D-dimension d: but $4 = \dim_Q \Lambda \otimes Q = d[D : Q]$, so that $[D : Q] \leq 4$. On the other hand, consider a nonsimple abelian surface $A \sim E_1 \times E_2$. If E_1 is not isogenous to E_2, we find $End^0(A) = End(E_1) \times End(E_2)$, which (being a nonsimple algebra of dimension at most 4) obviously does not contain B. So if A is a nonsimple QM surface, $A \sim E^2$ and $End^0(A) \cong M_2(End^0(E))$. If $End^0(E) = Q$, then it does not contain our nonsplit quaternion algebra B. We have therefore shown:

Proposition 52 *Let $(A, \iota, P)/C$ be a QM abelian surface. Then either A is simple, in which case $End^0(A) = B$, or $A \sim E^2$, where E/C is an elliptic curve with CM by a field K which splits B.*

In the latter case we refer to the corresponding point (A, ι, P) on the Shimura curve as a CM point. It is easy to see that that corresponds to the notion of a K-special point in our earlier terminology. In particular, the CM points on X^D form a countable subset, so that the generic QM surface is simple. Indeed, when we study QM surfaces as arithmetic-geometric objects, it is the simple ones that are of true interest, since the CM points are just repackaged CM elliptic curves. On the other hand, the skillful exploitation of "degenerate" objects in a moduli space is one of the classic tricks of the trade, and the reader should not be surprised that CM points will play an important role – all the more so since in our cocompact setting we have no truly degenerate objects (cusps) to work with.

$X_0^D(N)$ as a moduli space: Here we have our choice of moduli interpretations:

Proposition 53 *$X_0^D(N)/C$ is the coarse moduli space for each of the following moduli problems:*

M1: isomorphism classes of triples (A, ι, P) as above, but with $\iota : \mathcal{O}_N \hookrightarrow End(A)$, $\mathcal{O}_N$ a level N Eichler order.

M2: isomorphism classes of quadruples (A, ι, P, Q_N) where the first three components are as for X^D, and $Q_N \leq A[N]$ is a subgroup of order N^2, isomorphic as abelian group to $Z/N \oplus Z/N$ and cyclic as $\mathcal{O}$-module: there exists $P \in Q_N$ such that $\mathcal{O}P = Q_N$.

M3: isomorphisms classes of maps $\phi : A_1 \to A_2$, where ϕ is a QM-equivariant isogeny of the QM-surfaces $(A_1, \iota_1), (A_2, \iota_2)$, whose kernel is a cyclic $\mathcal{O}$-module of order N^2.

Proof: That $X_0^D(N)$ is the coarse moduli space for M1 is established by the same analytic construction as in the X^D case. Moreover, it is immediate to see that M2 and M3 are the same moduli problem: take $Q_N = Ker(\phi)$. To see the equivalence of M1 and M2, given an $\mathcal{O}$-QM abelian surface and a subgroup $Q_N \leq A[N]$ as in M2, let $\mathcal{O}' := \{x \in \mathcal{O} \mid xQ_N \leq Q_N\}$. We must check that $\mathcal{O}'$ is a level N Eichler order. Indeed $\mathcal{O}'$ is precisely the suborder of $\mathcal{O}$ consisting of elements which give well-defined endomorphisms of A_2, so ϕ induces a map $\mathcal{O}' \to \mathcal{O}_2 = End(A_2)$ and hence an automorphism of B. By Noether-Skolem, this automorphism is given as conjugation by some α, so $\mathcal{O}' = \mathcal{O}_1 \cap \alpha\mathcal{O}_1\alpha^{-1}$, hence it is an Eichler order. As for its level, we used earlier that $[\mathcal{O}^1 : \mathcal{O}_N^1] = \Pi_{p|N}(p+1)$, whereas it is clear from the defining property of $\mathcal{O}'$ that its norm 1 units have the same index in the norm 1 units of $\mathcal{O}$ (exactly as in the elliptic modular case). This completes the proof.

Remark: Writing a level N Eichler order as the intersection of two maximal orders $\mathcal{O}_N = \mathcal{O}_1 \cap \mathcal{O}_2$ gives by M1 two forgetful functors (degeneracy maps!) $q_1, q_2 : X_0^D(N) \to X^D$. The equivalence of M1 and M2 implies that $q_1 = w_N q_2$,

where w_N is the "main classical" Atkin-Lehner involution .

Modular interpretation of the w_d: Now that we can view our Shimura curves as moduli spaces, it is natural to ask for an interpretation of the w_d in terms of automorphisms of our moduli problem. This was done in [Jordan I]; we reproduce the work here. Choosing an Atkin-Lehner element α_d, we have $\alpha_d : (A_z, \iota_z, P_z) \mapsto (A_{\alpha_d z}, \iota_{\alpha_d z}, P_{\alpha_d z})$. Now notice

$$f : A_{\alpha_d z} = C^2/\mathcal{O}\begin{bmatrix}\alpha_d z\\ 1\end{bmatrix} \cong C^2/\mathcal{O}\alpha_d\begin{bmatrix}z\\ 1\end{bmatrix} = C^2/\alpha_d\mathcal{O}\begin{bmatrix}z\\ 1\end{bmatrix} \stackrel{\alpha_d^{-1}}{\to} C^2/\mathcal{O}\begin{bmatrix}z\\ 1\end{bmatrix},$$

so we find that α_d does not change the underlying abelian surface. On the other hand, $f \circ \iota_{\alpha_d z}(m) = \iota_z(\alpha_d^{-1} m \alpha_d) \circ f$, that is $\iota_{\alpha_d z} = \alpha_d^\star \iota$, where for any $\alpha \in N\mathcal{O}$, $\alpha^\star\iota$ twists the QM structure: $\iota(b) \mapsto \iota(\alpha_{-1} b \alpha)$. We also check that $\alpha^\star E_z = E_z(\alpha x\left[\begin{smallmatrix}z\\1\end{smallmatrix}\right], \alpha y\left[\begin{smallmatrix}z\\1\end{smallmatrix}\right])$ is the induced polarization. To summarize:

Proposition 54 *([Jordan I]) The Atkin-Lehner involutions w_d act on the moduli problem (A, ι, P) by preserving A, by twisting $\iota \mapsto \alpha_d^\star \iota$ and by carrying $E_z(x, y) \mapsto E_z(\alpha x, \alpha y)$.*

Shimura curves with level U structure: Finally, recall that from the perspective of Shimura varieties, any adelic level structure gives rise to a Shimura curve. That is, let U be a compact open subgroup of $B^\times(A_f)$. Then $X^D(U) = B^\times \backslash B^\times(A)/C^\times U$ is a C-manifold which is the coarse moduli space for (A, ι, P, u), where A, ι, P are as usual and u is a U-orbit of isomorphisms $TA \to \hat{\mathcal{O}}$ where TA is the full Tate module of A and $\hat{\mathcal{O}}$ is the profinite completion of $\mathcal{O}$. As examples, we take:

$U := \widehat{\Gamma_0(N)} := \{(x_p) \mid x_p \cong \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \pmod p \text{ for } p|N\}$

$:= \widehat{\Gamma_1(N)} := \{(x_p) \mid x_p \cong \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \pmod p \text{ for } p|N\}$

$:= \widehat{\Gamma(N)} := \{(x_p) \mid x_p \cong \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \pmod p \text{ for } p|N\}$

We see that, as long as N is prime to D, the notion of level N-structure is group-theoretically the same as in the elliptic modular case. By Proposition 53, the first of these curves is $X_0^D(N)$.

0.4 The canonical R-model

In this section we define an R-model for all our Shimura curves and discuss the real locus $X(R)$.

Recall that if X/L is a variety defined over a field L such that L/K is a finite Galois extension, then a K-model of X is a variety X_0/K such that $X_0 \times_K L \cong X/L$. One specifies a K-model via descent data: that is, for each $\sigma \in G_{L/K}$ we give an automorphism $f_\sigma : X \to X^\sigma = X \times_\sigma L$ such that

$f_{\sigma\tau} = \sigma(f_\tau) \circ f_\sigma$. In the special case that $L/K = C/R$, the descent data is specified by a single map $f : X \to \overline{X}$ with the property that $\overline{f} = f^{-1}$; such an f is said to be an antiholomorphic involution.

Now let $X/C = \Gamma\backslash\overline{\mathcal{H}}$ be any compact Riemann surface uniformized by a Fuchsian group. We can supply an antiholomorphic involution by giving a subgroup $\tilde{\Gamma}$ such that $\tilde{\Gamma} \leq GL_2(R)$ but not in $GL_2^+(R)$ and $\Gamma \leq \tilde{\Gamma}, [\tilde{\Gamma} : \Gamma] = 2$. In other words, $\tilde{\Gamma} = \langle \Gamma, \tilde{\gamma} \rangle$, where $\det(\tilde{\gamma}) < 0$ and $\tilde{\gamma}^2 \in \Gamma$. We claim that $\tilde{\Gamma}$ defines a real model of X/C. Indeed, define $f : \mathcal{H} \to \overline{\mathcal{H}}$ by $f(z) := \tilde{\gamma} z$; in other words, we have an antiholomorphic map $g : \mathcal{H} \to \mathcal{H}$ via $g(z) := \tilde{\gamma}\overline{z}$. Obviously Γ is normal in $\tilde{\Gamma}$, so $\tilde{\gamma}\overline{\Gamma z} = \tilde{\gamma}\Gamma\overline{z} = \Gamma\tilde{\gamma}\overline{z}$, and g descends to a map on X which is plainly an antiholomorphic involution.

The group $\tilde{\Gamma}$ for $X_0^D(N)$: since $X_0^D(N)/C = \mathcal{O}_N^1\backslash\mathcal{H}$ and $n : \mathcal{O}_N^\times \to \pm 1$, it is very natural to take $\tilde{\Gamma} := \mathcal{O}_N^\times$ to give a model $X_0^D(N)/R$. Notice that this choice is compatible with the (full) Atkin-Lehner group $N\mathcal{O}_N/\mathcal{O}_N^\times$, so that all the w_d are defined over R.

$X(R)$: The real locus of any Shimura curve is a compact 1-manifold, i.e., is a direct sum of circles. The only topological invariant therefore is the number of (analytic!) connected components $\#\Phi$. The first and most important result in this direction is due to Shimura:

Theorem 55 *(Shimura)* $X_0^D(N)(R) = \emptyset$.

Proof: Without loss of generality we may take $N = 1$, since the natural map $X_0^D(N) \to X^D$ would produce real points on X^D given any on $X_0^D(N)$. Let $\tilde{\gamma} \in \mathcal{O}^\times - \mathcal{O}^1$ be our antiholomorphic involution; a real point corresponds to a fixed point of $\tilde{\gamma}$. Say $\tilde{\gamma} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$; we must consider $z = \frac{a\overline{z}+b}{c\overline{z}+d}$, or $c|z|^2 + b = a\overline{z} - dz$. This has a solution for $z \in \mathcal{H}$ if and only if $a = -d$, in which case $\tilde{\gamma}^2$ is a homothety by $a^2 + bc$. It follows that $\alpha := \frac{\tilde{\gamma}}{a^2+bc}$ is an element of $B^\times$ such that $\alpha^2 = 1$. Since B is a division algebra, $\alpha = \pm 1$, which gives a contradiction since $n(\alpha) = \frac{n(\tilde{\gamma})}{(a^2+bc)^2} < 0$.

To gain some additional insight into this result, compare with the following

Proposition 56

a) There is no $(E, \iota)/R$, where E/R is an elliptic curve and $\iota : K \hookrightarrow End_R^0(E)$ is an R-rational CM-structure.

b) There is no $(A, \iota)/R$, where A/R is an abelian surface and $\iota : B \hookrightarrow End_R^0(A)$ is an R-rational QM-structure.

Proof: We prove b); a similar argument works for a). Write $V := H_1(A(C), Q)$ (singular homology!), so V/Q is a four-dimensional vector space. The R-structure on A gives rise to an antiholomorphic involution c on $A(C)$; let $W := V^{c=1}$; it is

a two-dimensional Q-vector space corresponding to the embedding $T_0(A(R)) \hookrightarrow T_0(A(C))$. Then, if ι is defined over R, we have $\iota : B \hookrightarrow End_Q(W) \cong M_2(Q)$, contradicting the fact that B is nonsplit.

Remark: To be sure, Proposition 56 is weaker than Shimura's theorem: since our moduli space is only coarse, it is possible that there could be R-points on $X_0^D(N)$ not induced by any structure $(A, \iota, P, Q_N)/R$. We will explore this issue in detail later.

On the other hand, Atkin-Lehner quotients of $X_0^D(N)$ may well have real points. Indeed the problem of the real locus of $X_0^D(N)/H$ was studied in [Ogg I]; he obtains the following

Theorem 57 *([Ogg I]) Let $1 \neq m|DN$. Let $v(m) = \sum_S h(S) \prod v_p(S, \mathcal{O}_N)$, S ranging over the set of orders of $Q(\sqrt{m})$ containing $Z[\sqrt{m}]$. Then the number $\#\Phi$ of analytic components of $X_0^D(N)/w_m(R)$ is $v(m)/2$, unless: $v(m) > 0$, $\sqrt{-1} \in \mathcal{O}_N$, $DN = 2t$ for odd t, $m = t$ or $2t$, and $x^2 - my^2 = \pm 2$ is solvable in integers x, y, in which case*

$$\#\Phi = \frac{v(m) + 2^{r+s-2}}{2},$$

where $r = \#\{p|D\}$, $s = \#\{q|N\}$.

We extract the case of interest to us in the sequel:

Corollary 58 *$X_0^{D+}(N)(R)$ is nonempty if and only if for all p dividing N, $(\frac{D}{p}) = 0$ or 1.*

0.5 The canonical Q-model and Shimura reciprocity

Finally in this section we introduce the basic arithmetic-geometric objects we wish to study: namely we define Q-models for our quaternionic Shimura curves. There are two approaches to this: by extending the moduli problem to Q-schemes, and by studying fields of moduli of CM points. The latter method was the one employed by Shimura in his seminal study of what are now called Shimura varieties. On the other hand, from a modern perspective the moduli method is a bit more familiar, so we shall give it first. In this way, Shimura's study of the CM points becomes an important theorem about the Shimura curves over Q. Still, knowing that the structure of the CM locus characterizes the Q-model is an instance of an important philosophy: that the arithmetic of a Shimura variety is controlled by the arithmetic of its special points.

Proposition 59 *There is a Q-model for $X_0^D(N)$ which is characterized as the coarse moduli scheme associated to the following moduli problem in the category of Q-schemes: $S/Q \mapsto \{(A, \iota, P, Q_N)/S\}/\cong$, where A/S is a two-dimensional*

abelian scheme, $\iota : \mathcal{O} \to End_S(A)$ is an $\mathcal{O}$-QM structure, P/S is a compatible principal polarization (with respect to a choice of auxiliary data $\mu \in \mathcal{O}, \mu^2 + D = 0$), and $Q_N \leq_S A[N]$ is an fppf-locally cyclic $\mathcal{O}$-module whose geometric fibres are of type $Z/NZ \oplus Z/NZ$.

We defer our discussion of the proof until the next section, in which we extend the moduli problem (verbatim) to $Z[\frac{1}{ND}]$-schemes.

Remarks: It is easy to see that the Atkin-Lehner involutions $w_m, m|DN$ have a moduli interpretation as in Section 3 – in particular the w_d preserve the underyling abelian variety and twist the quaternionic action, whereas the w_n are as in the classical case – hence they give automorphisms of $X_0^D(N)/Q$.
There is also an adelic formulation: for $U \leq \hat{\mathcal{O}}^\times$ a compact open subgroup, we have the curve X_U/Q which parameterizes level U-structures on the adelic Tate module of a QM-abelian surface (scheme). If $n : U \to \hat{Z}^\times$ is surjective, then X_U/Q is geometrically irreducible. In the general case, the finite group $\hat{Z}^\times/n(U)$ corresponds by class field theory to a finite abelian extension K/Q, and abelian extension gives the field of definition of a single connected component of X_U (i.e., exactly as in the elliptic modular case). In particular, $X_1^D(N)/Q$ is irreducible, whereas $X^D(N)$ has each connected component defined over $Q(\zeta_N)$, which explains why we do not consider the case of full level N-structure in the sequel.

Shimura reciprocity at the CM points: As alluded to above, we can explicitly compute the fields generated by the coordinates of CM points on Shimura curves. The fundamental result is:

Theorem 60 *(Shimura reciprocity law) Let $\varphi : \mathcal{H} \to X^D$ be the natural map. Let K be an imaginary quadratic field that splits B, and let $z \in \mathcal{H}$ be a point such that $\varphi(z)$ has CM by the maximal order R_K of K.*

a) We have $\varphi(z)K = K^1$, the Hilbert classfield of K.

b) The action of $G_{K^1/K} = Pic(R_K)$ on $\varphi(z) \in CM(R_K)$ is given by: $\sigma(\varphi(z)) = \varphi(\alpha^{-1}z)$, where z corresponds to the embedding $q : K \to B$ and if $\sigma = (a, K^1/K)$, then $q(a)\mathcal{O} = \alpha\mathcal{O}$ for $\alpha \in \mathcal{O}$ with $n(\alpha) > 0$.

In [Jordan I] this is pushed further: he studies the action of $W \times G_{K^1/K}$ on $CM(R_K)$, where W is the Atkin-Lehner group.

Proposition 61 *([Jordan I]) Let $d|D$ and $z \in \mathcal{H}$ represent a point on X^D having R_K-CM.*

a) If d is the norm of an ideal a of K, $w_d(\varphi(z)) = \sigma(\varphi(z))$, where $\sigma = (a, K^1/K)$.

b) Otherwise $w_d(\varphi(z)) \neq \sigma(\varphi(z))$ for any $\sigma \in G_{K^1/K}$.

Proposition 62 *([Jordan I]) Let $W'(K) \leq W$ be the subgroup of the Atkin-Lehner group generated by $\{w_p|p|D$ and p is inert in $K\}$. Then $W'(K) \times G_{K^1/K}$ acts simply transitively on the R_K-CM locus.*

Proposition 63 *([Jordan I]) Let $G = G(K^1/K)$; let $\pi : W'(K) \times G \to W'(K) \times G/G^2$ be the natural map. Let τ denote complex conjugation. For any fixed $z \in CM(R_K)$, there is a unique element $w_d \times \sigma$ such that $\tau(z) = w_d\sigma(z)$. Then the class of $w_d \times \sigma$ in $W'(K) \times G/G^2$ is independent of the choice of z, so τ corresponds to a well-defined element of $W'(K) \times G/G^2$. It is given as follows: $[\tau] = \pi(w_d, (a, K^1/K))$, where $B \cong (\frac{-s,dN(a)}{Q})$.*

Corollary 64 *Let $d'(K)$ be the product over those primes dividing D which are inert in K. Then $[\tau] = \pi(w_{d'(K)} \times (a, K^1/K)) \in W'(K) \times G/G^2$.*

Using these results, we can explicitly give all rational CM points on a degree two Atkin-Lehner quotient X^D/w_d. Indeed:

Proposition 65

a) If $x \in X^D/w_d(Q)$ is a rational R_K-CM point, the class number of R_K is 1 or 2.

b) Class number one case: every class number one R_K-CM point $x \in X^D(K)$ becomes rational on a unique degree 2 Atkin-Lehner quotient, namely on $X^D/w_{d'(K)}$. In particular, if D is prime to the discriminant of K, x induces a rational point on X^{D+}.

c) Class number two case: let $x \in X^D(K^1)$ be an R_K-CM point, where K is a class number two field. If D does not divide the discriminant of K, then x does not become rational on any twofold Atkin-Lehner quotient. Conversely, if D divides the discriminant of K then $x \in X^D/w_d(Q)$ unless:
$d \equiv 3(4), D = d$, and $K = Q(\sqrt{-D})$; or
$d \equiv 1(4), D = d, K = Q(\sqrt{D})$ or $d = D/2$, $K = Q(\sqrt{\frac{-D}{2}})$; or
$d \equiv 2(4), d > 2; D = d, K = Q(\sqrt{-D})$.
Notice that in all there are only finitely many rational CM points on twofold Atkin-Lehner quotients arising from class number two CM fields.

0.6 Fields of moduli and fields of definition

Let $A/\overline{K}$ be a structure defined over the algebraic closure of a field K. We define the field of moduli L of A to be the field cut out be the subgroup $H := \{\sigma \in G_K | A \times_\sigma \overline{K} \cong A\}$. In case we have a (coarse or fine) moduli space X/K for a set of structures, we can also characterize the field of moduli of an $A \in X(\overline{K})$ as the field of definition of the point A on X (i.e., as the field extension of K generated by the coordinates of A in any local affine model of X, or equivalently as the residue field of the local ring of the closed subscheme defined by A). Contrast this with the notion of a field of definition: we say L/K is a field of definition for $A/\overline{K}$ if there exists an L-model for A.

Let us discuss how the two notions are related: it is immediate that the field of moduli is contained in every field of definition. The field of moduli is unique;

there are many fields of definition. The most pleasant state of affairs would be if A could be defined over its field of moduli – in particular there would then be a unique minimal field of definition. Whether or not a variety of a certain type can be defined over its field of moduli can be an interesting question. Here are some examples of results in this direction:

Theorem 66

a) Any elliptic curve can be defined over its field of moduli.

b) (Shimura) More generally, the generic odd-dimensional principally polarized abelian variety can be defined over its field of moduli.

c) (Shimura) No generic even-dimensional principally polarized abelian variety can be defined over its field of moduli.

d) (Mestre) If $C/\overline{K}$ is a genus 2 curve with field of moduli K whose only non-trivial automorphism is the hyperelliptic involution, then C can be defined over K if and only if a certain obstruction in $Br(K)[2]$ vanishes.

e) (Cardona-Quer) If $C/\overline{K}$ is a genus 2 curve with larger automorphism group, then it can be defined over its field of moduli.

f) (Jordan) Let $(A, \iota, P)/\overline{K}$ be a QM-surface with field of moduli K. Then (A, ι, P) can be defined over a field L if and only if L is a splitting field for B.

g) (Shimura) Any CM abelian variety can be defined over its field of moduli.

Jordan's result will be fundamental for our study of Shimura curves and will be elaborated upon in Chapter 2. Notice that, like Mestre's, it is also obstruction-theoretic in nature: the obstruction to a QM surface being defined over a field L containing the field of moduli is the element $[B] \in Br(L)$. The necessity is rather clear: if $(A, \iota, P)/L$, then the action $B \hookrightarrow End_L^0(A)$ gives rise by $\Omega^1_{A/L}$ to a map $B \hookrightarrow M_2(L)$, so L splits B. The sufficiency is accomplished by a Galois descent argument, for which see [Jordan II].

0.7 The integral canonical model I: good reduction

In this section we give canonical models over Z for our Shimura curves, by providing moduli problems in the category of Z-schemes. In particular, we study the curve $X_0^D(N)/Z$; it turns out that the canonical model of this curve is smooth over $Z[\frac{1}{DN}]$, and we analyze the situation of QM abelian surfaces modulo a good prime p in this section. When p divides N then – recalling our convention that N be squarefree – the curve $X_0^D(N)$ has semistable bad reduction of a kind completely analogous to the reduction of $X_0(N)$ at N – we call this kind of reduction Deligne-Rapoport reduction and study it in the following section. When finally $p|D$, we find a phenomenon without an elliptic modular analogue: again the curve has semistable bad reduction, but in this case the special fibre of $X_0^D(N)$ is a reducible curve, each component of which

has geometric genus 0. We call this type of reduction Cerednik-Drinfeld reduction and will have (much) more to say about it in the last section of this chapter.

The moduli problem for $X_0^D(N)$ over Z: It is almost the same as the moduli problem over Q (and indeed would be verbatim over $Z[\frac{1}{D}]$): to a scheme S we associate the set of isomorphism classes of structures (A, ι, P, Q_N) where A/S is an abelian scheme of relative dimension 2, $\iota : \mathcal{O}_D \to End_S(A)$ is a quaternionic structure, P is the induced polarization, $Q_N \leq A[N]$ is a subgroup scheme which is, fppf locally on S, cyclic as $\mathcal{O}_D$-module and isomorphic as a constant group scheme to $Z/NZ \oplus Z/NZ$, and: the quaternionic structure satisfies the additional condition that for all $m \in \mathcal{O}_D$, the trace of $\iota(m)$ acting on the Lie algebra of A coincides with $t(m)$ (reduced trace). It is immediate to check that this holds in characteristic 0, so this is indeed an extension of our earlier moduli problem.

Remark: In much of the literature (e.g. [Milne]) one sees the moduli problem given in terms of weak polarizations (two polarizations P, P' give the same weak polarization if there exist positive integers m, n such that $mP = nP'$). Milne's proof of the uniqueness of the polarization compatible with the QM structure actually establishes the uniqueness of the weak polarization. But then work of Jordan (generalized by [Rotger II-IV]) shows that by a correct choice of the auxiliary data μ we can get the polarization to be principal. This is done by constructing Riemann forms, i.e., is a priori valid only in characteristic 0. However, we can argue for the existence of a principal polarization in positive characteristic: it is really a matter of X^D, not $X_0^D(N)$ so we have only good primes and Cerednik-Drinfeld primes $(p|D)$. For a good prime the mod p Shimura curve is smooth, hence is the reduction of some characteristic 0 point, and the principal polarization comes down to us from characteristic 0. Even for a Cerednik-Drinfeld prime there are only finitely many nonsmooth points at which the existence of a principal polarization is in doubt; thus we have a rational map from a curve into the space of principally polarized semi-abelian surfaces, a complete variety. But a semi-abelian QM surface is abelian. It follows that the map extends to these finitely many exceptional points and gives principal polarizations there as well.

But we are getting a little ahead of ourselves. The (coarse) representability of the moduli problem is actually a major

Theorem 67 *(Drinfeld) There exists a coarse moduli scheme $X_0^D(N)/Z$ attached to the above moduli problem. Moreover the scheme $X_0^D(N)/Z$ is flat, projective, integral, of relative dimension 1, and its restriction to $Z[\frac{1}{ND}]$ is smooth.*

This is not at all an easy theorem, and we do not discuss the proof here. Notice that the situation is fundamentally more difficult than in the elliptic modular case: one can define $X_0^D(N)/Q$ by relatively elementary means (e.g. via the "generic" elliptic curve over $Q(t)$; see [Rohrlich]), and one knows an integral

model for X^1/Z, namely A^1/Z. Thus we can at least define an integral model for $X_0(N)$ by taking the normalization of X^1 in $X_0(N)/Q$. One still has to check that this is the right object (but it is). In our case, however, we have a "two-dimensional modular tower" of curves, and there is never an A^1 at the bottom. We will in construct the integral model for $X_0^D(N)$ in Section 1.X, but we do not show that it is actually the coarse moduli scheme – in fact, we do not use this fact anywhere in the thesis; the existence of a Z_p-regular model is enough.

QM surfaces over $\overline{F_p}$: So let $A/\overline{F_p}$ be a QM abelian surface, where p does not divide D. We shall see that this is a rather different sort of object than a QM surface in characteristic 0.

The p-divisible group: let $\mathcal{A}$ be the associated p-divisible group, i.e., $\lim A[p^n]$, and let $D(\mathcal{A})$ be its associated Dieudonné module, a free rank 4 module over $W(\overline{F_p}) = Z_{p^\infty}$ endowed with semilinear actions of F and V. Notice that $\mathcal{O}_D \otimes W(\overline{F_p}) \cong M_2(Z_{p^\infty})$ acts on D, so that by choosing nontrivial idempotents $e_1 + e_2 = 1$ we get a splitting $D = D_1 \oplus D_2$ into isomorphic $Z_{p^\infty}[F,V]$-submodules. Comparing this splitting with the known list of possible slope sequences for the p-divisble group of an abelian surface – namely $\{0,0,1,1\}, \{0,1/2,1/2,1\}, \{1/2,1/2,1/2,1/2\}$, we find that only the first and the third are possible: that is, $A/\overline{F_p}$ is either ordinary – equivalently, its p-rank is equal to its dimension –or supersingular – equivalently, it is isogenous to a product of supersingular elliptic curves (from the short list of formal isogeny types we have just exhibited, it is equivalent in dimension 2 to require p-rank 0, but already in dimension 3 we have (1/3,1/3,1/3,2/3,2/3,2/3) and this is no longer the case).

Recall that for any nonsupersingular formal isogeny type, there exist (geometrically) simple abelian varieties of that isogeny type. However, an ordinary QM surface over $\overline{F_p}$ is in fact isogenous to the square of an elliptic curve. The reason for this is the substantially different theory of endomorphism algebras of abelian varieties in positive characteristic: B_D is not an acceptable choice for the full endomorphism algebra of an abelian surface over a finite field; however, it is large enough to be incompatible with any division endomorphism algebra. The formal proof of this requires the Honda-Tate theory of isogeny classes of abelian varieties over finite fields. See the Appendix for a self-contained account of Honda-Tate theory with applications to the problem at hand. In particular, we conclude:

Proposition 68 *Let $A/\overline{F_p}$ be a QM surface. Then A is isogenous to the square of an elliptic curve E. Accordingly, the full endomorphism algebra $M := End^0(A)$ is in the ordinary case $M_2(K)$, where K/Q is a CM quadratic field and in the supersingular case $M_2(B_{p,\infty})$, where $B_{p,\infty}$ is the quaternion algebra over Q ramified precisely at ∞ and p.*

Corollary 69 *Let $H := End_B(M)$ be the commutant of B_D in the full en-*

domorphism algebra M of A. In the ordinary case $H = K$, whereas in the supersingular case $H \cong B_{\infty,pD}$ is the quaternion algebra over Q ramified at ∞, p and at the primes dividing D.

Proof: Indeed, since $B \hookrightarrow M = M_2(End^0(E))$ is an embedding of a CSA over Q into a simple Q-algebra, we have $B \otimes_Q H$ is isomorphic to the centralizer of Q in M, i.e., to M itself. In the ordinary case, this tells us that the dimension of H is 2, and since certainly $K = Z(M)$ centralizes B, we must have $H = K$. In the supersingular case, we get that the dimension of H is 4 and moreover the equality $[B_D] + [H] = [B_{p,\infty}]$ in $Br(Q)[2]$, whence the result.

Isogeny classes in $X^D(U)(\overline{F_p})$, especially the supersingular class: The analysis of points mod p on quaternionic Shimura varieties has been carried much further, en route to computing the local factor of the Hasse-Weil zeta function at p. Indeed, it is no more difficult (and more natural) to work with a general adelic level U structure, assumed maximal at p: $U = U^pU_p$, where $U_p = G(Z_p) \cong GL_2(Z_p)$ – here $G = \mathcal{O}_D^\times$ viewed as a group over Z. The set $X^D(U)(\overline{F_p})$ parameterizes isomorphism classes of structures (A, ι, ϕ) as in Section 3. A key step in its determination is the forgetful map $(A, \iota, \phi) \to (A, \iota) \otimes Q$, where $(A, \iota) \otimes Q$ denotes the class of all QM surfaces which are B-equivariantly isogenous to (A, ι). The fibres of this map are called the isogeny classes (and are Frobenius and Hecke stable); the problem is reduced to determining how many fibres there are and then what is the structure of each fibre. This is all explained very carefully in [Milne] (see also [VFL] for a treatment of the case of higher-dimensional totally indefinite quaternionic Shimura varieties; the results are morally the same but the details are significantly more onerous); we content ourselves here with a summary of Milne's results:

Theorem 70 *The set $\mathcal{J}_p$ of B-isogeny classes is given as follows: for each quadratic imaginary field K which splits B and in which p splits there is a corresponding isogeny class; moreover there is a unique supersingular isogeny class.*

Let us at least remark that the field K corresponds to the endomorphism algebra of the ordinary elliptic curve E such that $A \sim E^2$; both splitting conditions are rather obviously necessary (the former since $B \hookrightarrow M_2(K)$ and the latter is Honda-Tate theory for an ordinary elliptic curve; cf. the Appendix). Also, it is true by definition that all supersingular abelian surfaces are isogenous, but it is not so obvious that they are B-isogenous; this is part of the proof.

Let $(A, \iota, \phi) \in X^D(U)(\overline{F_p})$. Let $D'A := DA \otimes Q$ be its rational Dieudonné module. Write X for the set of "suitable lattices" in $D'A$, namely for the set of $Z_{p^\infty}[F, V]$-submodules which are Z_{p^∞}-free of rank 4 and $\mathcal{O}_B$-stable. A very useful expression for its isogeny class $Z(A, \iota, \phi)$ is given by the following

Theorem 71 *The isogeny class of (A, ι, ϕ) can be given as a double-coset space:*

$$Z(A, \iota, \phi) \cong H(Q)\backslash G(A_f^p) \times X/U^p$$

Moreover Frobenius acts on the isogeny class by sending $M \in X$ to FM.

We are especially interested in the supersingular class: one finds that $X \cong \overline{G}(Q_p)/\overline{G}(Z_p)$, where $\overline{G} = End_{\mathcal{O}_B}(D'A)^\times$ is the unit group of the $\mathcal{O}_B$-equivariant endomorphisms of the Dieudonné module. This is the local version of the calculation performed in the last corollary, so in the supersingular case we get $\hat{B}_p^\times/\hat{\mathcal{O}}_p^\times \cong Z$ – the nonzero elements of the unique division quaternion algebra over Q_p modulo units in the maximal order.

Let us specialize these results to the cases of $\Gamma_0(N)$-level structure, i.e., we can take for our U the profinite completion of the units in a level N Eichler order $\mathcal{O}_N$.

Proposition 72 *Each point in the supersingular locus $X_0^D(N)(\overline{F_p})^{ss}$ is defined over F_{p^2}.*

Proof: Recall that the square of the unique prime $\mathcal{P}$ of $\widehat{\mathcal{O}_p}$ is the ideal generated by p, so F^2 acts as multiplication by p. Writing the supersingular point as $(g, x) \in G(A_f^{)} \times X$, since $p \in H(Q)$ we have $(g, px) \sim (p^{-1}g, x)$; but because of our choice of level structure, p^{-1} lies in U^p – one checks this componentwise using the descrption of Eichler orders in quaternion algebras over local fields given in Section 0.2. This completes the proof.

Corollary 73

a) The supersingular locus on $X^D(\overline{F_p})$ is isomorphic to the "Brandt set" $Pic_l(\mathcal{O})$, where $\mathcal{O}$ is a fixed maximal order in the definite rational quaternion algebra of discriminant pD.

b) The supersingular locus on $X_0^D(N)(\overline{F_p})$ is isomorphic to the "Brandt set" $Pic_l(\mathcal{O}')$, where $\mathcal{O}'$ is a fixed N-Eichler order in the definite rational quaternion algebra of discriminant pD.

Proof: We must emphasize the "interchange of indices" that is taking place: we are going from one 2-torsion Brauer group element, B_D, to another Brauer group element differing by $[p] + [\infty]$. But notice that $G(A_f^p)$ does not depend on p or ∞. This remark, together with the computation of X above and the "global-adelic dictionary" from Section 1.2, give the result.

Remark: In fact it would have been acceptable to take $D = 1$ throughout this section (subject to the proviso that our moduli spaces would no longer be projective due the presence of cusps) and we would reacquire familiar results, in particular the isomorphism of the Brandt set of ideals for a maximal (resp. N-Eichler) order in the definite quaternion algebra $B_{p,\infty}$ with the set of supersingular elliptic curves mod p (resp. supersingular $\Gamma_0(N)$-structured elliptic curves); we will find a use for this correspondence as well later on.

0.8 The integral canonical model II: Cerednik-Drinfeld reduction

Let B/Q be an indefinite rational quaternion algebra, $U \leq B^\times(A_f)$ a compact open subgroup of the finite adelic points which is maximal at a fixed ramified prime p of B. The p-adic uniformization theory developed by Cerednik and refined by Drinfeld furnishes us with a model X_U/Z_p of the associated Shimura curve. The goal of these notes is to give an overview with some details of this theory, and especially, in the case of a "connected" Shimura curve and a quaternion algebra of discriminant pq, to express the data of the special fibre in terms of supersingular elliptic curves.

We do not offer any indication of a proof of the main theorem (i.e., we shall not mention moduli of p-divisible groups). For this we refer the reader to the excellent treatment given in [Boutot-Carayol].

0.8.1 Preparation for Mumford curves

In the early 1970s, [Mumford I] did fundamental work on uniformization of certain curves over complete local rings; his theory is motivated simultaneously by Tate's analytic construction of semistable elliptic curves over complete rings and by older work of Schottky on uniformization of curves over the complex numbers by means of Schottky groups. By way of introduction, let us say a few words about each of these theories: Schottky starts from a discrete subgroup Γ of $PGL_2(C)$ acting discontinuously at at least one point of $P^1(C)$ and which as a group is free on n generators. He shows that the set of points $\Omega \subset P^1(C)$ on which Γ acts discontinuously is connected and open, and the quotient Ω/Γ is a compact Riemann surface of genus g. Now, working say over Q_p, Tate's elliptic curve is of the form $E_q = Q_p^\times/q^Z$, where $q \in Q_p^\times$ is some integral element. Then $j(E_q) = \frac{1}{q} + 744 + \ldots$ is nonintegral, so that Tate's elliptic curve has bad – indeed split multiplicative – reduction. Recall also that it is not quite true that any elliptic curve over Q_p with multiplicative bad reduction (i.e., nonintegral j-invariant) is isomorphic to a Tate curve *over the ground field* – rather, every semistable elliptic curve over a local field is isomorphic to a twist of a Tate curve.

To see the relation between Tate curves and Schottky curves, observe that $\begin{bmatrix} 1 & 0 \\ 0 & q^n \end{bmatrix}$ embeds q^Z as a discrete subgroup of $GL_2(Q_p)$, and the only points in $P^1(Q_p)$ at which this group acts discontinuously are the two fixed points $0, \infty$. To follow Schottky then, we take $\Omega = P^1(Q_p) - \{0, \infty\} = Q_p^\times$ and indeed $\Omega/\Gamma = Q_p^\times/q^Z$ gives us our Tate curve – notice that since 1 is at the same time the genus of the quotient curve and the rank of the free group q^Z, the analogy to the classical case is very strong. Mumford's work generalizes Tate curves to the higher genus case, in a way which we will now explain.

A p-adic upper halfplane: Let Δ be the Bruhat-Tits tree of $PGL_2(Q_p)$, whose vertices parameterize homothety classes of Z_p- lattices $M \subset Q_p^2$; recall that two classes $[M_1], [M_2]$ are defined to be adjacent if there exist representative lattices with bases related as follows: $M_1 = \{a, b\}, M_2 = \{a, pb\}$. This gives a tree in which each vertex has degree $p + 1$. On the other hand, consider the category $\mathcal{Z}$ of integral Z_p-schemes Z/Z_p endowed with an isomorphism $P^1(Q_p) \xrightarrow{\sim} Z_\eta$ (where we denote by η the generic fibre of a Z_p-scheme). Now each vertex of the Bruhat-Tits tree $M \in \Delta$ naturally gives rise to such a scheme $P(M)$, namely $P(M) := Proj(Sym(M^\vee))$. To spell this out a bit, if $M = aZ_p \oplus bZ_p \subset Q_p^2$, let $X, Y : M \to Z_p$ via $X(a) = 1, X(b) = 0$, $Y(a) = 0, Y(b) = 1$, then $P(M) = Proj(Z_p[X, Y])$, and the isomorphism we take on the generic fibre is the evident one given by tensoring the graded algebra to Q_p. In fact, this construction gives an embedding of Δ onto the subcategory of $\mathcal{Z}$ given by schemes which are abstractly isomorphic to $P^1(Z_p)$ – it is worth noting that the evident change of variables on X, Y given by a matrix in $PGL_2(Q_p)$ acts transitively (by abstract isomorphisms!) on this subcategory of schemes, but it is only the matrices in $PGL_2(Z_p)$ which are compatible with the choice of isomorphism on the generic fibres. We can give the entire category $\mathcal{Z}$ a partial ordering by decreeing $Z_1 > Z_2$ if and only if there exists a Z_p-morphism $\varphi : Z_1 \to Z_2$ whose restriction to the generic fibre is the identity on $P^1(Q_p)$ (to be interpreted with respect to the given isomorphisms of the generic fibres with P^1). Then any two elements Z_1, Z_2 of our category have a least upper bound, called their *join*: by construction we have a canonical isomorphism φ from the generic fibre of Z_1 to the generic fibre of Z_2 we take $J(Z_1, Z_2)$ to be the closure in $Z_1 \times_{Z_p} Z_2$ of the graph of φ – it has all the desired properties.

Example: If M_1, M_2 represent vertices of Δ whose distance in the tree is n, then $J(P^1[M_1], P^1[M_2])$ is given by the closure of the equation $Y_0X_1 - p^nX_0Y_1$ in $Proj(Z_p[X_iY_j])$; notice that its special fibre is a nodal curve, and the singularity is analytically isomorphic to $Z_p[[T_1, T_2]]/(T_1T_2 - p^n)$.

Proposition 74 *Let $\{Z_1, \ldots, Z_n\} \subset \mathcal{Z}$ be any finite subset. Then the join (i.e., lub) of these elements, $J(Z_1, \ldots, Z_n)$ exists in $\mathcal{Z}$; it is normal, proper and flat over Z_p, and generically isomorphic to P^1/Q_p.*

Proof: We construct the join as in the case of two elements, namely as the closure in the fibre product of the graphs of the generic isomorphisms between all the factors.
If C/Z_p is a proper, flat, normal curve, we say it is F_p-split degenerate if its special fibre is a reduced irreducible curve every component of which has geometric genus zero, and every singularity is nodal and occurs at an F_p-rational point. The special fibre of such a curve such a special fibre is essentially a combinatorial rather than a geometric object (there are no moduli!), and as such can be completely described via combinatorial means: the *dual graph* to such a curve is the finite graph whose vertex set is the set of irreducible components, and the edge set is the set of singular points. Now we can enunciate the

Proposition 75 *For any $M_1, \ldots, M_n \in \Delta$, the join $J(P^1[M_1], \ldots, P^1[M_n])$ has F_p-split degenerate special fibre, whose dual graph is a finite tree Δ' on the vertex set $M_1, \ldots, M_n$.*

Indeed, more is true:

Proposition 76 *The join $J(\{P^1[M_i] | i \in \Delta\})$ exists in $\mathcal{Z}$. Its closed fibre has dual graph Δ (so is not of finite-type!).*

Proof: Fix a vertex $M \in \Delta$, and write $\Delta = \bigcup_{n \geq 0} \Delta_n$, where Δ_n is the union over all the paths in Δ with origin M and length at most $n+1$. By construction of the join we have a morphism $P^1[\Delta_{n+1}] \to P^1[\Delta_n]$, and it is not hard to see that this birational morphism blows down the locus corresponding to the vertices of $\Delta_{n+1} \setminus \Delta_n$. Let $U_n \subset P^1[\Delta_n]$ be the complement of the finite set of (singular) points corresponding to the edges of $\Delta_{n+1} - \Delta_n$. Observe that the morphism $P^1[\Delta_{n+1}] \to P^1[\Delta_n]$ becomes an isomorphism on the preimage of U_n. Therefore we get open immersions $U_0 \hookrightarrow U_1 \hookrightarrow \ldots$, and we can glue to get the desired scheme.

We denote by $\mathcal{P}$ the Z_p-formal scheme obtained by completing $P^1[\Delta]$ along the special fibre. It will be our p-adic upper halfplane.

Remark: As the terminology suggests, the p-adic upper halfplane plays as basic a role in p-adic geometry as the usual upper half plane plays in complex geometry – it is a (non-algebraic!) analytic object which gives rise to many algebraic objects by an analytic construction (uniformization). A difference between complex analytic spaces and p-adic analytic spaces is that there is a universally agreed upon definition for the former, whereas there are at least three different frameworks for the latter: formal schemes up to admissible blowups, rigid analytic spaces *a là* Tate, and Berkovich analytic spaces. The p-adic upper halfplane exists in each of these categories and it is morally – but not exactly – the same as the object we have defined here. E.g., a Z_p-formal scheme gives rise to the structure of a rigid analytic space (Raynaud's "generic fibre" construction), but one generally understands the rigid p-adic upper half plane to have generic fibre $P^1(C_p) \setminus P^1(Q_p)$; equivalently, it is obtained from our p-adic upper halfplane by removing all the F_p-rational points from the special fibres of the basic objects $P^1(M)$.

Observe that, by construction, $PGL_2(Q_p)$ acts on $\mathcal{P}$.

0.8.2 Cocompact Schottky groups

Let $\Gamma \leq PGL_2(Q_p)$ be a finitely generated discrete subgroup such that $\Gamma \backslash PGL_2(Q_p)$ is compact. Consider the quotient map $\Gamma \backslash PGL_2(Q_p) \to \Gamma \backslash \Delta$; the image is at once compact and discrete, i.e., finite. By similar reasoning, it turns out that for any edge e of the Bruhat-Tits tree, the edge-stabilizer $\Gamma_e < \Gamma$ is finite. For any edge $\overline{e}$ of the finite graph $\Gamma \backslash \Delta$, we define its *length* $l(\overline{e})$ to be the cardinality

of its stablizer. We call this data of a graph together with a "length function" on its edges an *l-graph*.

The Mumford curve $\Gamma\backslash\mathcal{P}$: We are now going to fulfill our first goal, namely to our discrete cocompact subgroup $\Gamma < PGL_2(Q_p)$ we shall associate a curve X_Γ/Z_p, a Mumford curve. To start, one knows that every discrete subgroup $\Gamma \leq PGL_2(Q_p)$ is virtually torsionfree: there exists a finite index torison-free normal subgroup $\Gamma_1 \leq \Gamma$; moreover one can show that such a Γ_1 acts freely on Δ (Ihara's theorem) and consequently is a free group. One says that Γ_1 is a p-adic Schottky group. We can find another finite index normal subgroup $\Gamma_2 \leq \Gamma_1$ with the property that no γ in Γ_2 maps any vertex in Δ to an adjacent vertex – indeed, this amounts to choosing a finite Galois covering space of the finite graph $\Gamma_1\backslash\Delta$ which unwraps all the loops of $\Gamma_1\backslash\Delta$. Working now with Γ_2, we see that the special fibre of $\mathcal{P}$ can be covered by open affines $\mathcal{U}_i$ such that $\gamma U_i \cap U_i = \emptyset$ for any nonidentity element γ of Γ_2. Take now the induced formal open affines of $\mathcal{P}$ (complete preimages), which we continue to denote by U_i; these cover $\mathcal{P}$, and for any pair of indices i, j, there is at most one $\gamma \in \Gamma_2$ such that $\gamma U_i \cap U_j$ is nonempty. Thus we can glue to construct the quotient $\Gamma_2\backslash\mathcal{P}$. This quotient is a projective formal scheme, so it is algebraic, i.e., it is uniquely the completion along the closed fibre of a proper, normal, Z_p-flat curve X_{Γ_2}. But now the quotient of this projective scheme by the finite group Γ/Γ_2 can certainly be taken; as a result, we have realized $\Gamma\backslash\mathcal{P}$ as an F_p-split, degenerate, semistable curve over Z_p.

Having completed the basic construction, we pause for some remarks:

First, it should be clear that everything we have done so far would be valid with Q_p replaced by any locally compact non-Archimedean field K. (We have chosen to formulate the construction in terms of Q_p for the sake of specificity and also to point out a key point on the sort of curves which can arise as Mumford curves; this is coming up in the next section). Indeed Mumford's work is significantly more general: he works even with an arbitrary integral complete local ring A (not necessarily a DVR). Later authors seem not to have carried on this much generality, but the intermediate situation of the valuation ring of a complete, local field K with infinite residue field (e.g. C_p) is important. In this case, the Bruhat-Tits tree Δ is no longer locally finite, so one cannot hope to mod out by a discrete subgroup and get a finite graph. The solution here, as in the important case when the discrete group Γ is not cocompact, is to work with a subtree Δ_Γ associated to Γ, so that the quotient $\Gamma\backslash\Delta_\Gamma$ is once again finite. For the construction of this tree see e.g. [Schmecta].

0.8.3 Base extension and admissible curves

The theory of Mumford curves we have developed in the cocompact case is not enough to encompass the p-adic uniformization of modular curves. We know this already, e.g. we recalled above that if E/Q_p is an elliptic curve with multi-

plicative bad reduction, then E has a p-adic uniformization over Q_p if and only if the multiplicative reduction is split [Silverman]. In general, what we can say is that a semistable elliptic curve E/Q_p is a twisted form of a Tate curve.

Nonexample: Consider $X_0(p)/Z_p$. It is well-known that the special fibre is a "double helix" (see the next section, where the analogous phenomenon is explored in the Shimura curve case): it has two irreducible components, each isomporphic to $X(1)$, which intersect along the supersingular locus – however, one knows that the supersingular points are all defined over F_{p^2}, but they are in general not all defined over the prime subfield. That is, the special fibre of $X_0(p)/Z_p$ is degenerate but not F_p-split, hence is *not* a Mumford curve. But neither is it a Z_p^∞/Z_p-twisted form of a Mumford curve; indeed, for sufficiently large p, $\mathrm{Aut}(\mathrm{X}_0(p))$ is generated by the Atkin-Lehner involution w_p, so that there is a unique twisted form corresponding to the cocycle $\eta : F \mapsto w_p$, and this curve is not a Mumford curve either – its special fibre is irreducible over F_p. (Compare with page 118 of [Schmecta], which seems to be in error on this point.) Thus the behavior of classical modular curves at primes dividing the level is to be contrasted with the behavior of the curves exhibited in the remainder of this section.

Thus the useful notion for us is that of a twisted Mumford curve, the data for which is a Mumford curve X_Γ/Z_p and a twist $\alpha \in H^1(G(Q_p^\infty/Q_p), Aut(X_\Gamma/Z_p^\infty))$.

Remark: Compare with [Jordan-Livné I], who work with the notion of an *admissible curve*, which comes down to a *potential* Mumford curve. This is the suitable class of curves whose special fibres can be described by the combinatorial data of an ℓ-graph; see Chapter 4. Notice that $X_0(p)/Z_p$ is admissible.

Drinfeld's twisting: Drinfeld systematized this twsting process as follows: introduce the formal scheme $\mathcal{P}^\infty := \mathcal{P} \times_{Z_p} Z_p^\infty$ viewed as a formal scheme over Z_p(!) Whereas $\mathcal{P}/Z_p$ carries a natural action of $PGL_2(Q_p)$, we can equip $\mathcal{P}^\infty$ with an action of $GL_2(Q_p)$: namely $\alpha \in GL_2(Q_p)$ acts on (x, u) as $([\alpha]x, Frob^{-v_p(det\alpha)}u)$. This gets used as follows: let now $\Gamma \leq GL_2(Q_p)$ be a discrete cocompact subgroup containing a positive power of the scalar matrix p. The quotient $\Gamma\backslash\mathcal{P}^\infty$ can then be algebraicized and gives a twisted Mumford curve over Z_p. Indeed, say that $\alpha_n = p^n.1 \in \Gamma$ and n is minimal such that this occurs. Then α_n acts trivially on the first factor of $\mathcal{P}^\infty$ and as "translation by 2n" on the second factor, so that $\Gamma\backslash\mathcal{P}^\infty = \Gamma\backslash(\mathcal{P} \times Z_p^{2n})$. This last object is a finite-type Z_p-formal scheme with a possibly disconnected special fibre. Twisting will now come from nonscalar elements of Γ whose determinant has valuation indivisible by $2n$.

0.8.4 At last, the Cerednik-Drinfeld uniformization

We now return to the case of Shimura curves. Let $\mathcal{O} \leq B$ be an Eichler order in an indefinite rational quaternion algebra of discriminant D; let p be a prime dividing D, and write $\overline{G}$ for the unit group of the the definite quaternion algebra

of discriminant D/p. Motivated by our study of the supersingular isogeny class in the good reduction case (recall that *all* points are supersingular in characteristic p dividing D), let U be an adelic level structure which as maximal at p, and recall the isomorphism

$$G(A_f^p) \xrightarrow{\sim} \overline{G}(A_f^p)$$

already exploited in the proof of Corollary 73, and consider the p-adic space

$$Z_U := U\backslash\overline{G}(A_f)/\overline{G}(Q).$$

Notice that via a choice of isomorphism $GL_2(Q_p) = \overline{G}(Q_p)$, we have a natural action of $GL_2(Q_p)$ on Z_U. Now we have the main result of this section:

Theorem 77 *(Cerednik-Drinfeld) Let $X^D(U)/Z_p$ be the canonical integral model of the Shimura curve with level U structure. Then we have a canonical isomorphism of Z_p-formal schemes*

$$X^D(U)/Z_p \cong GL_2(Q_p)\backslash(\mathcal{P}^\infty \times Z_U).$$

Remark: We have not given such close attention to the moduli problem over Z_p. Since the right hand side is an algebraic formal scheme, it is acceptable for our purposes to take it as the *definition* of the canonical Z_p-model, and the merit of the Cerednik-Drinfeld theorem for us is the following

Corollary 78 *The Shimura curve $X^D(U)/Q_p$ has a canonical integral model whose special fibre is a twisted Mumford curve.*

Proof: We need to explain the appearance of twisted Mumford curves. Indeed, by the theory of algebraic groups, one knows that $GL_2(Q_p)$ has only finitely many orbits on our space Z_U (think of the quotient space as a zero-dimensional Shimura variety), and certainly each orbit contains an element x_i whose component at p is 1. One gets (again in analogy to Shimura varieties in characteristic zero) that the stablizer Γ_i of such an element is discrete and cocompact in $GL_2(Q_p)$ and accordingly contains a suitable power of the scalar matrix p. Let Γ_i' be the image in $PGL_2(Q_p)$ of the subgroup of Γ_i of elements of unit determinant. Then Γ_i' is a Schottky group and the Z_p-formal scheme in the corollary is isomorphic to a finite disjoint union of schemes of the form

$$\Gamma_i'\backslash(\mathcal{P} \otimes_{Z_p} Z_p^{2n_i}).$$

When U is small enough so that the complex curve $X^D(U)$ is connected, so is the special fibre, and we have ($i = 1$ and) a twisted Mumford curve.

Finally, in case $U = \Gamma_0(N)$ we record the following more explicit results which are needed in Chapter 4.

Let $B = B_D$ be an indefinite rational quaternion algebra, and let p be a prime dividing the discriminant D. Write $\overline{B}$ for the quaternion algebra obtained from B

by interchanging the local invariants p and $[\infty]$ (so that $\overline{B}$ is the *definite* quaternion algebra of discriminant D/p). Let $\mathcal{O} \leq \overline{B}$ be a level N Eichler order. We define subgroups $\widetilde{\Gamma_0}, \widetilde{\Gamma_+}$ of $GL_2(Q_p)$ as follows: $\widetilde{\Gamma_0} := (\mathcal{O} \otimes Z[1/p])^\times$, and $\widetilde{\Gamma_+}$ is the subgroup of $\widetilde{\Gamma_0}$ consisting of elements of whose determinant has even valuation; notice that $W = \{1, w_p\}$ is a set of coset representatives for $\widetilde{\Gamma_0}$ in $\widetilde{\Gamma_+}$, where w_p is any element of $\mathcal{O}$ of norm p. Also write $\Gamma_0 := \widetilde{\Gamma_0}/Z[1/p]^\times,\ \Gamma_+ := \widetilde{\Gamma_+}/Z[1/p]^\times$. The Cerednik-Drinfeld theorem then reads as follows in our case:

$$X_0^D(N)/Z_p = \widetilde{\Gamma_0} \backslash \mathcal{P}^\infty =$$

$$\Gamma_0 \backslash \mathcal{P} \times_{Z_p} Z_p^2 =$$

$$W \backslash (\Gamma_+ \backslash (\mathcal{P} \times_{Z_p} Z_p^2)),$$

so that $X_0^D(N)/Z_p$ is a Z_p^2/Z_p-twisted form of the Mumford curve $\Gamma_+ \backslash \mathcal{P}$ under the twist

$$Frob \mapsto w_p.$$

It is this fact which leads to a description of the special fibre which will be useful to us in Chapter 4.

0.9 The Integral Canonical Model III: Deligne-Rapoport reduction

In this section we discuss the reduction of one of our Shimura curves $X_0^D(N), X_1^D(N)$ at a prime dividing N. More precisely we consider the moduli problems of QM surfaces A/S equipped with a level N structure, where S is a $Z[1/D]$-scheme. This turns out to be significantly *easier* than the situation considered in the last section, since the theory of moduli of QM surfaces "away from characteristic dividing D" and with level structure prime to D is highly analogous to the analogous moduli problems in the elliptic modular case (i.e., with $D = 1$). That is, the special fibre at a prime p dividing N of $X_\bullet^D(N)$ has the same qualitative description of the special fibre at a prime dividing N of $X_\bullet(N)$, which is fortunate because the theory of arithmetic moduli of elliptic curves is very well-developed ([Deligne-Rapoport], [Katz-Mazur]). In fact it is easier because we do not need to worry about the modular interpretation of the cusps: every "generalized QM surface" is a QM surface.

0.9.1 Buzzard's work on "false elliptic curves"

In this section D is fixed and all schemes S are over $Z[1/D]$. The key fact that drives the analogy between QM-surfaces and elliptic curves "away from D" is the (already seen) fact that $\mathcal{O}_D \otimes Z_l \cong M_2(Z_l)$. Because of this, if N is prime to D, we have $\mathcal{O}_D \otimes Z/NZ \cong M_2(Z/NZ)$. We fix a compatible system of such

isomorphisms for all N prime to D and allow ourselves to pass between the left and right hand sides as equality without further comment. A *naive full level N structure* on a QM surface A/S is an isomorphism

$$\alpha : (\mathcal{O}_D \otimes Z/NZ)_S = M_2(Z/NZ) \xrightarrow{\sim} A[N]$$

which is compatible with the left-action of $\mathcal{O}_D$. The associated moduli problem, which takes A/S to the set of full level N-structures, is relatively representable by an étale (right) $(\mathcal{O}_D \otimes Z/NZ)^\times$-torsor on S. If now H is a subgroup of $(\mathcal{O}_D \otimes Z/NZ)^\times = GL_2(Z/NZ)$, we get an associated moduli problem by taking A/S to the H-orbits of full level N structures; this problem is likewise represented by an étale S-scheme, namely $(GL_2(Z/NZ)/H)_S$, just the same as for elliptic modular curves. In particular, one has notions of naive $\Gamma_0(N)$ and $\Gamma_1(N)$-structures, and the following analogous result to the elliptic modular case:

Theorem 79 *([Buzzard]) Let $N \geq 4$. Then the moduli problem of naive $\Gamma_1(N)$-structures on $\mathcal{O}_D$-QM abelian surfaces is representable in the category of $Z[1/ND]$-schemes by a smooth projective curve $X_1^D(N)/Z[1/DN]$. For all N, the moduli problem of naive $\Gamma_0(N)$-structures is coarsely represented by a smooth projective curve $X_0^D(N)/Z[1/DN]$.*

Sketch proof: This is quite formal – one knows that a solution to the moduli problem exsits as a stack; rigidity of $\Gamma_1(N)$-level structures with $N \geq 4$ then implies that the stack is associated to an algebraic space. But the morphism $X_1^D(N) \to Z[1/DN]$ is smooth, proper and 1-dimensional over a regular base, so it is a scheme [Knutson]. There is no problem deducing the result for $\Gamma_0(N)$, since we have only to take a quotient by a suitable finite group.

The reason that these structures are said to be naive is that they are *empty* unless N is invertible on S. One defines non-naive $\Gamma_0(N), \Gamma_1(N)$-structures over $Z[1/D]$-schemes using the notion of *cyclicity* as in [Katz-Mazur]. For this, we can exploit the isotypicality of the N-torsion of QM surfaces: if $G \leq_S A[N]$ is an $\mathcal{O}_D$-stable subgroup scheme, then G must split under the action of $\mathcal{O}_D \otimes Z/NZ = M_2(Z/NZ)$. More precisely, let e be the standard idempotent matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$; then $G = eG \oplus (1-e)G$. Then a $\Gamma_0(N)$-level structure is given by a rank N^2 $\mathcal{O}_D$-stable subgroup scheme G of $A[N]$ such that the rank N subgroup scheme eG has, fppf-locally on the base, a *generator* (cf. Chapter 1 of [Katz-Mazur]). Equivalently, we can give a $\Gamma_0(N)$-structure by a QM-isogeny $A \to A'/S$ (so that the kernel is an $\mathcal{O}_D$-stable subgroupscheme) of degree N^2. Similarly, a $\Gamma_1(N)$-level structure is the data of a $\Gamma_0(N)$-level structure together with a choice of generator.

Theorem 80 *(Buzzard) The moduli problem of $\Gamma_1(N)$ structures on QM surfaces over $Z[1/D]$-schemes extends the above naive moduli problem. When*

$N \geq 4$, $X_1^D(N)/Z[1/D]$ exists a fine moduli space. For all N, $X_\bullet^D(N)/Z[1/D]$ exists as a coarse moduli space.

Sketch proof: For the extension part of the theorem, we must check that over $Z[1/DN]$-schemes we have isomorphic functors, namely if we have $A/S/Z[1/DN]$ we must show that the naive $\Gamma_\bullet(N)$ structures are functorially in bijection with the non-naive $\Gamma_\bullet(N)$-structures. Both are étale sheaves on S, so (by passing to a surjective étale cover) we may assume that $A[N] \cong ((Z/NZ)^4)_S$. If α is a naive $\Gamma_\bullet(N)$-structure then choose β a naive full level N-structure lifting α. The bijection is obtained via the applying β to an appropriate subgroup: in the $\Gamma_0(N)$-case it is $\begin{bmatrix} 0 & 0 \\ 0 & * \end{bmatrix}$; in the $\Gamma_1(N)$-case it is $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Why does $X_\bullet^D(N)$ exist as a coarse moduli scheme at all? It is convenient to impose an additional rigidifying level structure U of level M_U prime to DN, so that by the previous theorem $X^D(U)/Z[1/M_U]$ exists as a fine moduli scheme. If we can show that the composite moduli problem $(\Gamma_\bullet(N), U)$ admits a solution as a fine moduli scheme over $Z[1/M_U D]$, then by standard stuff we'll get $X_\bullet^D(N)/Z[1/D]$. But observe that the moduli problem $\Gamma_\bullet(N)$ is *relatively* representable – this is true for $\Gamma_0(N)$ because it is a closed subscheme of a Grassmannian, and true for $\Gamma_1(N)$ by ([Katz-Mazur], Proposition 1.9.1). And one knows that the composite of a relatively representable moduli problem and a representable moduli problem is representable (Proposition 4.3.4 of [Katz-Mazur]).

Remark: Of course the same argument – with some additional attention at the cusps – also works in the $D = 1$ case.

Theorem 81 *(Buzzard+Katz-Mazur)*
a) The scheme $X_\bullet^D(N)/Z[1/D]$ is connected, proper and smooth away from the supersingular points in characteristics dividing N.
b) The modular forgetful map

$$c : X_\bullet^D(N) \to X^D$$

is finite flat.
c) The special fibre of $X_0^D(p^r)$ in characteristic p has the following more precise description: it has as irreducible components $a+1$ nonsingular curves, each (non-canonically) isomorphic to X^D, intersecting at the supersingular points. When $a = 1$ the intersection is transverse (the local ring at a supersingular point of the total space is analytically isomorphic to $Z_p[x,y]/(xy - p^n)$ for some $n \leq 3$).
d) The special fibre of $X_1^D(p^r)$ in characteristic p has the following more precise description: it has as irreducible components $a+1$ curves; index them as $C_0, \ldots, C_a$. All but one of these curves is nonreduced: indeed C_i has multiplicity $\phi(p^i)$ and its underlying abstract curve is (non-canonically) isomorphic to the level p^{n-i}-Igusa curve. The components intersect transversally at the supersingular points.

Remark: We will review Igusa level structures in the course of the proof.

Remark: We have not given a *completely* precise description of the singular fibres, since we have not explained which supersingular points to glue to which. In general, a supersingular point on one component is glued to a suitable F^k-conjugate of the corresponding Frobenius point on another component. In the case of $X_0^D(p)$ we just glue $P \mapsto FP$; the recipe in the general case does not concern us here.

Proof: The properness follows immediately from the potentially good reduction of QM surfaces together with the fact that if A/R is a QM surface over a discrete valuation ring R then a level N-structure on the generic fibre extends uniquely to a level N structure on A/R (since A/R is the Néron model of its generic fibre). Since the morphism is proper, the finitness can be checked on geometric fibres, i.e., we need only show that if A/k is a QM surface over an algebraically closed field, then there are only finitely many $\Gamma_\bullet(N)$ structures on A/k. When the characteristic does not divide N, there are the same number of $\Gamma_\bullet(N)$ structures as in characteristic 0. When the characteristic p divides N, assume for simplicity that $N = p^r$ (we can certainly reduce to this case). We have $A[p^r] \cong G \times G$, and we want to give a rank p^r-cyclic subgroup scheme of G. In the ordinary case, $G \cong Z/p^rZ \otimes \mu_{p^r}$, and there are $r+1$ such subgroup schemes – factor $p^r = p^a p^b$ and take an étale group scheme of rank p^a and a multiplicative group scheme of rank p^{r-a}. In the supersingular case, $G \cong \alpha_{p^{2r}}$, and the only rank p^r subgroup scheme is α_{p^r}. This shows the finiteness of $\Gamma_0(N)$-structures. In the $\Gamma_1(N)$ case it is similarly clear that there are only finitely many generators in each case.

The next step is to check all the assertions of the theorem except those that can be checked locally at each supersingular point: namely we want to see that the ordinary locus is smooth and that it has the precise description given in parts c) and d) of the theorem. For the remainder of the proof we assume $N = p^r$.

Consider first the $\Gamma_0(p^r)$ case. For $0 \leq i \leq r$, let $\mathcal{A}_i$ be the functor from $Z[1/D]$-schemes to sets sending S to the set of QM surfaces equipped with an $\mathcal{O}_D$-invariant subgroup scheme G which is étale-locally isomorphic to $(Z/p^iZ \times \mu_{p^{r-i}})^2$. The limiting cases $\mathcal{A}_r, \mathcal{A}_0$ classify "fully étale" and "fully multiplicative" subgroup schemes, respectively. We claim that each of the $\mathcal{A}_i$'s are naturally isomorphic as functors. Indeed, we can go from $\mathcal{A}_0 \mapsto \mathcal{A}_1 \mapsto \ldots \mathcal{A}_r$ by at each stage taking $(A, G) \mapsto (A/G', A[p^r]/G')$, where $G' \leq G$ is the unique subgroup scheme locally isomorphic to $(Z/pZ)^2$. Now $\mathcal{A}_0$ is representable (a special case of the Hilbert scheme), so all the $\mathcal{A}_i$'s are. We can check smoothness fibrewise, the only questionable fibre being p. But in characteristic p, an $\mathcal{A}_i$ is isomorphic as a functor to the functor coarsely represented by $(X^D)^{ord}$ – that is to say, the forgetful map which takes a pair $(A/S, G_i)$, $A/S/F_p$ an ordinary QM surface and $G_i \leq A[p^r]$ a cyclic subgroup scheme of type i to A/S is an isomorphism, since each ordinary QM surface has a unique cyclic structure of a

given type. This concludes the description of the ordinary locus of the special fibre of $X_0^D(N)$.

In the $\Gamma_1(p^r)$-case the proof is similar; one considers the modified functors $\mathcal{B}_i$ from $Z[1/D]$-schemes to sets which takes S to isomorphism classes of QM surfaces A/S equipped with a *generator* of the cyclic structure corresponding to $\mathcal{B}_i(S)$. However, in contrast to the previous case, the various $\mathcal{B}_i$'s are not isomorphic to each other because a finite multiplicative group scheme of any rank has a unique generator while a finite étale group scheme of rank p^a has $\phi(p^a)$-generators. What we find instead is that the functor $\mathcal{B}_i$ is isomorphic to the ordinary subfunctor of the moduli problem of p^a-Igusa level structure, which we now describe.

Definition: A level p^a-Igusa structure on a QM surface $A/S/F_p$ is given by a generator of $\mathrm{e}(\ker V^a)$, where e is our standard idempotent matrix and $V^a : A^{p^a} \to A$ is the a-fold Verschiebung. Note well that one of the characterizations of the ordinary/supersingular dichotomy for QM surfaces (and equally for elliptic curves) in characteristic p is that A is ordinary if and only if the kernel of (any $\iff$ all) V^a is an étale subgroup scheme of $A[p^a]$.

Basic properties of p^a-Igusa curves: that the moduli problem is relatively representable and flat follows immediately from the "Main Theorem on Cyclic Groups" ([Katz-Mazur], Theorem 6.6.1). To see that the ordinary locus is smooth is easy: indeed $X^D(Ig(p^a)) \to X^D$ it is an étale $(Z/p^aZ)^\times$-torsor. In more concrete language, the natural map

$$\varphi : X^D(Ig(p^a)) \to X^D$$

is flat and unramified of degree $\phi(p^a)$ away from the supersingular points. It is also smooth at the supersingular points – this is one of the local statements whose proof we deal with a little later. In any event, what we have recalled already about Igusa curves is enough to see that the ordinary locus of the special fibre of $X_1^D(p^r)$ is smooth but reducible with components as described in the proof of the theorem.

Definition: Fix a prime number p. A level p^N-Igusa structure on a QM-surface A/S – where S is now an F_p-scheme – is given by a generator of $e \ker V^N$, where $V^N : A^{p^N} \to A$ is the N-fold Verschiebung.

Let us summarize where we are in the proof: we have seen that the structure of the special fibre of $X_\bullet^D(p^r)$ is as we've claimed except possibly at the supersingular points, where we still must check the following: that the map $X_\bullet^D(p^r) \to X^D$ is flat at the supersingular points, that over each supersingular point of X^D we see all $r+1$ components intersecting transversally, and finally, for the Γ_1 case, that the p^a-Igusa curve is indeed smooth over each supersingular point of X^D. All of these statements can be checked on the *completed local rings*

of the curves involved. Because of this, their truth can be deduced directly from the truth of the corresponding statements for the $D = 1$ case, as we now explain.

We need some deformation theory: let $E/\overline{F_l}$ be a supersingular elliptic curve and $A/\overline{F_l}$ be a supersingular QM surface. Let W be the ring of Witt vectors of $\overline{F_l}$, let $E/W[[T]]$ be the universal formal deformation of E (to Artin local $\overline{F_l}$-algebras) and let $A/W[[T]]$ be the universal formal deformation of A. By ([Buzzard], Corollary 4.6(ii)), we have an $\mathcal{O}_D$-module isomorphism $A[p^\infty] \cong E[p^\infty]^2$ (this follows almost immediately from the Serre-Tate theorem). Now let $x \in X^D(U, Ig(p^N)(\overline{F_p})$ be a supersingular point, with corresponding QM surface $A/\overline{F_p}$ and similarly $y \in X(U, Ig(p^N))(F_p))$ a supersingular point with corresponding elliptic curve $E/\overline{F_p}$ – here U is a rigidifying level structure of level M_U prime to DN. Let S_x be the spectrum of the completed local ring of $X^D(U, Ig(p^N))$ at x and S_y the spectrum of the completed local ring of $X(U, Ig(p^N))$ at y. Then S_x is the scheme relatively representing $(U, Ig(p^N))$-structures on the universal deformation A, and S_y is the scheme relatively representing the same structures on the universal deformation E. So S_x is isomorphic to S_y.

That's the trick: the theories of moduli of Shimura curves and moduli of elliptic modular curves in characteristic p not dividing D and with level structure away from D are more than "analogous": there are canonical (given our choices at the beginning of this section) isomorphisms between the completed local rings of the one and the other. These isomorphisms allow us to "cheat" by transporting what are in some cases rather hard-earned theorems from the elliptic modular case to our Shimura context. In particular, the $D = 1$ analogues of all our local statements at the supersingular points are proved in wonderful detail in [Katz-Mazur], so they hold true for us. This completes the proof of the theorem.

0.9.2 A genus formula for rigidified Igusa-Shimura curves

As a further application of this "cheating," we will compute the genus of the rigidified Igusa curve $X_N/F_p := X^D(Ig(p^N), \Gamma_1(L))$ – here $L \geq 4$. The point is that for covers of degree divisible by p in characteristic p, we must apply Riemann-Hurwitz carefully: the degree of the ramification divisor at a point can equal or exceed the degree of the cover (wild ramification), so knowing that the forgetful map

$$\varphi : X_N \to X^D(\Gamma_1(L))$$

has degree $\phi(p^N)$ and is étale on the ordinary locus and totally ramified at each supersingular point is, when $N > 1$, *not* enough to tell us the genus. But we can cheat: the ramification will be the same as in the elliptic modular case, making the formula easy to compute. Here are the details: Riemann-Hurwitz reads

$$2g_N^D - 2 = \phi(p^N)(2g_0^D - 2) + \sum_{P \in ss} \deg R_P$$

where g_N^D, g_0^D are the genera of $X_N, X_1^D(L)$ respectively, and R_P is the ramification divisor at a supersingular point $P \in X_N$, i.e.,

$$R_P = length(\Omega_{X_N/X_1^D(N)})_P \; P.$$

This quantity can be computed with respect to the *completed* local rings, which allows for a key observation: the degree of the ramification divisor over any supersingular point is equal to the degree of the ramification divisor at a supersingular point for the classical covering $I_N(\Gamma_1(L)) \to X_1(L)$. Indeed, if $\tilde{x} \mapsto x$ are supersingular points on the map φ in the Shimura case, and if $\tilde{y} \mapsto y$ are supersingular points of the classical $D = 1$ Igusa covering, then denoting e.g. $S_{\tilde{x}}$ for the spectrum of the completed local ring at x, then the last parargraph of Section 0.9.1 leads to a commutative diagram

$$[ccc] S_{\tilde{x}} e, t \sim s S_{\tilde{y}} s S_x e, t \sim S_y$$

and we conclude that the ramification divisor at $\tilde{x}/x$ has the same degree as the ramification divisor of $\tilde{y}/y$. Notice that it also follows from this discussion that the degree of this ramification divisor does not depend upon the choice of supersingular point, either in our Shimura case or in the elliptic modular case; denoting this common degree by R and the number of supersingular points on X_N (which is the same as the number of supersingular points on $X_0 = X_1^D(L)$) as S^D, we get

$$2g_N^D - 2 = \phi(p^N)(2g_0^D - 2) + RS^D.$$

As for the common degree R, its value is implicit in the formulas given in [Katz-Mazur], since we have:

$$2g_N - 2 = \phi(p^N)(2g_0 - 2) + RS$$
$$p^N \phi(p^N) \deg(\omega) = 2g_N - 2 + \phi(p^N) c(\Gamma_1(L))$$
$$2 \deg(\omega) = 2g_0 - 2 + c(\Gamma_1(L))$$

where now g_N, g_0 are the genera of the classical curves $I_N(\Gamma_1(L))$ and $X_1(L)$; the quantities $c(\Gamma_1(L)), \deg(\omega)$ – defined in [Katz-Mazur] – cancel out, and we get

$$R = p^{N-1}(p^N - 2).$$

So

$$2g_N^D - 2 = \phi(p^N)(2g_0^D - 2) + p^{N-1}(p^N - 2)S^D.$$

On the other hand, from [Diamond-Taylor] we have

$$2S^D = (p-1)(2g_0^D - 2)$$

Substituting this in we get

Proposition 82 *Let D, L, p be pairwise coprime with $L \leq 4$, and let*

$$X_N := X^D(Ig(p^N), \Gamma_1(L))/F_p.$$

Write g_N for the genus of X_N and S^D for the number of (geometric) supersingular points on $X_1^D(L)$. Then we have

$$2g_N^D - 2 = p^{N-1}(p^N - 1)S^D.$$

Chapter 1

Moduli spaces of potentially QM surfaces

1.1 PQM surfaces

We maintain the notation of Chapter 0; especially $D > 1$ is the discriminant of a nonsplit indefinite rational quaternion algebra. Recall from Section 1.5 that X^D has no real points; a fortiori it has no rational points. It would seem to follow that the existence question for QM surfaces A/Q is settled in the negative – and of course, this is true, in the sense we have defined QM surfaces in Chapter 1. However, this is not necessarily the sense that is the most natural or interesting! Indeed, a moduli point in $X^D(K)$ comes from a structure (A, ι, P) where A/K is an abelian surface, $\iota : \mathcal{O}_D \to End_K(A)$: that is, it is part of the moduli problem that all the QM endomorphisms be defined over the field K. In the same section, we showed that an abelian surface never has a subring of endomorphisms isomorphic to $\mathcal{O}_D$ defined over the real numbers. We also recalled an analogous fact about CM elliptic curves: there does not exist a pair $(E, \iota)/R$ where E/R is an elliptic curve and $\iota : \mathcal{O}_K \hookrightarrow End_R(E)$ a subalgebra isomorphic to the maximal order in a CM quadratic field (the proof works for arbitrary orders). Clearly if we were to interpret this statement as telling us that there do not exist CM elliptic curves over Q, we would be missing out on very interesting geometric objects: of course there do exist elliptic curves E/Q with $\mathcal{O}_K$-CM (when K has class number 1) defined over a larger field (in fact, over K). Since CM elliptic curves much predate the formalism of moduli spaces, it did not happen that elliptic curves E/Q with "potential CM"' were excluded from study. Yet, to a large extent, this is what has happened for QM abelian surfaces.

So, we propose that our basic object of study should be a principally polarized abelian surface (A, P) over a field K which admits an $\mathcal{O}_D$-QM structure over the separable algebraic closure $\overline{K}$: there exists $\iota : \mathcal{O}_D \to End_{\overline{K}}(A)$. We call

this data a potentially quaternionic multiplication (PQM) surface A/K. The point is that there may well exist $\mathcal{O}_D$-PQM surfaces A/Q – indeed, the square of a CM elliptic curve E/Q will give such an example for all B_D split by the CM field K. More interesting then are geometrically simple PQM surfaces A/Q – an early example (possibly the first) of such a surface was given by Koike using modular forms. It turns out that the existence of such objects as a function of D, far from being ruled out a priori, is a deep question, explored (but by no means settled) in Chapter 2 of this thesis.

1.2 The $\mathcal{O}_D$-locus: travaux de Victor Rotger

Having defined PQM surfaces, we may ask: how are they related to Shimura curves? To make the question more precise, we can define $\mathcal{L}_D \subset \mathcal{A}_2/Q$ to be the locus of principally polarized abelian surfaces admitting geometric $\mathcal{O}_D$-QM. It is classical that $\mathcal{L}_D$ is a closed, one-dimensional subvariety of $\mathcal{A}_2$, loosely called a Shimura curve in the literature, but this is not quite correct. As we are about to explain, it is in fact *never* the case that $\mathcal{L}_D \cong_Q X^D$; moreover $\mathcal{L}_D$ will in general have several irreducible components. The precise relationship between $\mathcal{L}_D$ and X^D has been determined very recently by Victor Rotger. The key notion we need is that of a modular forgetful map. Indeed, recall that in order to interpret X^D as the moduli space for triples (A, ι, P), we chose a piece of auxiliary data $\mu \in \mathcal{O}_D, \mu^2 + D = 0$. The choice of μ enabled us to define a positive involution on B_D and thus gave a notion of compatiblity between ι and P. Given μ, there exists a unique principal polarization compatible with the QM structure. Therefore, dependent on μ, forgetting the QM structure: $(A, \iota, P) \mapsto (A, P)$ induces a morphism $F_\mu : X^D \to \mathcal{A}_2$, a forgetful modular map.

To describe Rotger's results we need some terminology. First, we call the pair $(\mathcal{O}_D, \mu)$ a principally polarized (maximal) order of B_D. Secondly, a nonzero element $\chi \in \mathcal{O}_D \cap N_{B^\times}(\mathcal{O}_D)$ is called a twist of $(\mathcal{O}_D, \mu)$ if χ is a pure quaternion (i.e., $t(\chi) = 0$, or equivalently $\chi^2 = -n(\chi)$) and $\mu\chi = -\chi\mu$. Notice that then $B_D \cong (\frac{-D,-n(\chi)}{Q})$. We say that $(\mathcal{O}_D, \mu)$ is twisting if it admits a twist by some χ, and that B_D itself is twisting if some principally polarized order is twisting. It is immediate that B_D is twisting if and only if $B_D \cong (\frac{-D,m}{Q})$ for some positive integer $m|D$. We also say that D admits a twist by m in this case.

Remark: D admits a twist by m if and only if D admits a twist by D/m (at every place v of Q, we have $(-D, m)_v(-D, D/m)_v = (-D, D)_v = 0$), and we will soon see that these twists are essentially the same. In general (if D is divisible by more than two primes), D can admit essentially different twists: e.g. if $\{p_1, \ldots, p_{2n}\}$ is an even cardinality set of primes such that for distinct i, j, $(\frac{p_i}{p_j}) = -1$, $D = p_1 \cdots p_{2n}$ admits twists by each p_i. Now, associated to $(\mathcal{O}_D, \mu)$ we define a subgroup of Atkin-Lehner involutions H_μ as follows: H_μ is generated by the main Atkin-Lehner involution w_D and all w_m such that

$(\mathcal{O}_D, \mu)$ admits a twist by a character χ of norm $-m$. Now we have:

Theorem 83 *([Rotger II-IV])*

a) If $(\mathcal{O}_D, \mu)$ is nontwisting, $H_\mu = \langle w_D \rangle$. If it is twisting, it admits an essentially unique twist: $H_\mu = \langle w_D, w_m \rangle$.

b) The forgetful maps F_μ are finite morphisms. More precisely, F_μ factors through X^D/H_μ and then gives a closed embedding

$$F_\mu : X^D/H_\mu \hookrightarrow \mathcal{A}_2.$$

c) The $\mathcal{O}_D$ locus $\mathcal{L}_D$ is obtained as the union of the images of the F_μ ranging over the finite set of $\mathcal{O}_D$-conjugacy classes of elements μ, $\mu^2 + D = 0$.

Let us now discuss some implications. First consider the simpler case where the entire quaternion algebra B is nontwisting. Then for each μ we find $F_\mu : X^{D+} \hookrightarrow \mathcal{A}$. Even in this favorable case it is not literally true that X^{D+} is the coarse moduli space for PQM abelian surfaces, since we will in general need several μ's to cover $\mathcal{L}_D$. Otherwise put, distinct points in $\mathcal{A}$ may correspond to the same point in X^{D+}: indeed, reflecting that the choice of μ is required only to define the polarization in the triple (A, ι, P), it follows that the ambiguity is precisely that we may have multiple W-orbits of principal polarizations on the same abelian surface (in this regard, it is useful to mention another theorem of [Rotger II-IV] which gives the Néron-Severi group of complex QM surface A as a certain group of pure quaternions of B; with this identification, the Atkin-Lehner group acts on the set of principal polarizations; elements in the same orbit correspond to the W-orbit of a point on the Shimura curve X^{D+}). In particular, it is true (but not very useful) that the more drastic forgetful map $F : (A, \iota, P) \mapsto A$ is surjective onto the *set* of principally polarizable PQM abelian surfaces. Nevertheless, in the nontwisting case, X^{D+} is "as good as" a coarse moduli space for PQM surfaces:

Corollary 84 *Assume D is a nontwisting discriminant. Then if $X^{D+}(K)$ is empty (resp. consists only of CM points), then there does not exist a PQM abelian surface A/K (resp. a geometrically simple PQM surface A/K).*

This is immediate. Moreover, we can use Jordan's theorem to say more precisely what kind of a point on $X^{D+}(K)$ will correspond to (at least one) PQM surface A/K; we take up this problem (along with the case of level structure) in the next section.

If on the other hand D is a twisting discriminant, then in general $\mathcal{L}_D$ will have some irreducible components isomorphic to X^{D+} and others isomorphic to X^D/H_μ, a further two-fold involutory quotient (and in general there will be more than one such μ, corresponding to the number of essentially different twists m of D; at least in the case $D = pq$ we don't have to worry about this). Since the second situation dominates the first, we conclude:

Corollary 85 *Assume D is a twisting discriminant and let $\{m_i\}$ be the set of essentially different divisors of D such that $(\frac{-D,m_i}{Q}) \cong B_D$. Then if for all i, $X^D/\langle w_D, w_{m_i}\rangle(K)$ is empty (resp. contains only CM points), there does not exist a PQM surface A/K (resp. a geometrically simple PQM surface A/K).*

In fact our work on Shimura curves centers around $X^{D+}, X_0^{D+}(N), X_1^{D+}(N)$. When D is nontwisting, this is appropriate for studying PQM surfaces, as just seen. In case D is twisting, our nonexistence results on $X_0^{D+}(N)(K)$ do not therefore preclude the existence of a PQM surface A/K, but only of such a surface whose QM becomes defined over a quadratic extension of K (rather than a biquadratic extension). Either way, let us call a QM surface A/K corresponding to a K-rational point on $X^{D+}(K)$ a PQM surface *of plus type*, and a PQM surface which is not of plus type *of (2,2)*-type. The terminology is, hopefully, explained by the following

Corollary 86 *Let A/F be a PQM abelian surface over a number field F. If A is of plus type, there exists a unique minimal extension K/F, at most quadratic, such that A/K admits a QM structure compatible with its polarization. Moreover, if F is real, K/F is necessarily nontrivial. If A/F is a PQM surface of $(2,2)$-type, there exists a unique minimal extension K/F, at most biquadratic, over which A admits a compatible QM structure, necessarily nontrivial if F is real.*

Proof: Indeed, by the Shimura curve geometry we have just surveyed, we know that a PQM A/F induces a point on an Atkin-Lehner quotient of degree 2 or 4. Taking the preimage we get a K-divisor on X^D of degree 2 or 4; the Galois group of the splitting field of this divisor is naturally a subgroup of the group H_μ of involutions. The only point which is not immediate is to see that one of these preimage points on X^D, which corresponds to a QM surface with field of moduli contained in K, can actually be defined over K. For this: let $B_D \leq End_C^0(A)$ be the QM subalgebra (this is well-defined, because a point on X^D/H corresponds to a QM-structure up to twisting by Atkin-Lehner elements, but the image of the QM-structure is invariant). B_D is stable under Galois, so there exists a unique minimal extension L/F cut out by the action of Galois on B_D. Let $M_1, M_2/K$ be two distinct quadratic extensions splitting B_D. By Jordan's theorem, M_1, M_2 are acceptable fields of definition for A as QM-surface, so $K \leq L \leq M_i$ for $i = 1, 2$ and we conclude $K = L$.

In fact there are good reasons to prefer PQMs of plus type. The following proposition, together with the generalized Taniyama-Shimura conjecture (already proved by [Ellenberg] in some special cases relevant to us), implies that A/Q a plus-type PQM surface is modular, i.e., isogenous to a Q-factor of $J_1(N)$.

Proposition 87 *Let A/Q be a simple B_D-PQM surface of plus type. Then $End_Q^0(A)$ is a quadratic field.*

Proof: By the preceding corollary, A admits the structure of a compatible QM surface over an imaginary quadratic extension L/Q. We thus have $End_{\overline{Q}}(A) =$

$End_L(A) = B$. The (analytic) representation of B on the complex cotangent space of A is faithful and Galois equivariant: viewing $\overline{Q} \hookrightarrow C$, the field of definition of any endomorphism of A is the same as the field of definition of its matrix coefficients in the analytic representation. So in particular we have $B \hookrightarrow M_2(L)$, and what we are trying to show is that some non-central endomorphism of B is defined over Q. That is, what we must show is that $M_2(Q) \cap B \subset M_2(L)$ is strictly larger than Q. This is easily seen as follows: view $M_2(Q), B$, and $M_2(L)$ as linear spaces over Q of dimensions 4,4 and 8. Notice that B and $M_2(Q)$ both lie in the subvariety V of $M_2(L)$ given by matrices whose trace and determinant lie in Q. This V is a smooth, 6-dimensional Q-subvariety of $M_2(L)$ containing the origin, so checking tangent spaces at the origin reveals that the intersection of the two linear spaces B and $M_2(Q)$ must have dimension at least 2.

Remark: [Rotger II-IV] actually shows more: that if A/Q is a nontwisting $\mathcal{O}_B$-PQM surface, then $End^0_Q(A)$ is an imaginary quadratic field.

1.3 Technical lemmas on moduli of PQM abelian surfaces with level structure

In this section – the technical core of the thesis – we intepret $X^{D+}, X_0^{D+}(N)$ and $X_1^{D+}(N)$ in terms of moduli of plus type $\mathcal{O}_D$-PQM surfaces with additional level structure. We also prove results about *moduli points* on these curves, i.e., we give the criterion for a point on $X_0^{D+}(N)(K)$ to be induced from a structure defined over K (rather than merely having field of moduli contained in K). As a starting point, recall from the last section that X^{D+} is the moduli space for $\mathcal{O}_D$-PQM abelian surfaces of plus type.

Proposition 88 *The curve $X_0^{D+}(N)$ is the moduli space for structures (A, P, Q_N), where Q_N is an $\mathcal{O}_D$-stable submodule of $A[N]$, cylic as $\mathcal{O}_D$-submodule and isomorphic to $Z/NZ \oplus Z/NZ$ as abelian group.*

Proof: In other words, the claim we are making is precisely that the data is the same as for $X_0^D(N)$ except we have forgotten the QM structure. In fact this is essentially immediate from the moduli interpretations of $X_0^D(N)$ and of the main Atkin-Lehner involution w_D. However, for later use we want to give an interpretation of the 2-1 map $(A, \iota, P, Q_N) \mapsto (A, P, Q_N)$ in terms of the N-torsion. Namely, since $(N, D) = 1$, $\mathcal{O}_D$ acts as endomorphisms on $A[N]$ through $\mathcal{O}_D \otimes Z/NZ \cong M_2(Z/NZ)$. Indeed, by choice of an idempotent $e \in \mathcal{O}_D \otimes Z/NZ$ we may decompose $Q_N = C_1 \oplus C_2$, where $C_1 = eA[N]$, $C_2 = (1-e)A[N]$, an instance of Morita equivalence; each C_i is a cyclic group of order N. Thus, with the QM defined, the data of Q_N is equivalent to the data of $C_1 = eQ_N$ (this moduli interpretation of $X_0^D(N)$ can be found in the literature). But we can take $w_D = \begin{bmatrix} 0 & -D \\ 1 & 0 \end{bmatrix}$ and check that $w_D^{-1} e w_D = (1-e)$, so that the Atkin-Lehner involution carries C_1 to C_2. It follows that, being in a state of twofold ambiguity as to the QM structure as we are on $X_0^{D+}(N)$, we cannot define the subgroup

C_1 by itself but only the pair $\{C_1, C_2\}$, and from this the full submodule $C_1 \oplus C_2$.

In the case of $X_1^{D+}(N)$ our moduli interpretation requires no justification:

Proposition 89 *The curve $X_1^{D+}(N)$ is the moduli space for structures (A, P, x_1, x_2), where $\langle x_i \rangle = Q_i$ and $w_D(x_1) = x_2$.*

We will now study when points $P \in X(K)$ are induced by a structure defined over K. Our point of departure is Jordan's theorem: let L be a field containing the field of moduli of (A, ι, P) as QM surface. Then this structure can be defined over L if and only if L splits B_D. Now fix $P \in X^{D+}(K)$. Let L be the splitting field of divisorial preimage of P in X^D, so as we have seen, L/K is either trivial or quadratic.

Proposition 90 *The point $P \in X^{D+}(K)$ is a moduli point, i.e., is induced by a PQM-abelian surface defined over K if and only if L splits B.*

Proof: Since we saw in the last section that if A can be defined over K as PQM-surface, (A, ι) can be defined over L as QM-surface, so by Jordan's theorem the necessity is clear. As for the sufficiency, the hypothesis together with Jordan's theorem implies that A can be defined over L as QM surface. We want to show that, as polarized abelian surface, the base field can be descended to K. But indeed the group $H = \langle w_D \rangle$ provides descent data for L/K: $w_D(A) = \sigma(A) \cong A$ as polarized abelian surface.

Theorem 91 *Let $U \leq GL_2(Z/NZ)$ be an arbitrary subgroup. Let $X^D(U)$ be the corresponding Shimura curve (with level U structure). Let $(A, \iota, P, \phi)/\overline{Q}$ be a U-structured $\mathcal{O}_D$-QM surface with field of moduli contained in L. Then this structure can be defined over L if and only if L splits B_D.*

Remark: This statement is the Shimura curve analogue of the famous "surjectivity" of the moduli problem of elliptic curves with level U structure: i.e., any point $P \in X^1(U)(L)$ is induced by at least one U-structured elliptic curve E/L; we are indeed about to copy the proof of [Deligne-Rapoport].

Proof: Let us abbreviate (A, U) for our U-structured QM-surface. If there exists $(A, U)/L$, then all such are given by the cohomology set $H^1(G_L, Aut((A, U)/\overline{L}))$ – indeed, since the QM-automorphism group of a QM-abelian surface is always abelian (even isomorphic to μ_2, μ_4, or μ_6, and necessarily to μ_2 if it is non-split) – this is a cohomology group. Moreover, the obstruction to the existence of an L-structure lies in $H^2(G_L, \mu_n)$ for $n = 2, 4, 6$. Write $\mu_m = Aut(A)$. There is a natural map $H^2(G_L, \mu_n) \to H^2(G_L, \mu_m)$ which is induced by the inclusion $\mu_n = Aut(A, U) \hookrightarrow \mu_m = Aut(A)$. By Jordan's theorem the obstruction vanishes when mapped to $H^2(G_L, \mu_m)$. We finish with the observation that $H^2(G_L, \mu_n) \to H^2(G_L, \mu_m)$ is injective! (Indeed, take cohomology of $1 \to \mu_n \to \mu_m \to \mu_{m/n} \to 1$; it is enough to show $H^1(\mu_m) \to H^1(\mu_{m/n})$ is surjective; but by Hilbert 90 we are looking at the map $L^\times/L^{\times m} \to L^\times/L^{\times m/n}$, i.e., a quotient map.)

If P is a K-valued point on any plus-quotient Shimura curve (with level structure), let P' be its image in $X^D(K)$. The canonical field of P is by definition the splitting field of the degree 2 divisor which is the preimage of P' in X^D – its compositum with K is an at most quadratic extension.

Corollary 92 *Let $P \in X_0^{D+}(N)(\overline{K})$ be a point with field of moduli contained in K. Then it is induced by some structure $(A, P, Q_N)/K$ if and only if the field $M = LK$ splits B_D, where L is the canonical field of P.*

Proof: This is immediate by the theorem and our earlier analysis of the moduli problem $X_0^{D+}(N)$: indeed we have just seen that (A, ι, P, Q_N) can be defined over M. Decompose $Q_N = C_1 \oplus C_2$, where $C_i = e_i Q_N$. We know that the main involution w_D interchanges C_1 and C_2, hence the nontrivial automorphism σ of $G_{L/K}$ does this as well. We thus find that σ preserves Q_N.

Remark: Let A/Q be a PQM of plus type. Then the proof of the corollary rules out the existence of a certain type of cyclic order N subgoup defined over Q (namely, the one that generates an $\mathcal{O}_D$-module of rank N^2). It is easy to give bounds on N in terms of D for the existence of the other type of order N subgroup defined over Q, whence we can get an (effective) bound on cyclic order N subgroups. We carry out this argument in detail in Section 4.5.

Chapter 2

Shimura curves with infinitely many rational points

2.1 Introduction

We are interested in studying the locus $X^D/H/Q$ of rational points on Atkin-Lehner quotients of Shimura curves with no level structure. Proposition 65 supplies rational CM points on many of these curves. Nevertheless, the congruence conditions necessary for the existence of rational CM points are not satisfied on an infinite (positive density) family of Shimura curves, leaving open the possibility that $X^D/H(Q) = \emptyset$ for these curves. With the respect to the quotient by the main Atkin-Lehner involution, we make the following

Conjecture 93 *For all sufficiently large D, $X^{D+}(Q)$ consists entirely of CM points. In particular, $X^{D+}(Q) = \emptyset$ for infinitely many D.*

One obstacle to an easy proof is provided by our Main Theorem 2, which tells us that $X^{D+}(A_Q)$ is nonempty – there are no local obstructions.

An immediate consequence of the conjecture would therefore be:

Conjecture 94 *There exist infinitely many discriminants D such that X^{D+}/Q violates the Hasse principle: it has points at every completion of Q but no Q-points.*

This latter conjecture is in turn related to a much more general conjecture about abelian varieties:

Conjecture 95 *(Finiteness conjecture for endomorphism algebras) For any positive integer d, there exist only finitely many isomorphism classes of semi-simple algebras arising as endomorphism algebras of principally polarized abelian varieties A/Q of dimension d.*

Remarks: a)There some plausible variants of this conjecture: we may conjecture a finiteness result for abelian varieties defined over any fixed number field K instead of Q, or even uniformly over number fields of bounded degree. We could also drop the requirement that the abelian varieties be principally polarizable (although it would, a priori, change the list: at the end of this section we give examples of QM surfaces which can be defined over Q as abelian surface but not as principally polarized abelian surface).

b) In any of its forms, this finiteness conjecture is very far from being resolved. Notice that it is true for $d = 1$ (i.e., for elliptic curves) due to the fact that the endomorphism algebra of E/C is either Q or a CM field of class number one, of which we know there are precisely 9. When $d = 2$, we have a similar classification of the possible CM endomorphism algebras, but already the QM case presents problems: it is Conjecture 93. Aside from Q, the other possible division algebra arising as the endomorphism algebra of an abelian surface in characteristic zero is a real quadratic field, and we are "reduced" to the problem of studying rational points on Hilbert modular surfaces. The higher dimensional versions of this conjecture lead us to the consideration of Q-points on various other families of higher-dimensional Shimura varieties of PEL-type.

A result of the form $X^{D+}(Q) = \emptyset$ would have very interesting consequences both in terms of properties of the curve X^{D+}/Q itself and for the moduli problem it is (coarsely) associated to. Unfortunately, such a result does not appear in this thesis. The best we can offer at the moment is two insights into why such a theorem should be *hard* to prove: first the existence of rational CM points in a positive density situation means that it will not be the case that $X^{D+}(Q) = \emptyset$ for all sufficiently large D, so any argument must take this into account. Second, by Jacquet-Langlands-Faltings-Ribet, $J(X^{D+}) = J_0(D)^{new, w_D=1}$, which according to the conjecture of Birch and Swinnerton-Dyer will yield no nontrivial quotient of rank zero, i.e., there can be no analogue of the winding quotient in this context.

Instead, we ask an easier question: for which discriminants D is it the case that there exist infinitely many A/Q (up to geometric isomorphism) with $End_{\overline{Q}}(A)$ a maximal order in B_D? It is here that we need to make use of the work of Rotger recalled in Section 2.2: assume more precisely that we are looking for discriminants D such that there exist infinitely many geometrically distinct structures $(A, P)/Q$ of principally polarized $\mathcal{O}_D$-PQM surfaces. Then such an (A, P) induces a Q-point on some Atkin-Lehner quotient X^D/w_d or $X^D/\langle w_d, w_m \rangle$ where m gives a twist of B_D. We can now state the main result of this chapter:

Main Theorem 1 *The list of discriminants such that there exist infinitely many principally polarized A/Q (up to geometric isomorphism) with $End_{\overline{Q}}(A) = \mathcal{O}_{B_D}$ is as follows:*

a) when $g(X^{D+}) = 0$:
$D = 6, 10, 14, 15, 21, 22, 26, 33, 34, 35, 38, 39, 46, 51, 55, 62, 69, 74, 86, 87, 94, 95, 111, 119, 134, 146, 159, 194, 206$

b) when $g(X^{D+}) = 1$:
$D = 58, 58, 65, 77, 82, 106, 118, 122, 129, 143, 166, 210, 215, 314, 330, 390, 510, 546$

c) when D *is twisting and* $g(X^{D+}) \geq 2$ *but* $g(X^D/H_R) = 0$:
$D = 85, 115, 202, 570, 690, 770$

d) when D *is twisting and* $g(X^{D+}) \geq 2$ *but* $g(X^D/H_R) = 1$:
$D = 91, 123, 185, 214, 218, 235, 262, 278, 298, 326, 335, 346, 362, 458$

We remark that there could exist further discriminants D such that there are infinitely many abelian surfaces A/Q with $End^0_{\overline{Q}}(A) = B_D$ – provided $End(A)$ is not a maximal order or A/Q does not admit a principal polarization over Q. Indeed, a slight modification of the methods of the theorem yield further discriminants D such that there exist infinitely many abelian surfaces A/Q with $End(A)$ a maximal order in B_D but are not PQM's for the technical reason that they do not admit principal polarizations over Q. This is made possible by the phenomenon (first studied in [Rotger II-IV], but already visible in Jordan's thesis) that a QM abelian surface over C can have more than one principal polarization.

2.2 The proof of Main Theorem 1

By work of Victor Rotger described in Chapter 1, every principally polarized abelian surface with geometric $\mathcal{O}$-QM lies in the image of a forgetful modular map $\varphi_\mu : X^D \to \mathcal{A}_2$, and the degree of φ_μ onto its image is either 2 or 4, according to whether the principally polarized order $(\mathcal{O}, \mu$ is nontwisting or twisting: either way, $X^D/H_R \hookrightarrow \mathcal{A}_2$. Say A/Q is a ppas lying in the image of φ_μ. Let K be the splitting field of the divisorial preimage of (A, P) in X^D, so K/Q is an abelian extension, either of degree 2 or of degree 4 and type $(2, 2)$ (in the nontwisting case it must be the former). We call K the canonical field of $(A, P) \in \mathcal{A}_2$; when A is simple it may equivalently be characterized as the field cut out by the Galois action on $End^0_{\overline{Q}}(A)$.

So it is enough to determine which curves $X^D/H_R/Q$ have infinitely many Q-points. By Faltings' celebrated theorem, we need only consider curves of genus 0 or 1.

Proposition 96 *The list of curves* X^D/H_R *of genus 0 or 1 is precisely as in the statement of Main Theorem 1.*

Proof: From Chapter 0 we have both a genus formula for an arbitrary Atkin-Lehner quotient of X^D and a guarantee that there are only finitely many curves whose genus is bounded. Using the fact that $\#H_R \leq 4$, it is trivial to make Corollary 50 effective and compute the entire list.

Proposition 97 *Each of the curves in the list of the previous proposition has infinitely many Q-points.*

Proof: First, by means of Proposition 65 we find at least one rational CM point on each X^{D+} (when $X^{D+} \to X^D/H_R$ has degree 2, this gives a fortiori a rational CM point on X^D/H_R). In case X^D/H_R has genus zero, the existence of a rational point implies $X^D/H_R \cong_Q P^1$,[1] and there is no more to say. Assume now that $g(X^D/H_R) = 1$. From Chapter 1, we recall that $J(X^D) \sim_Q J_0(D)^{new}$, so $J(X^{D+}) \sim J_0(D)^{new,+}$, and $J(X^D/H_R)$ is isogenous to the appropriate Q-factor of $J_0(D)^{new,+}$. Also, the fact that we have a Q-point allows us to identify the genus one curve X^D/H_R with its Jacobian $J(X^D/H_R)$; since having infinitely many rational points is a Q-isogeny invariant of elliptic curves, it is enough to show that every Q-factor of $J_0(D)^{new,+}$ has infinitely many rational points. By the classical Atkin-Lehner theory of signs of functional equations, this implies that X^D/H_R has odd analytic rank, so the infinitude is predicted by the conjecture of Birch and Swinnerton-Dyer. Of course we need not assume BSD: we can look in Cremona's tables and verify that for our list of D, every elliptic curve with conductor D and w_D-eigenvalue 1 really does have positive rank, completing the proof of the proposition.

The rest of the proof: We have seen that whenever X^D/H_R has genus 0 or 1, $\#X^D/H_R(Q)$ is infinite. However, because of field of moduli versus field of definition issues, this does not in itself tell us that there exist infinitely many $\mathcal{O}_B$-QM surfaces A/Q. Indeed, let $(A, P) \in \mathcal{A}_2$ be a point with field of moduli Q lying in the $\mathcal{O}_B$-QM locus and more specifically in the image of φ_μ. We saw in Chapter 1 that (A, P) is a moduli point – i.e., is induced by an abelian surface A/Q – if and only if the canonical field K splits B. So it remains to be seen is that, among the infinitely many Q-points living on X^D/H_R, there exists an infinite subset of points all of whose canonical fields split B. It may sound as if we would need to look at the equations defining the map $X^D \to X^D/H_R$ in order to check this, and this would be bad news: computing the equations of over 100 Shimura curves would be many (many)hours of hard toil. The key is to exploit the fact that we have at least one moduli point on each curve – the rational CM point we found above – and show that all points which are "sufficiently close" to the CM point will also be moduli points. We again give separate attention to the cases of genus 0 and genus 1.

Genus 0 case: Consider the map $X^D \to X^D/H_R \cong P^1$, and fix $P_0 \in P^1$ a rational CM point. Fix any $p|D$. Let $K_p(P_0)$ be the canonical field for P_0 over the base Q_p, i.e., the extension of Q_p cut out by the coordinates of the preimages of P_0 in X^D/Q_p. By Krasner's Lemma, if $P \in P^1/Q_p$ is sufficiently close to P_0, $K_p(P_0) = K_p(P)$. By weak approximation for Q, there exist infinitely many points $P_n \in P^1/Q$ which are simultaneously p-adically close to P_0 for all p dividing D such that $K(P_n)$ has the same p-adic completions at all $p|D$ as $K(P_0)$. By the Hasse Principle in the Brauer group of Q, the $K(P_n)$'s are all

[1] This follows from our discussion of Severi-Brauer conics in Chapter 0.

splitting fields for B.

Genus 1 case: We can implement the same idea of simultaneous p-adic approximation in a slightly different way. Let us assume that the canonical field K/Q is biquadratic; the proof is exactly the same (but slightly easier) in the plus quotient case. Fix a rational CM point $P_0 \in X^D/H_R(Q)$; we may thus view $E/Q := (X^D/H_R, P_0)$ as an elliptic curve, which by Jacquet-Langlands-Faltings, has semistable bad reduction at all primes $p|D$. Fix such a p, and recall the exact sequence

$$0 \to E^{ns}(Q_p) \to E(Q_p) \to \Phi \to 0,$$

where Φ is the finite abelian group corresponding to the component group of the special fibre of the Néron model at p [Silverman]. The subgroup $A_p := E^{ns}(Q_p) \cap E(Q)$ of Q-points with the same reduction as P_0 is thus of finite index in $E(Q)$. Let $A = \cap_{p|D,podd} A_p$; it is clearly an infinite subgroup of $E(Q)$. I claim that every P in A is a moduli point.

Proof of the claim: By assumption, the extension of function fields $Q(X^D)/Q(E)$ is biquadratic; let $L_i = Q(E)(\sqrt{f_i})(1 \leq i \leq 3)$ be the three intermediate quadratic fields. We need an innocuous lemma (whose proof we omit) on the splitting of quaternion algebras in $(2, \ldots, 2)$-extensions:

Lemma 98 *Let $B = B_D/Q$ be a nonsplit indefinite rational quaternion algebra, and let K/Q be a $(2, \ldots, 2)$-extension (i.e., a compositum of quadratic fields). Then K splits B if and only if for every odd $p|D$, there exists a quadratic subfield $Q \leq L \leq K$ such that p is nonsplit in L.*

Now, let $P \in A$, so that P has the same mod p reduction as P_0. By the lemma, we must show that for each odd $p|D$, then for at least one of the three quadratic subfields L_i of K, p is nonsplit in L_i, i.e., the Legendre symbol $(\frac{L_i}{p}) \neq 1$. Because $K(P_0)$ splits B, there exists an i such that p is nonsplit in $L_i(P_0) = Q(\sqrt{f_i(P_0)})$, so $(\frac{L_i(P_0)}{p}) = (\frac{f_i(P_0}{p}) \neq 1$. But by definition of the subgroup A, $f_i(P) \equiv f_i(P_0)$ modulo p and since p is odd, the Legendre symbol depends only on the mod p reduction of the numerator (a "tameness" property). Thus $(\frac{L_i(P)}{p}) = (\frac{f_i(P)}{p}) = (\frac{f_i(P_0)}{p}) = (\frac{L_i(P_0)}{p}) \neq 1$. It follows that there exist infinitely many moduli points on E, completing the proof.

2.3 A result on QM surfaces without Q-rational principal polarizations

Heretofore in this chapter we have studied low-genus quotients of Shimura curves by the subgroup H_R of Atkin-Lehner involutions; as we saw, this was the appropriate subgroup to study principally polarized abelian surfaces. On the other hand, the methods of Section 2 would apply equally well to those Shimura curves

such that the full Atkin-Lehner quotient X^D/G has genus 0 or 1. We have the following variant of Proposition 90:

Proposition 99 *Let $P \in X^D/G(Q)$, and let K be the field cut out by the divisorial preimage of P in X^D. Then if K splits B, P is a moduli point, i.e., is induced by a structure $(A, G.P)/Q$, where A/Q is an abelian surface and $G.P/Q$ is the G-orbit of a principal polarization.*

The proof is the same as for Proposition 90, i.e., by Galois descent. As in Section 3.2 we may establish the result:

Proposition 100 *The curve $X^D/G/Q$ has infinitely many Q-points if and only if its genus is zero or one (and in each case there is a rational CM point). Such curves are finite in number by Corollary 50. The list of D includes those D from Main Theorem 1 and in addition the following D:*
genus zero: $93, 161, 178, 183, 237, 462, 714, 798, 858, 870, 910, 930, 966, 1110,$
$1122, 1190, 1218, 1230, 1254, 1290, 1302, 1326, 1410, 1590, 1722$
$1770, 1794, 1914, 1938, 1974, 2010, 2130.$

genus one: $141, 142, 155, 158, 201, 203, 209, 219, 226, 254, 274, 309,$
$327, 381, 446, 1155, 1330, 1430, 1482, 1518, 1554, 1610, 1785, 1806, 1830, 2046$
$2090, 2170, 2190, 2210, 2226, 2262, 2370, 2415, 2442, 2478, 2490, 2670, 2706, 2838, 2910$
$3030, 3090$

Consider now the additional discriminants listed in Proposition 8. For any one of these D, the arguments of Section 3.2 generalize to produce infinitely many Q-rational moduli points. We know that for such a D not included in the list of Main Theorem 1, there are only finitely many principally polarized A/Q with $\mathcal{O}_{B_D}$-QM. On the other hand, we know from Chapter 0 that any abelian surface A/C with $\mathcal{O}_B$-QM is principally polarizable. We conclude:

Theorem 101 *For each D listed in Proposition 100, there exist infinitely many $\mathcal{O}_{B_D}$-abelian surfaces A/Q which are geometrically principally polarizable but do not admit principal polarizations over Q.*

Such an abelian surface A/Q corresponds to a Galois orbit of genus 2 curves C/K, where K/Q is a $(2, \ldots, 2)$-extension whose Galois group is naturally a quotient of G; the curves C cannot be defined over Q, but their common Jacobian $A = J(C)$ can be. It would be interesting (although probably difficult) to compute a particular example.

Chapter 3

Local points on Shimura curves

In this chapter we study local points on the curves X^{D+} and $X_0^{D+}(N)$; recall that, as always, N is squarefree and prime to D. Keeping in mind Ogg's results on the R-points on these curves from Section 1.4, we are left to studying the Q_p-valued points for various primes p. We will show the following results:

Main Theorem 2 *For all primes p, $X^{D+}(Q_p)$ is nonempty.*

Main Theorem 3

a) Assume $D = pq$ is a product of two primes and that N is a prime number. Then $X_0^{pq+}(N)(Q_p)$ is nonempty if and only if N is a norm from $Q(\sqrt{-q})$.

b) For fixed (arbitrary) D and sufficiently large prime N, $X_0^{D+}(N)(Q_N)$ is nonempty.

Remarks: That $X^{D+}(Q_p)$ is nonempty for all p dividing N was also proved by Andrew Ogg [Ogg II] and by Srinath Baba [Baba] (see also [Jordan-Livné III]). The proof given here is a little different in that it exploits a modular interpretation (due to Ribet) of the edges as well as the vertices of the finite graph dual to the special fibre. It is interesting to note that Ogg, Baba, and the author all prove more general results reducing to our Main Theorem 2: Baba determines when any Atkin-Lehner quotient X^D/w_d has Q_p rational points for p dividing D (and applies his result to the oddness of the Jacobian in the sense of [Poonen-Stoll]), whereas Ogg shows that $X_0^D(N)/w_{DN}(Q_p)$ is nonempty for all p dividing N.

In the present form, the proofs in Section 4 have one foot in the theory of enhanced CM and supersingular elliptic curves and one foot in the theory of Brandt-module categories. A more systematic use of the latter would lead to stronger results: Main Theorem 3 should remain valid for arbitrary D and squarefree N prime to D, as well as Ogg's generalization of Main Theorem

2. To be honest, I feel that arguments involving canonical lifting of supersingular elliptic curves are more appealing than arguments involving traces of Eichler-Brandt matrices. I am hopeful that the general situation can be made "geometric" by using the definite analogue of Shimura curves due to Gross and Roberts (see e.g. [Bertolini-Darmon]), and with any luck the final form of the results of this chapter will be couched in this language.

The organization of this chapter is as follows: in Section 1 we recall a technical (but extremely useful) result on the number of fixed points of an Atkin-Lehner involution on a Brandt-module category which can be found (albeit in somewhat disguised form) in [Vignéras]. In Section 2 we show that $X^{D+}(Q_p)$ is nonempty for all p prime to D. In Section 3 we show that $X_0^{D+}(N)(Q_N)$ is nonempty for fixed D and sufficiently large N. And in Section 4 we discuss the locus $X_0^{D+}(N)(Q_p)$ at primes p dividing N; the proof uses the Cerednik-Drinfeld reviewed in Chapter 0.

Finally, we should point out that we do not offer a result on the non/emptiness of the locus $X_0^{D+}(N)(Q_p)$ for primes p not dividing DN – these are precisely the primes of good reduction! Such a result could be put to good use in the context of the Hasse principle violations of the next chapter, so any ideas in this direction would be especially warmly received.

3.1 The fixed point formula

Let $\overline{B}/Q$ be a *definite* rational quaternion algebra of discriminant Dp. Choose, as usual, a squarefree positive integer N prime to Dp, and fix $\mathcal{O} \leq \overline{B}$ a level N Eichler order. Consider the Brandt set $Pic_r(\mathcal{O})$ of (right) classes of (left) $\mathcal{O}$-ideals; this is a finite set, and the free abelian group $M := Z[Pic_r(\mathcal{O})]$ is called the *Brandt module*. In a highly appropriate way it is a module over the Hecke algebra T,[1] but for our purposes here we are concerned only with the action of the Atkin-Lehner group on M. For this, observe that for any ring R, the automorphism group of R acts on $Pic_r(R)$ by "transport of structure." In the present case, this comes down to saying that a representative γ_m of an element of the Atkin-Lehner group acts on the Brandt set by conjugating the $\mathcal{O}$-module structure map: $\iota \mapsto \gamma_m^{-1} \circ \iota \circ \gamma_m$.

Clearly the automorphism induced by w_m on $Pic_r(\mathcal{O})$ is involutory; at several points in this chapter we will find ourselves in need of a formula for the number of fixed points (and especially, the criterion for when there are fixed points at all). We have the following result from [Vignéras, p. 152]:

Proposition 102 *(The fixed-point formula)*
a) The "main" Atkin-Lehner involution w_{pDN} always has fixed points.

[1] It is precisely Brandt module computations which are at the heart of MAGMA's modular forms package, so we have them – as well as David Kohel and William Stein – to thank for the ease and depth of modular forms calculations available to us in the present day.

b) An arbitrary Atkin-Lehner involution w_m has fixed points if and only if every prime dividing the discriminant pD is nonsplit in $Q(\sqrt{-m})$ and every prime dividing the level N is noninert in $Q(\sqrt{-m})$.

c) When $N = 1$ the number of fixed points of the main Atkin-Lehner involution w_{pD} is

$$\frac{h'(-D) + h'(-4D)}{2},$$

where $h'(m)$ is to be interpreted as the class number of the quadratic order of discriminant m if such exists (i.e., if m is 0 or 1 mod 4) and 0 otherwise.

Remark: Recall that when $D = 1$ the Brandt set is isomorphic (as Hecke module) to the category of supersingular elliptic curves in characteristic p in such a way that Frobenius corresponds to w_p, so when $N = 1$ we recover the classical formula for the number of supersingular elliptic curves defined over F_p and for general N we get a formula for the number of "enhanced supersingular elliptic curves" defined over F_p.

Remark: Comparing this result with Proposition 48 suggests that there should be a unified geometric proof.

3.2 Local points on X^{D+} at good primes

Recall from Chapter 0 that indeed X^D (*a fortiori* X^{D+}) is smooth at all primes p not dividing D. Thus, by Hensel's Lemma, it will be enough to show the existence of an F_p-valued point. We claim that in fact there will be an F_p-valued supersingular point. But this claim follows almost immediately from our identification of the supersingular isogeny class in Section 0.7 and the fixed point formula of the previous section: we have $X^D(F_{p^2})^{ss} = X^D(\overline{F_p})^{ss}$ in bijection with the Brandt set $Pic_r(\mathcal{O})$; from this we deduce $X^{D+}(F_{p^2})^{ss}$ is in bijection with the w_D-orbits of $Pic_r(\mathcal{O})$, and then that $X^{D+}(F_p)^{ss}$ corresponds to those w_D-orbits stable under w_p, i.e., to ideals I such that $w_p\{I, w_D I\} = \{I, w_D I\}$. But by the fixed point formula, there exists at least one ideal I such that $I = w_{pD} I = w_p w_D I$, so that $w_p\{I, w_D I\} = \{w_D I, I\}$, completing the proof.

3.3 Local points on $X_0^{D+}(N)$ at Deligne-Rapoport primes

We exploit the description of $X_0^D(N)/F_N$ as being two copies of the smooth curve X^D/F_N intersecting transversely along the supersingular points. By Hensel's Lemma, it is enough to show that there is an *ordinary* point $P \in X^D(F_N) - X^D(F_N)^{ss}$. But by the fixed point formula of Section 1, we have good control over the number of supersingular points: for fixed D, $\#X^D(F_N)^{ss} =$

$O(\sqrt{N})$. On the other hand, letting g be the genus of X^D, by Weil we have $|\#X^D(F_N) - (N+1)| \leq 2g\sqrt{N}$. So, obviously, for fixed D and sufficiently large N there exist ordinary F_N-valued points on X^D. These lift to give Q_N-points, and we're done. Note well that we've worked with X^D instead of X^{D+}, obtaining a stronger result that is an ingredient in our Main Theorem 5.

3.4 Local points at Cerednik-Drinfeld primes

3.4.1 Preliminaries

Notation: When our choice of level structure is clear from the context, we will allow ourselves to write X for either of: $X^{pq}(1), X_0^{pq}(N)$,$\mathrm{X}^{(p)}$ for either of: $X^{pq}(1)/w_p, X_0^{pq}(N)/w_p$, and $X^{(pq)}$ for either of:
$X^{pq}(1)/w_{pq}, X_0^{pq}(N)/w_{pq}$. Finally, we write Z_{p^∞} for $W(\overline{F_p})$ (i.e., for the integer ring of the completion of the maximal unramified extension of Q_p).

First we need to recall some of [Jordan-Livné I]: work of Cerednik-Drinfeld gives us canonical Z_p-models for all of our curves, such that the special fibres are admissible curves in the sense of Jordan-Livné: they are (reduced) semi-stable curves $/F_p$, every irreducible component of which has geometric genus 0. To such a curve we can associate a finite graph, its dual graph, as follows: the vertex set of the graph corresponds to the irreducible components of the curve, and edges $e : v_1 \to v_2$ correspond to intersection points of the components corresponding to v_1, v_2. Recall also that these curves are not Mumford curves; rather thery are twists of Mumford curves under $\mathrm{Frob}_p \mapsto \mathrm{w}_p$. That is, a component (respectively, a singular point) is defined over F_p if and only if it is fixed under the action of w_p. We will use:

Proposition 103 *(Hensel's Lemma): Let X/K be a projective curve over a field K which is complete with respect to a non-Archimedean valuation. Let $\mathcal{X}/O_K$ be any regular model for X. Then $X(K)$ is non-empty if and only if $\mathcal{X}(O_K/mO_K)$ has a smooth point.*

The canonical models $\mathcal{X}/Z_p$ need not be regular. Indeed, let $P \in \mathcal{X}/Z_{p^\infty}$ be a point lying on the closed fiber whose reduction into $\mathcal{X}(\overline{F_p})$ is singular. Then there is an analytic neighborhood of P in $\mathcal{X}/Z_p^\infty$ isomorphic to $Z_{p^\infty}[[X,Y]]/(XY - p^a)$. It is not hard to check that this local ring is regular if and only if $a = 1$. In general, the integer a is associated to the edge e and called its *length*. Thus, we have in all the structure of a "finite ℓ-graph," i.e., a finite graph to which each unoriented edge is associated a positive integer. To perform the regularization of the arithmetic surface $\mathcal{X}$, we repeatedly blowup at the points on the special fibre corresponding to edges e with length > 1. In terms of ℓ-graphs, this corresponds to replacing the single edge of length m with m edges of length 1. Thus, from the data of an ℓ-graph we can construct the special fibre of a regular model for X/K (and indeed, we could go on to construct the minimal model, if we needed it; see [Jordan-Livné I] for details.)

Now we consider the problem of examining the ℓ-graph G to determine whether $X(Q_p)$ is empty. A smooth F_p-point on the special fibre of the regularized surface comes from one of the following:
a) a w_p-fixed vertex v of G, or
b) an edge e of even length, which is flipped by w_p – i.e., such that

$$w_p : v \to w \mapsto w \to v.$$

(The length must be even so that the w_p-Frobenius action fixes the middle vertex in the chain created by blowing up the single edge e.) In fact, our first step is that we need not worry about b) in our situation. Let us agree for now that $G, G^{(p)}, G^{(pq)}$ stand for the dual graphs of the special fibres of the curves $X^{pq}(1), X^{pq}(1)/w_p, X^{pq}(1)/w_{pq}$ respectively.

Proposition 104 *There is no two-sided even length edge e of $G^{(pq)}$ which is flipped by w_p.*

For the proof, we need:

Lemma 105 *The sets $l(G^{(p)}), l(G), l(G^{(pq)})$ of lengths of the dual graphs are all the same. Moreover, when counted with multiplicity, the multisets of lengths of oriented eges of $G^{(p)}$ and $G^{(pq)}$ coincide.*

Proof of the lemma: The graph G admits a natural bipartition coming from the bipartition of the Bruhat-Tits tree Δ given by considering vertices of even and odd distance from any given vertex. From e.g. [Kurihara], one knows that w_p interchanges the two subsets of the bipartition (and in particular acts freely on the vertices and oriented edges of G) and w_q preserves each subset of the bipartition, so clearly w_{pq} again interchanges the two subsets and acts freely on vertices and oriented edges. From this we see that the at-most-(2-1) maps $G \to G^{(p)}, G \to G^{(pq)}$ are precisely 2-1 on the sets of oriented edges of G. Since an equivalent characterization of the length of an edge G is the cardinality of the stablizer of that edge inside the discrete subgroup of $GL_2(Q_p)$ associated to the curve [Jordan-Livné I], this lack of ramification immediately implies the lemma.

Proof of the proposition: From [Kurihara], we see that the only edges of even length in $G^{(p)}$ are those of length 2, and moreover the number of such oriented edges is either 0,1 or 2. In case there is only one such oriented edge, it must be one-sided. Its lift to G is therefore an edge $e : v \to w$ such that $w_p(e) = \overline{e} : w \to v$. Since all the Atkin-Lehner involutions preserve the length-structure of the graphs, we must have $w_q\{v, w\} = \{v, w\}$; since moreover v and w are adjacent vertices and w_q preserves the bipartition, we must have $w_q(v) = v$ and $w_q(w) = w$. Thus $w_{pq}(e) = \overline{e}$ as well, so that in G^{pq} the unique edge of length 2 is again one-sided. But according to the recipe for computing the special fibre of an admissible curve from its dual graph, one-sided edges are removed [Kurihara], [Jordan-Livné I], showing the proposition in this case. In case there are two oriented edges of length 2 in $G^{(p)}$ and they are both one-sided,

the above argument again shows that the corresponding pair of length 2 edges in $G^{(pq)}$ are one-sided. In the remaining case, we can consult [Kurihara] to see that there will be a unique vertex admitting edges of length 2, so the two geometric edges of length 2 are the two orientations of a loop at that distinguished vertex. Choosing an orientation, this edge lifts to two geometric edges in G, $e_1 : v \to w$ and $e_2 : w \to v$ such that $w_p(e_1) = e_2$. As above, we find that $w_q(v) = v$ and $w_q(w) = w$, so that the unique length 2 edge of $G^{(pq)}$ is again a loop at a single vertex; this edge can be represented in the quotient graph as a pair $\{e_1, e_2\}$, and evidently $w_p(\{e_1, e_2\}) = \{e_2, e_1\} = \{e_1, e_2\}$, so the edge is fixed by w_p and not flipped by it.

In fact it is not the proposition itself that we need in this section (it points in the direction of the emptiness of the Q_p-points, not the non-emptiness), but rather its analogue in the case of level structure:

Corollary 106 *There are no one-sided edges of even length in the dual graphs $G(X_0^{pq}(N)/w_{pq})$ which are flipped by w_p.*

In order to prove the corollary we must recall work of [Cerednik-Drinfeld], made explicit in our context by [Ribet] that interprets our graphs in terms of structures on supersingular elliptic curves. It is convenient to work adelically: let $U \leq B^\times(A_f)$ be a compact open subgroup which is maximal at p and q. Following [Ribet], we introduce the p-adic space $X_U := \Pi_{l \neq p}\ \mathrm{U}_l \backslash \widehat{\overline{B}}^\times / \overline{B}^\times$, where $\overline{B}/Q$ is the quaternion algebra obtained by "interchanging the invariants at p and ∞," i.e., $\overline{B}$ is definite of discriminant q. Notice that the spaces $\widehat{\overline{B}}^\times$ and $\widehat{B}^\times$ differ only at their p-components; thus, the prime-to-p part of U is an acceptable prime-to-p-level structure for the zero-dimensional Shimura variety $\overline{B}^\times$. Now we have two key observations:

Proposition 107 *(Ribet, (4.3))*
a) $GL_2(Z_p)\backslash X_{\Gamma_0(N)}$ is canonically isomorphic to the vertex set of $G(X_0^{pq}(N)/w_p)$. Moreover the vertex set of $G(X_0^{pq}(N))$ is canonically given as the disjoint union of two copies of $GL_2(Z_p)\backslash X_{\Gamma_0(N)}$.
b) The edge set of $G(X_\bullet^{pq}(N))$ is canonically isomorphic to $\Gamma_0(p)\backslash X_{\Gamma_0(N)}$. Moreover, the attaching map for the graph is as folows: to an element $e \in \Gamma_0(p)\backslash X_{\Gamma_0(N)}$, we associate its initial vertex $e(0)$ via the natural projection

$$\pi_1 : \Gamma_0(p)\backslash X_{\Gamma_0(N)} \longrightarrow (GL_2(Z_p)\backslash X_{\Gamma_0(N)})^1,$$

and its terminal vertex via $\pi_2 \circ (m^{-1}x)$, where

$$\pi_2 : \Gamma_0(p)\backslash X_{\Gamma_0(N)} \longrightarrow (GL_2(Z_p)\backslash X_{\Gamma_0(N)})^2$$

is the same map as before but formally landing in the second copy, and $m \in \widehat{\overline{B}}^\times$ is an idèle which is everywhere locally trivial except at p, and such that

$$m_p = \begin{bmatrix} 1 & 0 \\ 0 & p \end{bmatrix}.$$

Proposition 108

a) $GL_2(Z_p)\backslash X_{\Gamma_0(N)}$ *is in bijection with the set of supersingular elliptic curves with* $\Gamma_0(N)$*-structure, i.e., with* $X_0(N)(\overline{F_q})^{ss}$*.*
b) $\Gamma_0(p)\backslash X_{\Gamma_0(N)}$ *is in bijection with the set of supersingular elliptic curves with* $\Gamma_0(p) + \Gamma_0(N)$*-structure.*
c) Under these correspondences, w_q *on the dual graph of* $X_0^{pq}(N)/w_p$ *corresponds to the* q*-power Frobenius morphism* $Frob_q$ *on* $X_\bullet(N)(\overline{F_q})^{ss}$*. In particular, the set of* w_q*-fixed vertices corresponds to the locus* $X_0(N)(F_q)^{ss}$*.*

Proof: a),b) are standard adelic double coset constructions of the sort we recalled in Section 0.5 For c), see [Ribet].

Proof of the corollary: Let $\tilde{e} \in G(X_0^{pq}(N)/w_{pq})$ be a two-sided edge of even length that is flipped by w_p, and let $e \in G(X_0^{pq}/w_{pq})$ be the image of $\tilde{e}$ under the quotient map. I claim that the quotient map $\pi : G(X_0^{pq}(N)/w_{pq}) \to G(X^{pq}/w_{pq})$ preserves the orientation of edges; equivalently, $G(X_0^{pq}(N)) \to G(X^{pq})$ is orientation-preserving. Indeed, the initial vertex of the edge $\tilde{\tilde{e}} = (E, C_N)$ is evidently just (E, C_N), so that the initial vertex of the image of $\tilde{\tilde{e}}$ in $G(X^{pq})$ is unambiguously E (and indeed there are no loops in these bipartite graphs, so this analysis suffices). Now, it is clear that since w_p flips $\tilde{e}$, it must either flip or fix the image vertex e. Suppose first that it does both, i.e., that e is one-sided. Then, because π is orientation-preserving, $\tilde{e}$ must be one-sided as well, contrary to our assumption. Otherwise, the orientation-preserving nature of π implies that since $\tilde{e}$ gets flipped upstairs, e is a two-sided edge of even length that gets flipped downstairs, contradicting Proposition 104.

3.4.2 The proof of Main Theorem 2

In this subsection, we work with no level structure, and we write $G, G^{(p)}, G^{(pq)}$ for the three dual graphs in question. Because of Proposition 98, we must show that $X^{(p)}(F_p)$ is nonempty. First observe that a w_p-fixed vertex of $G^{(pq)}$ can be viewed as a pair $\{v, w_{pq}v\}$ such that $w_p\{v, w_{pq}v\} = \{v, w_{pq}v\}$, i.e., $\{w_pv, w_qv\} = \{v, w_{pq}v\}$. So either $v = w_pv$ or $v = w_qv$. The former is impossible, since w_p has no fixed points on the vertex set of G; hence $v = w_qv$. That is, fixed points of w_p on $G^{(pq)}$ correspond to pairs of fixed points of w_q on G. Similarly, under the quotient map $G \to G^{(p)}$, each pair of w_q-fixed points gets mapped to a single w_q-fixed point of $G^{(p)}$, which by Proposition 108c), corresponds to an element of $X(1)(F_q)^{ss}$, i.e., to a supersingular elliptic curve defined over the prime subfield F_q. It is well-known that this set is non-empty; e.g., by Honda-Tate theory the Weil q-number $\sqrt{-q}$ provides us with at least one such element. (This is a special case of the fixed point formula of Section 1.) We see that there are certainly components of $X^{(pq)}(F_p)$ which are defined over F_p. The only way that such a component could not yield a smooth F_p-point is if the component had the maximum number of singular points, namely $p + 1$, and if all of these singularities were themselves defined over F_p. Since singular

points on the special fibre correspond to edges in the dual graph, it will be enough to show the following

Claim: For any w_q-fixed vertex v of $G^{(p)}$ such that $p+1$ edges emanate from v, the w_q-action on this set of edges is nontrivial.
Sufficiency of the claim: if e is an edge of a w_q-fixed vertex, then $w_q(e(1)) = e(1)$ if and only if the terminal vertex of the associated vertex in $G^{(pq)}$ is w_q-fixed.

Proof of the claim: Because of Propositions 107 and 108, we can rephrase in terms of supersingular elliptic curves: namely, it is enough to show that if E/F_q is a supersingular elliptic curve, it is *not* the case that the Galois (=Frobenius)-action on the set of order p-subgroups of E is trivial. First assume that $q > 3$, so the characteristic polynomial of Frobenius must be $X^2 + q$, and the trace of Frobenius acting on $E[p]$ is zero, hence the matrix is nonscalar and accordingly moves some one-dimensional F_p-subspace, establishing the claim in this case. Now assume that q is 2 or 3, so the characteristic polynomial of Frobenius, if not $X^2 + q$, is $x^2 \pm qX + q$, whose rational canonical form as an endomorphism of $T_p(E)$ is $\begin{bmatrix} 0 & -q \\ 1 & \pm q \end{bmatrix}$, so that the action on $E[p]$ is again non-scalar. This completes the proof of the claim, and hence of Main Theorem 2.

3.4.3 The proof of Main Theorem 3

Having done all the necessary analysis in the preceding section, we can immediately reduce to the realm of $\Gamma_0(N)$-structures on supersingular elliptic curves over F_q. Indeed, using the discussion at the beginning of Section 3.3 and using Corollary 105, a sufficient condition for $X_0^{pq}(N)/w_{pq}(Q_p)$ to be empty is for the action of w_p on the dual graph $G(X_0^{pq}(N)/w_{pq})$ to have no fixed points. By the same argument used at the end of the proof of Theorem 1, if v is a w_q-fixed vertex of $G^{(p)}$, then either the cardinality of the star of v is less than $p+1$ or the action of w_q on the set of edges emanating from v will be nontrivial, implying that at least one of the $p+1$ F_p-rational points on the rational curve corresponding to v is smooth, whence we have established:

Proposition 109 *$X_0^{pq}(N)/w_{pq}(Q_p)$ is empty if and only if the action of w_q on $G^{(p)}$ is fixed point-free.*

But now we are done, since the proposition implies that $X_0^{pq+}(Q_p)$ is nonempty if and only if the Atkin-Lehner involution w_q has a fixed point on the Brandt set associated to an Eichler order of level N in the definite quaternion algebra of discriminant q. That is, we have reduced to the fixed point formula of Section 1, which completes the proof of the theorem.

Nevertheless, it is more "educational" to work with supersingular elliptic curves, and we indicate how most (no doubt all, with a bit more technique) of the theorem can be proved in this way. First, we can equally well state the proposition in terms of supersingular elliptic curves, getting:

Proposition 110 *$X_0^{pq}(N)/w_{pq}(Q_p)$ is empty if and only if there is no pair $(E,C)/F_q$, E/F_q a supersingular elliptic curve, $C \leq E$ a cyclic order N (hence étale) subgroup scheme of E defined over F_q.*

Proof of the theorem: Assume first that $q \equiv 1$ modulo 4, so in particular $q > 3$. Suppose there exists a pair $(E,C)/F_q$ as in the proposition. Since $q > 3$, the characteristic polynomial of Frobenius is necessarily $X^2 + q$, so $End_{F_q}(E) = Z[\sqrt{-q}]$, and this is the maximal order in the CM quadratic field $Q[\sqrt{-q}]$. By Deuring's lifting lemma, there exists $\tilde{E}/\overline{Q_q}$ an elliptic curve with CM by $Z[\sqrt{-q}]$ and $(\tilde{E}, \sqrt{-q})$ reduces to $(E, Frob_q)$ modulo q. Moreover, using the isomorphism of prime-to-q- adic Tate modules $T^{q'}(\tilde{E}) \to T^{q'}(E)$, we can lift C to a subgroup $\tilde{C} \leq \tilde{E}$. Since this isomorphism respects the CM-structures, we get moreover that $\tilde{C}$ is stable under $End(\tilde{E}) = Z[\sqrt{-q}]$. This implies that $Z[\sqrt{-q}]$ acts on the quotient $\tilde{E}/\tilde{C}$; since this is the maximal order, it must be that $\tilde{E}/\tilde{C}$ has precisely $Z[\sqrt{-q}]$-CM. By the theory of complex multiplication, there exist integral ideals a, b of $Z[\sqrt{-q}]$ so that over C the isogeny $\tilde{E} \to \tilde{E}/\tilde{C}$ may be realized as $C/a \to C/ab^{-1}$. The degree of this isogeny is on the one hand N and on the other hand the norm of b, which is what we wanted to show.

Now assume $q \equiv -1$ modulo 4. Again the ring generated by the Frobenius map inside $End^0(E)$ is $Z[\sqrt{-q}]$, and by Deuring we can lift to get a pair $(\tilde{E}, \tilde{C})$, where $\tilde{E}$ has precisely $Z[\sqrt{-q}]$-CM and $\tilde{C}$ is stable under this ring. So again $\tilde{E}/\tilde{C}$ has at least $Z[\sqrt{-q}]$-CM, but since this is no longer the maximal order, a priori it could have more CM. We thus distinguish two cases: in the first case, the quotient $\tilde{E}/\tilde{C}$ has precisely $Z[\sqrt{-q}]$-CM, so running through the above argument we get that N is the norm of an ideal in the ring $Z[\sqrt{-q}]$. Pushing this ideal forward to the full ring of integers, we get that either N or $2N$ is a norm from $Q(\sqrt{-q})$. In the second case, $\tilde{E}/\tilde{C}$ has CM by the full ring of integers. We can write $\tilde{E} = C/a$ and define a new elliptic curve $F = C/aZ[\frac{1+\sqrt{-q}}{2}]$, the "improvement" of $\tilde{E}$ to an elliptic curve with maximal CM. Obviously the quotient map $\phi : E \to F$ is a degree 2 isogeny; the composite $F \stackrel{\hat{\phi}}{\to} \tilde{E} \to \tilde{E}/\tilde{C}$ is then an isogeny between elliptic curves with CM by the maximal order, showing that $2N$ is a norm from $Q(\sqrt{-q})$.

We now prove the converse in case $q \equiv 1$ modulo 4. Namely, we must show that if N is a norm from $Q(\sqrt{-q})$, there exists a pair $(E,C)/F_q$, where E is a supersingular elliptic curve and $C \leq_{F_q} E[N]$ is an order N cyclic subgroup. Indeed the argument is very similar to the above: starting with a supersingular elliptic curve over F_q, we may lift $(E, Frob)$ to an elliptic curve E/Q_q and an endomorphism generating $Z[\sqrt{-q}]$. By our hypothesis on q, this is the maximal order, so this is the full endomorphism ring of E. Moreover, our assumption that N is a norm from $Q(\sqrt{-q})$ implies that there exists an endomorphism η of E of degree N. Let C be the kernel of η. Observe that η (hence C) can be defined over $Q_q[\sqrt{-q}]$ – a totally ramified extension of Q_q, so that the reduction of C can be defined over F_q, qed.

Chapter 4

Global points on Shimura curves

In this chapter we consider the locus $X_0^D(N)(L)$, where L is a number field. Observe that, by the results of the last chapter, depending on D, N and L $X_0^{D+}(N)$ need not have points rational even over a completion of L. But, if we believe in Conjecture 93 from Chapter 3 bounding the non-CM rational points on X^{D+}, we must believe that there will be correspondingly few points on Shimura curves $X_0^{D+}(N)$ over number fields. Indeed we make the following

Conjecture 111 *(Boundedness Conjecture) For any number field L there are numbers $N(L), D(L)$ such that $N \geq N(L),\ D \geq D(L)$ implies that $X_0^D(N)(L)$ is empty.*

The conjecture is best known when $D = 1$; for $D = 1,\ L = Q$ and N prime, we get the celebrated theorem of Mazur [RI] on the boundedness of prime degrees for rational isogenies of elliptic curves over Q. In that same paper the result was shown for ($D = 1$ and) imaginary quadratic fields K in which N is inert; to my knowledge the case in which N splits in K remains open.

In this chapter we explore the case of *fixed* $D > 1$. We adapt some of the methods of [RI] to our QM situation. Notice that since $X^D(R) = \emptyset$ the conjecture holds trivially for all number fields with a real place. We are really interested in $X_0^{D+}(N)(Q)$ but this turns out to be awkward to study directly. Instead we work in the following context: let F be a number field with a real place, and assume there exists $\overline{P} \in X^{D+}(F)$. (Recall from Chapter 2 that when the genus of X^{D+} is zero or one, $X^{D+}(Q)$ is infinite; this is the most interesting case for us.) Then the compositum of F and the splitting field of the divisorial preimage $\{P, w_D(P)\}$ of $\overline{P}$ in X^D generates a totally imaginary quadratic extension field K/F, the *canonical field* of $\overline{P}$. We will investigate the locus of points $X_0^D(N)(K)$ such whose image in $X_0^{D+}(N)$ becomes F-rational.

Definition: Say a number N is F-amenable if

• N is prime and prime to D.
•$N \equiv 1$ modulo 4.
•N remains prime in the number field F.

When $F = Q$ an amenable N is just a prime number which is 1 mod 4.

Main Theorem 4 *For fixed D and fixed K/F a totally imaginary extension of a number field with a real place, there is an absolute bound on F-amenable N such that there exists $\overline{P} \in X_0^{D+}(N)$ with canonical field K.*

Remark: As mentioned above, the theorem is most interesting when X^{D+} has genus zero or one. For when the genus is at least two, it follows by Faltings' theorem that $X^D(L)$ is a finite set, and to prove the boundedness conjecture it is enough to show that for any particular QM surface A/L there is an absolute bound on N such that A admits an L-rational QM N-isogeny. But this is known to be true due to the "largeness" of the adelic Galois representation on a QM surface, a Serre-type theorem due to [Ohta] (and independently proved by the author); for the statement, see Theorem 117 in Section 4.5.

Perhaps it will be helpful to give a few words about the strategy of the proof before plunging in. It is unabashedly based on [RI] – we assume the reader has a good familiarity with (and ready access to) this paper. The proof is much easier on the geometric side, as a key point of Mazur's argument is to show that having a rational N-isogeny for even moderately large N forces the elliptic curve to have potentially good reduction; the argument uses the Eisenstein ideal and the cuspidal geometry of modular curves. But Shimura curves have precisely no cuspidal geometry, and accordingly (as we have already seen) their potential good reduction is automatic. Thus we can skip to the analysis of the isogeny character, and we warn the reader that at this point our proof becomes more involved, due to the more slippery nature of the moduli problems at hand.

4.1 Preparation for Main Theorem 4: the Shimura Covering of $X_1^{D+}(N) \to X_0^{D+}(N)$

The object of this section is to prove the following result, an analogue of [Mazur, Corollary 2.3]:

Theorem 112 *The morphism of arithmetic surfaces $X_1^D(N) \to X_0^D(N)/SpecZ[\frac{1}{N}]$ admits a factorization $X_1^D(N) \stackrel{t}{\to} X_2^D(N) \stackrel{n}{\to} X_0^D(N)/SpecZ[\frac{1}{N}]$, where the second map is finite étale and cyclic. The index t divides 6; precisely $t = m_2 m_3$, where $m_2 = 2$ if $Q(\sqrt{-1})$ splits B_D and $(\frac{-1}{N}) = 1$; otherwise $m_2 = 1$; $m_3 = 3$ if $Q(\sqrt{-3})$ splits B_D and $(\frac{-3}{N}) = 1$; otherwise $m_3 = 3$. Finally, all of the above statements remain true for $X_1^{D+}(N) \to X_0^{D+}(N)$.*

Before giving the proof, we recall the general theory of Shimura coverings as presented by [Ling-Oesterlé]. So: let $f : Y \to X$ be a degree n morphism of algebraic curves over C (though any field of characteristic zero would work just as well, and give a Galois-equivariant theory). We define $\Sigma(f)$ as the kernel of $f^* : Pic^0 X \to Pic^0 Y$. It is a finite subgroup of $J(X) = Pic^0 X$ – indeed, since $f_* \circ f^* = [n]$, visibly $\Sigma(f)$ is contained in $J(X)[n]$; $\Sigma(f)$ is called the Shimura subgroup. We can also define a finite abelian group associated to f as follows: let $g : Z \to X$ be the maximal abelian unramified covering through which f factors; let A be its Galois group. The finite abelian groups $\Sigma(f)$ and A are canonically in duality. Indeed, the theory of line bundles on abelian varieties (Appell-Humbert theorem) together with the identification of $H^1(X, Z)$ as the lattice of covering transformations of the universal cover of $J(x)$ furnishes us with a canonical isomorphism

$$J(X) \xrightarrow{\sim} Hom(H_1(X, Z), S^1)$$

On the other hand, a monodromy argument gives us that A is the maximal abelian quotient of $\pi_1(X)$ to which $\pi_1(Y)$ maps to zero, i.e., A is isomorphic to the cokernel of $f_* : H_1(Y, Z) \to H_1(X, Z)$. We get a commutative diagram

$$[Hom(\Gamma, S^1)]J(Y)es, lf^*Hom(H_1(Y, Z), S^1)s, rf_*^\vee J(X)eHom(H_1(X, Z), S^1)$$

which exhibits the duality.

The case of modular curves: now let $\Gamma \leq GL_2^+(R)$ be a Fuchsian group of the first kind, i.e., such that $X_\Gamma = \Gamma\backslash\overline{\mathcal{H}}$ has the structure of a complex algebraic curve. Choosing any basepoint $\tau \in \overline{\mathcal{H}}$, we get a canonical surjection $\Phi : \Gamma \to \pi_1(X_\Gamma, \overline{\tau})$ as follows: for $\gamma \in \Gamma$, let c be a path in $\overline{\mathcal{H}}$ carrying τ to $\gamma\tau$; let $\Phi(\tau)$ be the homotopy class of this loop in X_Γ. One knows that the kernel of Φ is generated by the elliptic and the parabolic points of Γ. Passing to homology eliminates the dependence on the basepoint, and we get $\Gamma \to H_1(X_\Gamma, Z)$. Dualizing and composing with the above isomorphism, we get

$$\Psi : J(x) \hookrightarrow Hom(\Gamma, S^1)$$

the image of Ψ consists of homomorphisms whose kernel contains all the elliptic and parabolic elements of Γ. Now let $\Gamma' \leq \Gamma$ be a finite index normal subgroup. We have an induced map $w : X_{\Gamma'} \to X_\Gamma$ and $w^* : J(X_\Gamma) \to J(X_{\Gamma'})$. It is easy to see that the following diagram commutes:

$$[Hom(\Gamma, S^1)]J(X_\Gamma)es, lw^*Hom(\Gamma, S^1)s, ri^\vee J(X_{\Gamma'})eHom(\Gamma', S^1)$$

Let $\Sigma = \Sigma(w)$ be the associated Shimura subgroup. Since Γ/Γ' is the Galois group of the function field extension, we wish to identify Σ as being dual to a certain quotient of Γ/Γ'. Using the last diagram and the above image condition, we get:

Proposition 113 *(Shimura coverings of modular curves): With $\Gamma' \leq \Gamma$ as above, we have $\Sigma = Hom(\Gamma/N, S^1)$, where N is the normal subgroup generated by Γ' and by all the elliptic and parabolic elements of Γ.*

The case of Shimura curves: Take $\Gamma = \Gamma_0^D(N), \Gamma' = \Gamma_1^D(N)$; then $\Gamma/\Gamma' \cong (Z/NZ)^\times$. Notice that there are no parabolic elements. Moreover, by the basic theory of Shimura curves, we find that the elliptic elements can have orders only 1,2,3 in $\Gamma_0^D(N)/+/-1$ (notice that -1 is an elliptic element of $\Gamma_0^D(N)$ according to our setup). A straightforward analysis of when these elliptic points arise now gives the "generic fibre" part of the first part of our theorem.

To complete the proof of the theorem in the $X_1^D(N) \to X_0^D(N)$ case, we must look in positive characteristic l not equal to N. First note that $(Z/N/Z)^\times/+/-1$ acts as automorphisms of $X_1^D(N)$ over $SpecZ$. It will be enough to show that the covering $X_2^D(N) \to X_0^D(N)/\overline{F_l}$ remains unramified. When l does not divide D, all of curves remain smooth in characteristic l. The Riemann-Hurwitz formula implies that any degree d morphism of smooth curves $Y \to X$ is unramified if and only if $1 - g(Y) = d(1 - g(X))$. Since we are unramified in characteristic zero and none of these invariants change in good residue characteristic, we are equally well unramified in characteristic l. (A direct analysis of the ramification of $X_1^D(N) \to X_0^D(N)/\overline{F_l}$ in terms of the points of $X_0^D(N)$ with automorphism group larger than $\{+/-1\}$ would also succeed.) In characteristic l dividing D, the curves are split degenerate: every irreducible component has normalization P^1; such a curve is specified by its dual graph. From [Kurihara], if C/Z_p^∞ is an admissible curve of generic genus g, then

$$1 - g = \#V(\widetilde{C/\overline{F_p}}) - \#E(\widetilde{C/\overline{F_p}})$$

where the tilde indicates an unpleasant feature of the theory: when the action on the Bruhat-Tits tree identifies an oriented edge with its inverse, the quotient graph has a one-sided edge; these edges do not contribute to the Euler characteristic. One way to ensure that this phenomenon does not occur is to choose a uniformizing discrete subgroup which is sufficiently small so as to preserve the natural bipartition of the vertex set of the Bruhat-Tits tree into vertices of mutually even/odd distance. From e.g. [Kurihara], we know that the graph of $X^D(1)$ enjoys this property, hence has no one-sided edges; a fortiori neither does $X_0^D(N), X_1^D(N)$. Notice then that $\#V - \#E$ is the Euler characteristic of the dual graph. Now, letting $d = \frac{N-1}{2}$ be the degree of $X_1^D(N) \to X_0^D(N)$, we have $1 - g(X_1^D(N)) = d(1 - g(X_0^D(N))$ from characteristic zero, hence the Euler characteristic of the covering graph is the degree of the covering times the Euler characteristic of the quotient graph, which implies we have an unramified morphism of finite graphs, so the morphism of degenerate curves is unramified. This completes the proof of the theorem for the Shimura cover of $X_1^D(N) \to X_0^D(N)$.

Lemma 114 *The involution w_D acts trivially on the Shimura subgroup Σ of $X_1^D(N) \to X_0^D(N)$.*

Proof: Via the diagram

$$[Hom(\Gamma, S^1)]J_0^D(N)es, lw^* Hom(\Gamma_0^D(N), S^1)s, ri^\vee J(X')eHom(\Gamma_1^D(N), S^1)$$

this comes down to the evident fact that the modular involution w_D acts trivially on $\Gamma_0^D(N)$ modulo $\Gamma_1^D(N)$.

Applying the lemma to the diagram

$$[X_2^D(N)++]J_0^D(N)e, t/w_D s, lw^* J_0^{D+}(N)s, rw^* J_1^D(N)e, t/w_D J_1^{D+}(N)$$

we get that modding out by w_D induces a bijection on Shimura subgroups. Therefore, over C the Shimura cover of $X_1^{D+}(N) \to X_0^{D+}(N)$ is precisely $X_2^D(N)/w_D \to X_0^{D+}(N)$, and the groups involved are naturally isomorphic. Finally we must verify the same conclusion in positive characteristic l not dividing N. As above, we get this formally when l does not divide D. When l divides D, consider the commutative diagram of degenerate curves:

$$[Hom(\Gamma, S^1)]X_2^D(N)esX_0^D(N)sX_2^{D+}(N)eX_0^{D+}(N)$$

Now the two horizontal maps are obtained by modding out by images of the same finite subgroup. Since the top horizontal map does not reverse edges, neither does the bottom horizontal map. This completes the proof.

4.2 Preparation for Main Theorem 4: Galois representations arising from $\Gamma_0(N)$- structures

Let $\overline{P} \in X_0^{D+}(N)(F)$ be a rational point, with associated canonical field K. The basic dichotomy that we shall be wrestling with throughout the proof of the main theorem comes from the fact that $\overline{P}$ need not be induced by a PQM structure definable over F (recall that this occurs precisely when K splits B) but we want our theorem to apply to these "non-modular" points as well. Many of our arguments work more naturally in the modular case, and at several points we will give the argument first in this case and then discuss what modifications are necessary in the non-modular case.

So, suppose we are in the modular case – so K splits B – and choose a structure $(A, \iota, C_N)/K$ which induces $P \in X_0^D(N)(K)$. Associated to the cyclic subgroup C_N we have an isogeny character

$$r_K : G_K \to (Z/NZ)^\times.$$

From the work of Section 1.3 we know that since we started with an F-valued point $\overline{P}$, we have a canonical structure $(A, Q_N)/F$, where Q_N is $\mathcal{O}$-submodule generated by C_N. Since $Q_N \cong Z/NZ \oplus Z/NZ$, we have a two-dimensional Galois representation

$$r_F : G_F \to GL_2(Z/NZ)$$

with (at most) dihedral image with the property that the above isogeny character is the (diagonal) restriction of r_F. The fact that the isogeny character "comes from F" in this way gives us key information: by the definition of amenability, N is inert in F, so there are at most two places of K over N. Suppose that

there are exactly two places. Label them v_1, v_2, write I_1, I_2 for the respective inertia groups with respect to these two places, and write e_1, e_2 for the orders of the images of $r_K|_{I_1}$ and $r_K|_{I_2}$ respectively. Then the fact that the representation comes from F tells us that $e_1 = e_2$; in particular, if the representation is unramified at either place, it is everywhere unramified over N.

Suppose now we are in the nonmodular case – i.e., K does not split B, so our geometric point (A, ι, C_N) corresponding to $\overline{P}$ cannot be defined over K. Nevertheless it can be defined over many quadratic extensions of K, as follows: let M/F be any quadratic extension such that the compositum KM splits B (there are infinitely many). Then there is a KM-structure on (A, ι, C_N) and an M-structure on (A, Q_N). Accordingly, we have an isogeny character

$$r_{KM} : G_{KM} \to (Z/NZ)^\times$$

coming from a Galois representation

$$r_M : G_M \to GL_2(Z/NZ).$$

The character r_{KM} is "almost independent of the choice of M" in the following sense:

Lemma 115 *Suppose that for some (splitting) choice of M_1/F, the character r_{KM_1} is unramified at all places of KM_1 lying above N. Then for any other (splitting) choice of M_2/F, the character $r_{KM_2}^{24}$ is unramified at all places of KM_2 lying above N.*

Proof: Form the compositum $W = KM_1M_2$; since $(A_1, \iota_1, C_1)/W$, $(A_2, \iota_2, C_2)/W$ induce the same point P in moduli space, it follows as in the elliptic curve case that $r_1^{12}|_W = r_2^{12}|_W$ (because the group of automorphisms of a QM abelian surface is also cyclic of order dividing 12). So by our hypothesis on r_1 we have that $r_2^{12}|W$ is unramified at every place of W over N. Suppose that v is a place of KM_2 such that $r_2|_{I_v}$ is nontrivial. Choosing a place w of W over v we know that $r_2|_{I_w}$ has order dividing 12. Since $W_w/(KM_2)_v$ is at most a quadratic extension, this shows that $r_2|_{I_w}$ has order dividing 24, which was to be shown.

Notation: If K is a number field, we write $h(K)$ for the class number of K.

4.3 Beginning of the proof of Main Theorem 4

Let $\overline{P} \in X_0^{D+}(N)(F)$ and assume first that $\overline{P}$ is a modular point induced by some structure $(A, Q_N)/F$ with associated isogeny character r_K. We may assume that $N > 3$ and that N is sufficiently large so that it is unramified in K. Let K_{v_i} be the completions of K over N (so $i = 1$ or 2). As in [RI, Lemma 5.2], we get a factorization of r_K into $\alpha \cdot \chi^k$, where χ is the mod N cyclotomic character and α_i is unramified at K_{v_i}. We claim that (when $i = 2$) this factorization is independent of i, namely that $\alpha_i = \alpha_j, k_i = k_j$. This is immediate from the

considerations of the previous section: consider the character $r\chi^{-k_1}$; its restriction to I_1 is trivial, so its restriction to I_2 is also trivial, i.e., $\chi^{k_2-k_1}$ is trivial. Since N is unramified in K, we conclude $k_1 = k_2$ and thus $\alpha_1 = \alpha_2$. In view of this claim, we allow ourserves to write K_N for either one of the K_{v_i} and α for α_i.

A theorem of [Jordan II]) asserts that A acquires good reduction over a totally ramified extension K'/K_N of degree 4 or 6, so that $A[N]/\mathcal{O}_{K'}$ is a finite flat group scheme over a Henselian base whose absolute ramification index is 4 or 6. So [Raynaud, Corollaire 3.4.4] applies to the subgroup C_N exactly as in [RI, Proposition 5.1], and we get the conclusion that the values of k modulo $m := \frac{N-1}{2}$ are restricted to

$$k \equiv 0, 1, \frac{1}{3}, \frac{2}{3} \; mod \; m.$$

Note that we *cannot* have $k \equiv 1/2$ since we have assumed that $\frac{N-1}{2}$ is even.

Claim: The order of α is bounded independent of N.
Proof: Consider the Shimura covering

$$X_1^D(N) \stackrel{t}{\to} X_2^D(N) \stackrel{n}{\to} X_0^D(N);$$

recall that t divides 6 and $X_2^D(N) \stackrel{n}{\to} X_0^D(N)$ is finite étale over $SpecZ[\frac{1}{N}]$. Therefore, a direct modification of the twisting argument given in [RI][Lemma 5.3] shows that α^{72} is everywhere unramified, so $\alpha^{72h(K)}$ is trivial. For the convenience of the reader we reproduce the argument here: there is a cyclic field extension K'/K of order dividing nt whose ramification index at any place v of K over $p \neq N$ is divisible by t and hence by 6; also there is a K'-rational point $P' \in X_1^D(N)(K')$ projecting down to $P \in X_0^D(N)(K)$. Because $X_1^D(N)$ is a fine moduli space, P' corresponds to a unique $\Gamma_1(N)$-structured QM surface $(A', \iota', x' \in C_1')$; since the induced $\Gamma_0(N)$-structured QM surface $(A', \iota', < x' >)$ has the same modulus as (A, ι, C_1), they differ by an element of $H^1(G_{K'}, Aut(A))$. As we recalled in the previous section, the automorphism group of a QM surface (a fortiori of a $\Gamma_0(N)$-structured QM surface) is cyclic of order dividing 12; it follows that the 12th power of the isogeny character of (A, ι, C_1) equals the 12th power of the isogeny character of $(A', \iota', < x' >)$ so is trivial. This shows that r_K^{72} is unramified away from N, so that indeed it is everywhere unramified and $r_K^{72h(K)}$ is trivial.

Let us now consider the case when $\overline{P}$ is not a modular point. We will need to exploit the lemma of the previous section as follows: choose M_N/F to be a quadratic field extension such that KM_N splits B and N remains prime in M_N. Then, arguing as in the first paragraph of this section, we can write $r_{KM_N} = \alpha\chi^k$ where α is everywhere unramified over N.

Also by the same argument as in the modular case, it is true that the order of α is bounded by $72h(KM_N)$, but this visibly depends on N. To get around

this we use the considerations of the previous section: put $M_1 := M_N$ and take as M_2 some fixed choice of a splitting field. Once we observe that the lemma of the previous section is valid (with the identical proof) under $r_i \mapsto \chi^{-k} r_i$, we deduce that α_2^{12} is unramified everywhere over N, hence (using the Shimura covering as above) $\alpha_2^{12 \cdot 72} = \alpha_2^{864}$ is everywhere unramified. Hence the order of $\alpha := \alpha_2$ is bounded by $C = C(K, D) := 864h(KM_2)$, independent of N. (The 864 is clearly too large – we have shown complete "defensive indifference" in giving away powers of 6 – but we don't trouble ourselves to improve it here.)

4.4 End of the proof of Main Theorem 4

To achieve a unified presentation between the modular and nonmodular cases, in the former we put $M_2 := Q$, so that in both cases we are given a structure $(A, \iota, C_N)/KM_2$ with a corresponding factorization of the isogeny character into $\chi^k \cdot \alpha$, where α has order bounded independent of N. Now let p be a prime dividing D and $\mathcal{P}$ a prime of KM_0 lying over p. We have potential good reduction, so after making a totally ramified base extension we realize A as an abelian surface over the finite field $k := \mathcal{O}_{KM_2}/\mathcal{P}$.

Claim: Since p divides D, A/k is supersingular.

Proof: Writing V for the étale part of the p-adic Tate module tensored up to Q_p, we have that V is a representation space for $B \otimes Q_p$, which since p is a ramified prime, is a division quaternion algebra. But since V has dimension at most 2 as a Q_p-vector space and a division algebra admits no nontrivial representation of degree less than 4, we conclude that $V = 0$, which in dimension 2 is enough to ensure that $A \sim E^2$, where E is a supersingular elliptic curve.

From the general theory of QM surfaces we know that the isogeny is k-rational. In this way we identify the G_k-module structure on Q_N as being the direct sum of two identical copies of a cyclic order N subgroup $C_N \leq_k E[N]$. Then, for any base extension of k of cardinality q we have

$$\beta_{\mathcal{P}}(\sigma_q) q^k + \beta_{\mathcal{P}}(\sigma_q)^{-1} q^{1-k} \equiv a(F_q) \mod N$$

where $\beta_{\mathcal{P}}$ is the unramified $\mathcal{P}$-adic part of the character, and $a(F_q)$ is the trace of a *supersingular* elliptic curve. We take a residue extension of cardinality q^2 where q is sufficiently large so that both of the following hold:
• $\beta_{\mathcal{P}}$ is trivial on the q^2-Frobenius
• the F_{q^2}- rational endomorphism algebra of the associatd elliptic curve has stabilized and hence has Frobenius polynomial $(X - q)(X - q)$.
Indeed we can attain the first via taking the $864h(KM_2)$th power of k and the second by taking the 12th power, so it is clear that the necessary power is independent of N. We get

$$q^{2k} + q^{2-2k} \equiv 2q \mod N$$

Since k is not $\frac{1}{2}$, it is clear that this congruence is only a congruence and not an equality – hence it will be satisfied for only finitely many values of N. This completes the proof of the theorem.

4.5 A family of Shimura curves violating the Hasse principle (Main Theorem 5)

In this section we use the local analysis of the previous chapter, together with a "largeness" result on the Galois representation originally due to [Ohta] (and proved independently by the author) to deduce the following

Main Theorem 5 *Assume D is sufficiently large so that the genus of X^{D+} is at least 2. Then, for all sufficiently large primes l (with respect to D), there exist infinitely many imaginary quadratic fields K such that $X_0^D(l)/K$ violates the Hasse principle.*

Beginning of the proof: The hypothesis $g(X^{D+}) \geq 2$ ensures that the Shimura curve X^D has only finitely many *quadratic* points. This is a theorem of [Rotger I, Theorem 9]. Just to say a few words about it: the proof in turn relies on a pretty theorem of Abramovich and Harris that asserts that there are only two ways C/K a curve of genus at least 2 defined over a number field can have infinitely many quadratic points: either
i) there exists a degree 2 $\varphi/K : C \to P^1$ or
ii) there exists a degree 2 $\varphi/K : C \to E$, where E/K is an elliptic curve of positive Mordell-Weil rank.
In particular, such curves must be K-rationally hyperelliptic or bielliptic. The hyperelliptic Shimura curves X^D were computed by [Michon]; Rotger determines the bielliptic ones (using methods very close in spirit to those of this thesis, namely a combination of CM points and Cerednik-Drinfeld uniformization). Notice on the other hand that it is very easy to see that the set of D for which X^D is either Q-hyperelliptic or Q-bielliptic can be effectively bounded: one uses the fact that modulo an auxiliary prime of good reduction l, the supersingular locus provides a large enough supply of quadratic points to establish the following

Proposition 116 *Fix a positive integer d and a non-negative integer g. Then the set $\mathcal{D}_{d,g}$ of QM discriminants D such that the Shimura curve X^D admits a finite Q-morphism of degree d to a curve of genus g is finite and can be effectively bounded.*

Next we need to recall the theorem on largeness of Galois representations. Let A/K be a QM surface defined over a number field. Then the action of Galois commutes with the quaternionic action, so if $\rho_l : G_K \to Aut(V_l(A))$ is the associated ℓ-adic Galois representation, necessarily its image $\rho_l(G_K)$ is contained in the group of units of the commuting algebra of $B_D \otimes Q_l$ in $Aut(V_l(A))$ – i.e., the unit group of $(B_D \otimes Q_l)^{opp}$. If we restrict attention to l not dividing D, then as

we well know by now, $B_D \otimes Q_l \cong M_2(Q_l)$, so the Galois representation lands in a group isomorphic to $GL_2(Q_l)$, and in the compact subgroup $GL_2(Z_l)$. When l divides D, we $\rho_l(G_K)$ lands inside a group isomorphic to the unit group $\mathcal{O}_H^\times$ of the unique maximal order of the quaternion algebra $B_D \otimes Q_l$.

We can formally consolidate these two cases as follows: let $B/Z := (\mathcal{O}_D^\times)^{opp}$ viewed as a constant group scheme. Then compiling the various ℓ-adic Galois representations gives a homomorphism

$$\rho_{\hat{Z}} : G_K \to B(\hat{Z}),$$

the adelic Galois representation.

Theorem 117 *([Ohta])*
The image of the adelic Galois representation is open in $B(\hat{Z})$.

In particular, for every sufficiently large prime l prime to D, the image of G_K in the automorphism group of $A[l]$ is isomorphic to $GL_2(Z/lZ)$. Note that, as in the classical case, the existence of a K-rational $\Gamma_0(l)$-structure on A/K implies that the image of G_K is contained in a Borel subgroup of $GL_2(Z/lZ)$, so that we conclude:

Corollary 118 *Let A/K be a QM surface defined over a number field, and let P be the corresponding K-rational point of X^D. Let $\pi_l : X_0^D(l) \to X^D$ be the canonical map. Then, for all sufficiently large l, all the G_K orbits on $\pi_l^*(P)$ are nontrivial.*

Corollary 119 *When the genus of X^{D+} is at least 2, then for any number field K there is an absolute bound on primes l such that $X_0^{D+}(l)(K) \neq \emptyset$.*

Proof: By Faltings' theorem, $X^{D+}(K)$ will be a finite set of points $\{P_1, \ldots, P_n\}$. One can choose a (possibly much larger) number field L such that L is a field of definition for each of the P_i's; apply the previous corollary with L in place of K.

Remark: This explains why our Main Theorem 4 is only new when the genus of X^{D+} is 0 or 1.

Proof of the theorem: suppose that the genus of X^{D+} is at least 2, and consider the curve $X_0^D(l)$. Since X^D has only finitely many quadratic points, it follows from Ohta's theorem that for sufficiently large l $X_0^D(l)$ has *no* quadratic points. We will show that for our fixed D and all sufficiently large l there exist infinitely many imaginary quadratic fields K such that $X_0^D(l)/K$ violates the Hasse principle; it suffices to choose K such that the curve has points over every completion of K.

Since K is imaginary quadratic, $X_0^D(l)$ certainly has points over the Archimedean place. Let v be a place of K dividing D. If v is inert in K, then K_v/Q_p is an

unramified quadratic extension, so that $X_0^D(l)/K_v$ is an (untwisted) Mumford curve – in particular the Frobenius action on the components of the special fibre is trivial, so each component is isomorphic to P_1/F_{p^2}. Referring now to the dual graph, the degree of the corresponding vertex is a priori less than or equal to $p+1$, certainly there exists a point on P_1/F_{p^2} which is not an intersection point, and by Hensel's Lemma this implies that $X_0^D(l)(K_v)$ is nonempty. Consider now $X_0^D(l)/Q_l$; in Section 4.2 we saw precisely that for fixed D $X_0^D(l)(Q_l)$ is nonempty for all sufficiently large l; *a fortiori* by assuming that $l >> 0$ we get $X_0^D(l)(K_v) \neq$ for all v dividing l. Finally, consider a place v of K which does not divide Dl, so that $X_0^D(l)/K_v$ has good reduction. But now notice that if K_v/Q_p is a proper (quadratic) extension, then $X_0^D(l)(K_v)$ is non-empty, because the entire special fibre is smooth and the supersingular points are always defined over F_{p^2}. On the other hand, by the Riemann hypothesis for curves over finite fields, certainly $X_0^D(l)(Q_p)$ is nonempty for all sufficiently large primes p not dividing DL. Therefore, if we take l sufficiently large to ensure i) the nonexistence of quadratic points on $X_0^D(l)$ and ii) the existence of Q_l-rational points on $X_0^D(l)$ and take K to be any of the (infinitely many) imaginary quadratic fields which are inert at the primes dividing D and at the finitely many good primes p such that $X_0^D(l)(Q_p)$ is empty, then $X_0^D(l)/K$ has points everywhere locally but not globally, which was to be shown.

4.6 Bounds on cyclic torsion for PQM surfaces

So far in this chapter we have considered quaternionic $\Gamma_0(N)$-structures (A, Q_N), so that $Q_N \leq A[N]$ is a subgroup which is cyclic as $\mathcal{O}_D$-module and isomorphic as abelian group to $C_1 \oplus C_2 = Z/NZ \oplus Z/NZ$. Over a field of definition for the QM, this data is equivalent to the data of a direct summand $C_1 = e_1 Q_N$. As we observed in Corollary 87, assuming that (A, Q_N) is defined over a real number field K, the canonical field generates a proper quadratic extension $M = LK/K$ and the Atkin-Lehner w_D acts as on $A[N]$ as the nontrivial element of $G_{M/K}$ – in particular, it interchanges C_1 and C_2. In this section we exploit the restrictions this fact places on the possible rationally defined order N cyclic subgroups $C \leq_K A[N]$.

Proposition 120 *Let A/Q be an $\mathcal{O}_D$-PQM abelian surface of plus type, and let N be prime and prime to D, and let K be the canonical field (cf. Chapter 2). Suppose that there exists an order N (cyclic) subgroup $C \leq_Q A[N]$. Then the Galois group G_K acts on $A[N]$ by scalar matrices.*

Proof: Let $C \leq_Q A[N]$ be as in the statement of the theorem. We work now with the basechange of (A, C) to the canonical field K, over which the QM is defined. Let M be the $\mathcal{O}_D$-submodule generated by C. The upshot of the discussion preceding the statement of the theorem is that M *cannot* have rank 2 as a Z/NZ-module. We claim that, as a matter of elementary algebra, M is therefore forced to have rank 4, namely $M = A[N]$. Assuming this for the

moment, the result follows immediately: if $C' \leq A[N]$ is any other cyclic subgroup, then there exists $\alpha \in \mathcal{O}_D$ such that $C' = \alpha C$. Since G_K stabilizes C and $\mathcal{O}_D$ commutes with G_K, it follows that G_K stabilizes every cyclic subgroup of $A[N]$.

Proof of the claim: The action of $\mathcal{O}_D$ on $A[N]$ can be viewed as an action of $\mathcal{O}_D \otimes F_N \cong M_2(F_N)$ on a four-dimensional F_N-vector space $V = A[N]$. Therefore any $M_2(F_N)$-submodule $W \leq V$ decomposes into isomorphic subspaces $W = W_1 \oplus W_2 = e_1 W \oplus (1 - e_1)W$. So any nonzero $M_2(F_N)$-submodule has even F_N-dimension, qed.

Remark: This proposition – simple as it is – serves to reinforce that the *quaternionic* $\Gamma_0(N)$-structure Q_N is the interesting object in our situation. On the other hand, it *is* natural to be interested in the structure of the rational torsion subgroup of a PQM abelian surface A/Q. Since the existence of a rational point P of order N on A implies the existence of a rational cyclic structure $\langle P \rangle$, one expects that the structure of the rational torsion *prime to D* on a QM surface should be very restricted. This leads us to the considerations of the final chapter.

Chapter 5

Strong bounds on rational torsion for certain abelian varieties

5.1 Strong boundendess of rational torsion over local fields (Main Theorem 6)

Let K be a field. We say that *torsion is strongly bounded* for abelian varieties of dimension d over K if for all finite field extensions L/K and all d-dimensional abelian varieties A/L, there exists an $N = N(d, [L : K])$ such that $\#A(L)[tors] \leq N$. It would be equivalent to require the order of any individual torsion point on $A(L)$ to be bounded dependent only on d and $[L : K]$. (We would say that torsion is (merely) bounded if the N above were allowed to depend on L itself and not just its degree over K.)

For example, torsion is strongly bounded in every dimension d for abelian varieties over a finite field $K = F_q$, since if L/K is a field extension of degree n then the Weil conjectures imply $\#A(F_{q^n}) \leq (1 + q^{n/2})^{2d}$.

A celebrated theorem of [Merel] asserts that torsion is strongly bounded when $d = 1$ and $K = Q$. It is natural to conjecture that torsion is strongly bounded for abelian varieties A/Q in every dimension d, but this seems much beyond present reach.

The torsion is not bounded for elliptic curves over Q_p: the point p has exact order n on the Tate curve $E_{p^n} = G_m/\langle p^n \rangle$. However there is a subclass of abelian varieties over a non-Archimedean local field – containing the quaternionic surfaces – for which the torsion is strongly bounded. The goal of this section is to prove the following Let K be a non-Archimedean local field with

residue cardinality q and absolute ramification index e. Assume that $p > e-1$. Then the torsion is strongly (and effectively) bounded for d-dimensional abelian varieties A/K with potentially good reduction.
Proof of the theorem: we treat separately the cases of a point of order prime to p and of p-power order. For the former we can easily reduce to the above remark about finite fields:

Proposition 121 *Let A/K be a d-dimensional abelian variety with potentially good reduction over a non-Archimedean local field K of residue characteristic q. Then the order of the prime-to-q-torsion subgroup of $A(K)$ is bounded by $(1+\sqrt{q})^{2d)}$.*

Proof: According to [Serre-Tate], we can find a finite totally ramified extension L/K such that A/L has good reduction. We recall the argument: letting $\Gamma := G_{K^{unr}}$, we have the familiar short exact sequence

$$1 \to \Gamma \to G_K \to \hat{Z} \to 1,$$

and since $\hat{Z}$ is projective, the sequence splits, allowing us to (noncanonically) choose a subgroup Z of G_K complementary to Γ. Letting $M := \overline{K}^Z$, we get that M/K is a totally ramified extension, and using the fact that formation of the Néron model commutes with étale base change, it must be that A/M has good reduction; by definition of potentially good reduction, some finite subextension L of M will do. But now the prime-to-q torsion of $A(K)$ is contained in the prime-to-q torsion of $A(L)$, which is isomorphic to the prime-to-q torsion of the good reduction $A(F_q)$. As mentioned above, the Weil bound on the order of $A(k)$ is $(1+\sqrt{q})^{2d}$.

Next we must bound the p-power torsion. Recall the low ramification hypothesis: $p > e-1$. It follows that if $P \in A(K)[tors]$ is a point of order $N = p^k$, then the scheme-theoretic closure of $\langle P \rangle$ in the Néron model $\mathcal{A}/\mathcal{O}_K$ is still the constant group scheme Z/NZ, and we get an N-torsion point in the Néron special fibre. Switching notation slightly, we have the exact sequence

$$1 \to A^0 \to A \to \Phi \to 1$$

where A^0 is the connected component, an extension of an α-dimensional abelian variety by a β-dimensional connected linear group ($\alpha + 2\beta = d$) and Φ/F_q is the component group. For a finite commutative group G, write $e_p(G)$ for the largest integer i such that there exists an element of order p^i in G. We have

$$k = \log_p N \leq e_p(\#A^0(F_q)) + e_p(\#\Phi(F_q)).$$

Recall that U/F_q is a commutative unipotent groupscheme of dimension at most d, $U(F_q)$ is a p^d-torsion group (this follows from the fact that all such U are products of Witt vector schemes). On the other hand, the full order of an at most d-dimensional abelian variety over F_q is explicitly bounded as in the proposition, we only have to worry about the exponent of the Néron component group. So we are finished by the following

Theorem 122 *([McCallum]) Let A be an abelian variety over a non-Archimedean local field K with potentially good reduction. Let ϵ be the exponent of $\Phi(\overline{k})$, and let λ be the additive dimension of A^0. Factor $n = p_1^{e_1} p_2^{e_2} \dots p_r^{e_r}$, and let $L(n) = p_2 - 1$ if $n = 2p_2$, p_2 odd and $L(n) = \sum_{i=1}^{r} p^{e_i - 1}(p_i - 1)$ otherwise. Write $\epsilon = p^a \epsilon'$, with $(\epsilon', p) = 1$. Then $\max\{L(\epsilon'), L(p^a)\} \leq 2\lambda$.*

Remark: We conjecture that if K is any non-Archimedean local field then for every d torsion is strongly bounded among abelian varieties A/K of toric rank zero. As the proof of the theorem makes clear, in the low ramification case, all that needs to be shown is the uniform boundedness of the p-part of the exponent of the component group.

5.2 Bounds on rational torsion for abelian varieties with everywhere potentially good reduction over number fields

Let A/K be an abelian variety over a number field K with everywhere potentially good reduction. By the main theorem of the previous section, for any finite place v of K such that $p(v) - 1 > e(v)$, $A(K_v)[tors]$ can be uniformly bounded – so of course the torsion is uniformly bounded over K! In this section we roll up our sleeves and actually exhibit a relatively short list containing all the possible orders of the groups of rational torsion for an abelian surface A/Q with everywhere potentially good reduction. Let us be clear that from a theoretical perspective there is nothing new here – indeed, one can find in the literature the (easy) proof that torsion is strongly bounded among abelian varieties over number fields with everywhere good reduction. However, the published bounds are vertiginously large; in contrast the bounds that we obtain are small enough so that one is actually tempted to find examples to show that the list is complete.

The first and quickest thing to say is that one can simply apply Proposition 112 to two different completions (that is, we can bypass the nontrivial part of the theorem of the previous section); we get

Corollary 123 *(Explicit strong boundedness over number fields) Let A/K be a d-dimensional abelian variety over a degree $n = [K : Q]$ number field. Assume that A has potentially good reduction at places v_2,v_3 of K over 2 and 3. Then*

$$\#A(K)[tors] \leq [(1 + 2^{n/2})^{2d}][(1 + 3^{n/2})^{2d}].$$

Remarks: Assume $K = Q$. When $d = 1$ the bound obtained on the rational torsion is 35 and when $d = 2$ we get a bound of 1815.

But we have lost a lot of information by multiplying the prime-to-2 torsion by the prime-to-3-torsion. To overcome this, we first record that $\#A(F_2) \leq 33$ and $\#A(F_3) \leq 55$. The first inequality already tells us that the largest possible

prime dividing the order of the rational torsion group is 31; working prime-by-prime, the largest possible prime-powers dividing the torsion subgroup could be:

$$2^5, 3^3, 5^2, 7, 11, 13, 19, 23, 29, 31,$$

and we can already assemble a relatively small list of possible orders: they will be of the form $2^a \cdot y$, where $0 \leq a \leq 5$ and y lies in the set

$$S_{odd} = \{1, 3, 5, 7, 3^2, 11, 13, 3 \cdot 5, 17, 19, 3 \cdot 7, 23, 5^2, 3^3, 29, 31, 3 \cdot 11\}.$$

But there is a further improvement to make: once $d > 1$, the Weil bound is not the last word on the possible orders of the Mordell-Weil groups of abelian varieties over F_p. Rather we can use Honda-Tate theory to give complete lists of $\#A(F_p)$ for various small p. Indeed one can explicitly parameterize the F_p-rational isogeny classes of abelian surfaces by the Galois conjugacy classes of Frobenius roots, and the lattter by means of Weil numbers (for all this see the Appendix). Notice also that if $A_1 \sim_{F_p} A_2$, then $\#A_1(F_p) = \#A_2(F_p)$ – indeed the number of rational points of an abelian variety over a finite field is computed as the determinant of $F - 1$ acting on any ℓ-adic Tate module (choose an l prime to the degree of the isogeny). So a little honest toil yields the following useful information:

$$\#A(F_2) \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 25\}.$$

$$\#A(F_3) \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25$$
$$28, 29, 30, 34, 35, 36, 42, 49\}.$$

$$\#A(F_5) \in \{4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28$$
$$29, 30, 31, 32, 33, 34, 35, 36, 37, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 54, 55, 56, 58, 59,$$
$$60, 61, 62, 63, 64, 70, 71, 72, 79, 80, 81, 90, 100\}.$$

A couple of comments: Up to F_p-rational isogeny there are three essentially different kinds of abelian surfaces over F_p. The first is a product of two elliptic curves; since it is known (Hasse-Deuring-Waterhouse) that for elliptic curves E/F_p, any order in between $p + 1 - 2\sqrt{p}$ and $p + 1 + 2\sqrt{p}$ can be attained as $\#E(F_p)$, it is easy to write down such orders. There is one F_p-simple abelian surface with a real quadratic Weil number – namely $\sqrt{p}$ – which has $(1 - \sqrt{p})^2(1 + \sqrt{p})^2 = (p - 1)^2$ points on it and the weird (unstable) endomorphism algebra $B_{\infty_1,\infty_2}/Q(\sqrt{p})$ (Case 2 in our classification of Weil numbers). The remaining abelian surfaces are F_p-simple and their rational endomorphism algebra, a quartic CM field, is generated by Frobenius. These quartic Weil numbers are computed via their associated real quadratic $\beta = \pi + p/\pi$.

Taking into account these three lists, we can reduce the possible orders to

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 22, 24, 25, 30, 36\}$$

This is a pretty short list! Notice that since one knows that there exist elliptic curves E/Q with potentially good reduction (aka integral j-invariant) and torsion of orders 1 through 6, it follows that we can produce Q-split examples of many of the elements of the list. The orders which do not arise in this way are

$$S_{nonsplit} \subset \{7, 11, 13, 14, 19, 22\}$$

We have checked that all of these orders do indeed arise over all finite fields of cardinality < 100 – in short, none of our methods can rule out their existence, and we may as well be optimistic and conjecture that they do arise.

A last-minute remark: I spoke recently (the end of April, 2003) on this material in the number theory seminar at Harvard. When I put up the above list, Barry Mazur and Noam Elkies each immediately observed that $J_1(13)/Q$ is an abelian surface with everywhere potentially good reduction (as follows by combining the theorem of [Katz-Mazur] that $J_1(p)/J_0(p)$ has potentially good reduction at p with the fact that $X_0(13)$ has genus zero) and a rational torsion point of order 19. Indeed $J_1(13)$ is an example of a potentially quaternionic surface with $D = 1$ – i.e., its geometric endomorphism algebra is $M_2(Q)$ and the splitting takes place over the (real) degree six extension $Q(\zeta_{13} + \zeta_{13}^{-1})$. It would be interesting to look for similar examples.

5.3 Bounds on the order of a torsion point on a PQM surface

If A/Q is a PQM surface, it has everywhere potentially good reduction, so the results of the previous section apply to give bounds on the possible order of the rational torsion subgroup. Of course, since we have acquired a thesis worth of information about PQM surfaces we should expect to be able to say more in this case! Indeed we have the following

Proposition 124 *Let A/Q be an $\mathcal{O}_D$-PQM surface (not necessarily of plus type). Then the possible orders for $A[Q][tors]$ lie in the following set:*

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 24, 25, 36\}.$$

Proof: To be sure, what remains to be done after the work of the previous section is to show that the orders $11, 13, 15, 19, 22, 30$ cannot arise. Now, we know that the QM becomes defined over at worst a $(2, 2)$-extension. But the only $(2, \ldots, 2)$-extension of a finite field is a 2-extension, so whether A/Q is of plus type or not, the QM still becomes defined over at worst F_{p^2}. By Honda-Tate theory (see Appendix) we know there exists an elliptic curve E/F_{p^2} such that $A \sim_{F_{p^2}} E^2$. As in the previous section it follows that $\#A(F_{p^2}) = \#E(F_{p^2})^2$. It follows that if N is a squarefree odd number dividing the order of the rational torsion, then N less than or equal to the maximum possible order of an elliptic curve over F_{2^2}, namely 9. This allows us to eliminate everything in our list.

If we ask instead about the possible orders of a given torsion *point* and restrict to N prime to the quaternionic discriminant D (so that there is a connection with $X_1^D(N)$) we can further reduce the list, as follows:

Proposition 125 *Let A/Q be an $\mathcal{O}_D$-PQM abelian surface of plus type. Then:*
a) If 2 does not divide D there is no 8-torsion point in $A(Q)$.
b) If 3 does not divide D there is no 9-torsion point in $A(Q)$.
c) If 5 does not divide D there is no 5-torsion point in $A(Q)$.
d) If 7 does not divide D there is no 11-torsion point in $A(Q)$.

Proof: Let N be prime to the quaternionic discriminant D. Let $P \in A[N](Q)$ be a Q-rational point of order N on our plus-type PQM surface A/Q. Let $C_N := \langle P \rangle$ be the corresponding cyclic group. By Proposition 120, there exists an imaginary quadratic field K such that G_K acts on $A[N]$ by scalar matrices; but G_K acts trivially on P, and it follows that G_K acts trivially on $A[N]$. Assume $N = p^r$ is a prime-power; we know that the determinant of the representation of G_K on $A[N]$ is the square of the mod N cyclotomic character, so that N is such that $\chi_N^2 | G_K = 1$. The field cut out by χ_N^2 is $Q[\zeta_N + \zeta_N^{-1}]$ which is totally real; hence, if were a proper extension of Q it would be disjoint from K. Therefore the putative trivialization can occur only when $N = 1, 2, 3, 4$ or 6.

On the other hand, if p divides D then A/F_p necessarily has supersingular reduction, and we get a smaller list of Mordell-Weil groups to check. One does not get quite as much mileage out of this as one might suppose, but we get the following additional

Proposition 126 *Let A/Q be an $\mathcal{O}_D$-PQM abelian surface of plus type. If 3 divides D then there is no rational torsion point of order 7. If 5 divides D there is no rational torsion point of order 14.*

Proof: If 3 divides the QM discriminant, A/F_3 is necessarily supersingular. The possible Mordell-Weil groups for a supersingular elliptic curve E/F_9 are $Z/4, Z/10, Z/13$ and $Z/4 \oplus Z/4$. Thus we can cross off 7 and 14 from the list. If 5 divides the QM discriminant, A/F_5 is necessarily supersingular, and the possible Mordell-Weil groups for a supersingular elliptic curve E/F_{25} are $Z/6 \oplus Z/6, Z/31, Z/21, Z/4 \oplus Z/4$ and $Z/31$. Thus we can cross 14 from the list.

5.4 Applications to $X_1^D(N), X_1^{D+}(N)$

Proposition 127 *Let K/Q_p be a p-adic field with residue field F_q and absolute ramification index e, and assume that $p > e - 1$. Then for sufficiently large N (depending only on q) we have $X_1^D(N)(K)$ is empty. If we restrict to N prime*

to p the same statement is true without restriction on the absolute ramification index.

Proof: If $N \geq 4$, $X_1^D(N)/K$ is a fine moduli space for QM abelian surfaces together with a K-rational (QM) point of order N. Thus the result follows immediately from Main Theorem 6.

Remark: Of course the same result holds for $X_1^{D+}(N)$ (for N prime to the residue characteristic) since a K-rational point on $X_1^{D+}(N)$ comes from an L-rational point on $X_1^D(N)$ for some quadratic extension L/K.

To summarize, the Shimura curves $X_1^D(N)$ give a family of curves as remarkably resistant to having rational points (even!) over local fields as any family I have ever seen. In the remainder of this section we show that the family $X_1^D(N)/F_q$ where N is prime and $(q, DN) = 1$ is "within a factor of 4 of being optimally pointless," in a certain sense that we are about to make precise:

Let C/K be a (smooth irreducible projective) genus g curve over a field K. Define its *pseudoindex* $s(C/K)$ to be the least degree of a field extension L/K over which C acquires an L-rational point. This quantity is to be compared to the *period* $p(C/K)$ and the *index* $i(C/K)$ which are respectively the minimal positive degree of a K-rational divisor class and of a K-rational divisor (in these latter two cases, it would be equivalent to take the gcd instead of the minimum). Notice that we could recast the definition of the pseudoindex in terms of the least positive degree of an *effective* K-rational divisor on C, which hopefully explains the "pseudo" (a neologism due to the author). One has

$$p(C/K) \mid i(C/K) \leq s(C/K) \leq 2g - 2,$$

where the last inequality – only valid for $g \geq 2$ – comes by applying Riemann-Roch to the canonical class. The period/index/pseudoindex problem is to understand how these quantities relate to each other in terms of the field K. When $g = 1$ the index and the pseudoindex coincide and this is nothing else than the period-index problem in the Weil-Chatelet $H^1(K, E)$ of an elliptic curve E/K. When $g > 1$ it is still far from understood (e.g., one does not know a necessary and sufficient condition for a hyperelliptic curve over Q_p to have index 1!) and the relationship between the index and the pseudoindex seems almost completely unexplored. The problem is still of some interest in the case of finite fields; we have the following

Proposition 128 *Let C/F_q be a curve of genus g over a finite field. Then $i(C) = 1$ and $s(C) - 1 \leq 2\log_q(g) + log_q(4)$.*

Proof: By the Weil conjectures, any variety V/F_q will have points over F_{q^d} for all sufficiently large d. In particular if d is large enough it has points over F_{q^d} and $F_{q^{d+1}}$, hence it has F_q-rational divisors D_1, D_2 of degrees d and $d+1$. Since $(d, d+1) = 1$ there exist integers m, n (not both positive!) such that

$D = mD_1 + nD_2$ is an F_q-rational divisor of degree 1, showing $i(V/F_q) = 1$. Let $d = s(C) - 1$, so that C/F_{q^d} is a genus g curve without any rational points. By the Weil bound we have $q^d \leq 2gq^{d/2}$ or $d \leq \log_q(4g^2)$.

Proposition 129 *Let q be a prime power and N a prime number prime to qD. Then the (smooth) Shimura curve $X_1^D(N)/F_q$ has pseudoindex $1/2\log_q(g) - O(1)$.*

Proof: The genus of $X_1^D(N)$ is $O(N^2)$, and taking $N \geq 4$ to get a fine moduli space, we know that $X_1^D(N)(F_{q^n}) \neq \emptyset$ implies the existence of an order N^2 subgroup in the Mordell-Weil group of an abelian surface A/F_{q^n}, so that $N^2 \leq (1 + q^{n/2})^4$ or $\log_q(N) \leq n + O(1)$. Since there exists a constant C' such that $N \geq C'\sqrt{g}$, we get $n \geq \log_q(g) - O(1)$ as claimed.

Remark: Improvements of this proposition are discussed in [Clark-Elkies].

Appendix: Explicit Honda-Tate theory for abelian surfaces

This appendix was written two years before the rest of the thesis[1] – that is, much earlier in the author's graduate career. We give a quite down-to-earth treatment of the Honda-Tate theory of the "isogeny category" of abelian varieties over finite fields with an eye towards ready computability of the endomorphism algebra of a principally polarized abelian surface arising as the Jacobian of a hyperelliptic curve. Assuming that one has the Hasse-Weil zeta function (which we compute in a completely naive way, i.e., by directly counting F_q- and F_q^2-rational points), the Honda-Tate theory immediately gives us the F_q-rational endomorphism algebra. But it is at least as desirable to have the *geometric* endomorphism algebra, i.e., $End^0_{\overline{F_q}}(A)$ (actually, it is nice to have both; computing these algebras for various mod p reductions of an abelian surface A/Q is a reasonably good technique for getting at the endomorphim algebra of A itself, the computation of which is an open problem; see [Poonen]). For this, it is enough to compute the F_{q^n}-rational endomorphism algebra for some sufficiently large n; the main point of the appendix is to give a reasonable "universal" value of n. An explicit, nearly optimal value of n was found by David Savitt at the author's request circa spring 2000. I thank him again lo these many years later.

Weil numbers

Let $q = p^a$, p a prime number. A *Weil q-number* is an algebraic integer π such that for every Archimedean place $|\ |$ of $Q(\pi)$, $|\pi| = \sqrt{q}$. Let A be an abelian variety defined over the finite field $k = F_q$. It is known [Milne] that the roots of $P_A(T)$, the characteristic polynomial of Frobenius $/F_q$, are Weil q-numbers. Let E_k be the algebra of endomorphisms of A which are *defined over k*; write $A \sim_k B$ if A and B are isogenous over k. Honda-Tate theory consists of the following assertions (and their proofs!):

[1] We ignore the paradoxical issues arising from this clause!

a) $A \sim_k B$ if and only if $P_A(T) = P_B(T)$; in particular, $A \sim_k B_1 \times B_2$ if and only if $P_A(T) = P_{B_1}(T)P_{B_2}(T)$.
b) Assume A is k-simple, i.e., A is not k-isogenous to a nontrivial product. Then (easily) E_k is a division algebra and the minimum polynomial of Frobenius $/k$ is irreducible, so $P_A(T)$ is a power of an irreducible polynomial. Thus any two roots of $P_A(T)$ are Galois conjugates. Let π be any root. Then $A \mapsto \pi$ gives an equivalence of categories, from the category of k-simple abelian varieties up to k-isogeny to the category of Weil q-numbers up to Galois conjugacy.
c) The center of E_k is $Q(\pi)$. Writing $f = [Q(\pi) : Q]$, $e^2 = [E_k : Q(\pi)]$, we have that $P_A(T) = P_1(T)^e$, where $P_1(T)$ is an irreducible polynomial. Moreover, the local invariants of the $Q(\pi)$-central division algebra E_k are given as follows:
$i_\infty = 1/2$ at every real Archimedean place (if any) of $Q(\pi)$
$i_{\mathcal{P}} = 0$ for every prime ideal $\mathcal{P}$ of $Q(\pi)$ not lying over p
For $\mathcal{P}/p$, $i_{\mathcal{P}} = \frac{f_{\mathcal{P}} ord_{\mathcal{P}} \pi}{a}$, where $f_{\mathcal{P}}$ is the inertial degree at $\mathcal{P}$ and $ord_{\mathcal{P}}$ is the normalized valuation at $\mathcal{P}$. Finally, $2 \dim A = ef$.

In this section we carry out a classification of Weil numbers as far as we need to classify abelian surfaces defined over a finite field up to isogeny. In Section 2 we consider the question of stability of Weil numbers. In Section 3 we see how the (very classical) theory of endormorphism algebras of elliptic curves over a finite field follows easily from our analysis, and finally in Section 4 we classify the endomorphism algebras that can occur for an abelian surface.

Stability of Weil numbers: For most applications one is interested in the $\overline{F_p}$-endomorphism algebra (sometimes called the "geometric" endomorphism algebra for emphasis); the k-rational endomorphism algebra is a means to this end. Certainly given any A/F_q, $End_{\overline{F_p}} A = End_{F_{q^n}} A$ for sufficiently large n. Upon extending the base field $F_q \mapsto F_{q^n}$ we take Frobenius to its nth power and hence $\pi \mapsto \pi^n$ and $Q(\pi) \mapsto Q(\pi^n)$. Thus, knowing the k-rational endomorphism algebra gives enough information to compute the geometric endomorphism algebra. Upon extending the base, the following phenomena can occur: the rank of E_k can increase (but not decrease, clearly), the rank of the center can decrease (but not increase, curiously), and A may become nonsimple. Let us write $E = End_{\overline{F_p}} A$. If $E_k \neq E$ we call E_k *unstable*. We call the process of making sure that k is sufficiently large to ensure equality *stabilization*. For computational purposes it is key to know explicitly a base extension large enough to ensure stability. We take up this issue in Section 2.

Classification of Weil numbers: We begin with the following useful observation: since $f_{\mathcal{P}} ord_{\mathcal{P}} \pi$ is integral, the central simple algebra E_k is unramified at $\mathcal{P}$ if and only if $a/f_{\mathcal{P}} ord_{\mathcal{P}}$. Thus we have equivalent conditions: E_k is commutative if and only if $e = 1$ if and only if $a/f_{\mathcal{P}} ord_{\mathcal{P}}$ for all $\mathcal{P}$ and $Q(\pi)$ is totally imaginary. In particular, when A is defined over F_p (i.e., $a = 1$), then when $Q(\pi)$ is totally imaginary, the F_p-rational endomorphism algebra of A is always

commutative.

First, consider the case of a Weil q-number π such that $Q(\pi)$ has a real embedding. Then under $Q(\pi) \hookrightarrow R$, $\pi^2 = |\pi|^2 = q$, i.e., $\pi = +/- \sqrt{q}$.
Case 1: $Q(\pi) = Q$. Then $q = p^a$ with a even. We have a unique real Archimedean place ∞ at which $i_\infty = 1/2$. On the other hand, since the sum of the local invariants is 0 $(modZ)$ and all the invariants away from p vanish, we must have $i_p = 1/2$. Then e, being the lcm of the denominators of all the local invariants, is 2, so $f = 1$, so the unique k-simple abelian variety A associated to π has dimension 1 and $E_k = B_{p,\infty}/Q$, the quaternion algebra over Q ramified at precisely p and ∞. That is, A is a supersingular elliptic curve with all endormorphisms defined. Note well that we have just shown that this situation does not occur over any odd-degree extension of F_p.
Case 2: $Q(\pi) = Q(\sqrt{p})$, so $\pi = \sqrt{q}$, $q = p^a$ with a odd. There are two real Archimedean places ∞_1, ∞_2 at which $i_{\infty_1} = i_{\infty_2} = 1/2$. As p ramifies in $Q(\pi)$, there is a unique prime $\mathcal{P}$ over p with $i_{\mathcal{P}} = \frac{f_{\mathcal{P}} ord_{\mathcal{P}} \pi}{a} = \frac{ord_{\mathcal{P}} \sqrt{q}}{a} = \frac{1/2 a ord_{\mathcal{P}} p}{a} = 1 \equiv 0(Z)$. Therefore $e = 2$ and $f = 2$ so $\dim A = 2$, i.e., the associated k-simple A is an abelian surface with $E_k \cong B_{\infty_1,\infty_2}/Q(\sqrt{p})$. But E_k is *unstable*: $q \mapsto q^2$ takes us back to Case 1; over the extended field, $P_A(T) = (T - p^a)^2$, so by the Honda-Tate theory above, A is geometrically isogenous to the square of a supersingular elliptic curve.

Thus it remains to consider Weil numbers π with $Q(\pi)$ a totally imaginary field. In this case, it turns out that $Q(\pi)$ is a CM-field. Indeed, put $\beta = \pi + q/\pi$. Fix any embedding $Q(\pi) \hookrightarrow C$ and write $\overline{x}$ for the complex conjugate of x. We then have $\pi\overline{\pi} = q$, so $\beta = \pi + \overline{\pi}$ is totally real. Moreover the equation $\pi^2 - \beta\pi + q = 0$ exhibits $Q(\pi)$ as a quadratic extension of $Q(\beta)$. (Conversely, if β is a totally real algebraic integer with $|\beta| \leq 2\sqrt{q}$ in every embedding, then we can *define* π by $\pi^2 - \beta\pi + q = 0$ and then π is a Weil q-number. This is often useful in the construction of Weil numbers; see [Waterhouse].) In particular, unless π is rational, $Q(\pi)$ has even degree.

Let us try to classify Weil numbers π such that $Q(\pi)$ is an imaginary quadratic field and see what happens.

Case 1: p is inert in $Q(\pi)$. There are no real Archimedean places. Moreover, $i_p = \frac{f_p ord_p \pi}{a} = \frac{2 ord_p \pi}{a}$. Now, since $\pi\overline{\pi} = q$, we have $ord_p(\overline{\pi}) = ord_{\overline{p}}(\overline{\pi}) = ord_p(\pi)$, so $ord_p(\pi) = 1/2 ord_p(q) = a/2$. We conclude i_p is integral, so $e = 1$ and $dim A = ef/2 = 1$, so A is an elliptic curve with $E_k = Q(\pi) = K$ an imaginary quadratic field. However, we claim that π is unstable, so that the elliptic curve A is supersingular. Indeed, $ord_p(\pi^2/q) = 0$, and since $ord_l(\pi^2/q) = 0$ for all l not equal to p, π^2/q is a unit in the ring of integers of the imaginary quadratic field K, and hence a root of unity, which shows that some power of π is rational.

Case 2: p ramifies in $Q(\pi)$. Let $\mathcal{P}$ be the unique prime over p; by an argu-

ment as above we have $ord_{\mathcal{P}}(\pi) = 1/2ord_{\mathcal{P}}(q) = 1/2aord_{\mathcal{P}}(p) = a$, so $i_{\mathcal{P}}$ is integral and $e = 1$. A is again an elliptic curve with $E_k = Q(\pi)$ an imaginary quadratic field. Again π is unstable, by the same argument as above, and A is supersingular.

Case 3: p splits in $Q(\pi)$. Let $\mathbf{p}_1, \mathbf{p}_2$ be the two primes lying over p. In this case $f_{\mathbf{p}_1} = f_{\mathbf{p}_2} = 1$ and $ord_{\mathbf{p}_1}(\pi)$ is not uniquely determined: if $x = ord_{\mathbf{p}_1}(\pi), \pi\overline{\pi} = p^a$ implies $ord_{\mathbf{p}_2}(\pi) + ord_{\mathbf{p}_2}(\overline{\pi}) = a$. Using $\overline{\mathbf{p}_1} = \mathbf{p}_2$, $ord_{\mathbf{p}_2}(\overline{\pi}) = ord_{\overline{\mathbf{p}_1}}(\overline{\pi}) = ord_{\mathbf{p}_1}(\pi) = x$, so $ord_{\mathbf{p}_2}(\pi) = a - x$ and this is consistent: $i_{\mathbf{p}_1} + i_{\mathbf{p}_2}$ is integral. Write $D(\frac{x}{a})$ for the reduced denominator of $\frac{x}{a}$, we have $e = \frac{1}{D(\frac{x}{a})}$. Then $dimA = e$ and E_k is the central simple algebra over the quadratic imaginary field $K = Q(\pi)$ which has index e ramified only at p with invariant x/a. Since π^2/q is a root of unity if and only if its valuations at the two primes $\mathbf{p}_1$ and $\mathbf{p}_2$ are equal, we see by the same considerations as in Cases 1 and 2 above that π is stable if and only if $x = ord_{\mathbf{p}_1}(\pi) \neq \frac{a}{2}$. To get an idea of what can happen in this case, assume for example that $\mathbf{p}_1$ and $\mathbf{p}_2$ are principal ideals, so that we know there is an element π corresponding to any choice of x; then, as $q \to \infty$ the Weil numbers considered here correspond to k-simple abelian varieties of arbitrarily large dimension. We will call this the split case, and such a π a *split quadratic Weil number*.

Stabilization of Weil numbers

The Stabilization Problem: Let π be a Weil q-number with $[Q(\pi) : Q] = 2n$. Recall that π is *stable* if $Q(\pi^k) = Q(\pi)$ for all positive integers k, and that N is a *stabilizer* of π if π^N is a stable Weil q^N-number. For computational prurposes it is convenient, given n, to have an N that stabilizes every degree $2n$ Weil number. This is always possible, as the following simple argument shows.

Proposition 130 *There is a (readily computable) function $N = N(n)$, not depending on q, such that $N(n)$ stabilizes every degree $2n$ Weil q-number.*

Proof: We first consider the inherently easier case $n = 1$. If π is a quadratic Weil number, it is unstable if and only if some power of π is rational if and only if some power of π^2/q is rational. But π^2/q has absolute value 1, so π is unstable if and only if $\pi^2/q = \zeta_k$, a kth root of unity. Visibly ζ_k is at most quadratic, so $k \in \{1, 2, 3, 4, 6\}$ and ζ_k^6 is rational, so π^{12} is rational. We can thus take $N(1) = 12$.
Now assume $n > 1$. Let N be any positive integer. Then $Q(\pi^N)$ is a proper subfield of $Q(\pi)$ if and only if π^N has fewer than $2n$ distinct Galois conjugates. Assuming this is so, since π certainly has $2n$ Galois conjugates and the Galois conjugates of π^N are the Nth powers of the Galois conjugates of π, there must exist two distinct Galois conjugates $\pi' \neq \pi$ such that $\pi'^N = \pi^N$. Then $\pi = \zeta_k\pi'$, where ζ_k is a kth root of unity for some k dividing N. Moreover $\zeta_k \in Q(\pi, \pi')$. Since $n > 1$ the latter field has degree $2nl \leq 2n(2n-2) = b(n)$, since l is an even number less than $2n$. Therefore k is such that $\phi(k) \leq b(n)$. Let N_0 be the

least common multiple of all such k. Then $\pi^N = \pi'^N$ implies $\pi^{N_0} = \pi'^{N_0}$, so N_0 is a stabilizer for π independent of q. $\diamondsuit$

Let us examine the bound given by Proposition 1. When $n = 1$, what we have shown is equivalent to: if A/F_q is an elliptic curve with Frobenius root π, A is (geometrically!) supersingular if and only if π^4 or π^6 is rational. Conversely, $\pi = 1+i$ is a Weil 2-number which does not become rational until raised to the fourth power, and $\pi = \frac{3+\sqrt{-3}}{2}$ is a Weil 3-number becoming rational only when raised to the sixth power. This shows that $N(1) = 12$ is sharp. (On the other hand, if we allow N to depend on q we can get sharper bounds, e.g. it will follow from the work of Section 3 that if $q = p > 3$ we can take $N(1, p) = 2$. See [Waterhouse] for a comprehensive discussion of this and other fine points concerning endomorphisms of elliptic curves.)
When $n = 2$, $b(n) = 8$, and the bound given by Proposition 1 is
$lcm\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 24, 30\} = 5040$, which is already a bit far from the truth. In our own computations with Weil numbers of abelian surfaces we have used the bound $N(2) = 120$, so we explain how this improvement can be derived using an auxiliary result which is interesting in its own right.
Let π be a $2n$-dimensional Weil number for $n > 1$ and let L be the Galois closure of $Q(\pi)$ viewed as a subfield of C. Label the Galois conjugates of π as $(\pi_1 = \pi), \pi_2, \ldots, \pi_{2n}$ such that for $1 \leq k \leq n$, $\pi_{2k+2} = \overline{\pi_{2k+1}} = \frac{q}{\pi_{2k+1}}$. We observe that a Galois automorphism $\sigma \in G = Gal(L/Q)$ is uniquely specified by its action on $\pi_1, \pi_3, \ldots, \pi_{2n-1}$ and indeed induces a permutation on the n element set of C-conjugate pairs $\{\{\pi_1, \pi_2\}, \ldots, \{\pi_{2n-1}, \pi_{2n}\}\}$. The set of all such permutations of the Galois conjugates forms an order $2^n n!$ subgroup $\mathcal{D}$ of the full permutation group in which our Galois group G is constrained to lie, so we deduce in particular that $\#G \;/\; 2^n n!$. We say that the Weil number π is *maximal* if we have equality, i.e., $G = \mathcal{D}$.

Proposition 131 *Maximal Weil numbers are stable.*

Proof: Let $H := Gal(L/Q(\pi))$, so H is the subgroup of $\mathcal{D} = Gal(L/Q)$ consisting of elements which fix π_1 (and hence also π_2). If π is unstable, then for some N, π^N generates a proper subfield of $Q(\pi)$ and thus gives rise to a subgroup J, $H \subset J \subseteq \mathcal{D}$. We claim that J contains the permutation (12) (where we have identified i with π_i). Indeed, since J properly contains H, J has a permutation $\sigma : 1 \to a$, $a > 1$. If $a = 2$, then since H contains every element of $\mathcal{D}$ fixing 1 and 2, $\sigma(12) \in J$, so $(12) \in J$. If $a > 2$, then the two-cycle $\tau = (a\overline{a}) \in H$, and $\sigma^{-1}\tau\sigma$ is again an element of J taking $1 \to 2$. We conclude that J contains $\langle (12), H \rangle$. But $L^{\langle (12), H \rangle}$ is the totally real subfield $Q(\pi_1 + \pi_2)$ of $Q(\pi)$. Thus π^N is a totally real Weil number and π^{2N} is rational. Arguing as in the first part of the proof of Proposition 1, we have $\pi \in Q(\sqrt{q}, \zeta_{2N})$, hence $Q(\pi)$ is a subfield of the cyclotomic field $Q(\zeta_{4pN})$. Thus $Q(\pi)/Q$ is itself a Galois extension, so the degree of the Galois closure is $2n$, not $2^n n!$. This contradiction implies that π is stable. $\diamondsuit$

Putting together Propositions 1 and 2, we can get a better uniformizer for quartic Weil numbers π: with notation as in Proposition 1, consider the extension $Q(\pi, \pi')/Q$; it has degree 4 or 8. But if it has degree $8 = 2^2 2!$, π is a maximal, hence stable, Weil number. So if π is unstable we must have $[Q(\pi, \pi') : Q] = 4$, and as in the proof of Proposition 1, we can then take N to be the least common multiple of all k such that $\phi(k) \leq 4$, which is $lcm\{1, 2, 3, 4, 5, 6, 8, 10, 12\} = 120$. On the other hand, by taking Weil p^2-numbers of the form $p\zeta_k$, with $k = 3, 5, 8$ we see that a stabilizer N for quartic Weil numbers must be at least $3.5.4 = 60$, so $N = 120$ is sharp to within a factor of 2.

Acknowledgement: The proof of Proposition 1 and the statement and proof of Proposition 2 were generously supplied by David Savitt upon the request of the author.

Applications to Elliptic Curves

The main goal of these notes is to apply the theory of the preceding sections to study abelian surfaces. To do this it is indispensable to have the corresponding theory for elliptic curves over finite fields, so we develop them here for completeness. Most of the results we obtain are very well-known and obtainable by more direct methods; nevertheless it is instructive to see how easily they can be derived from the Honda-Tate theory.

So let A/k be an elliptic curve. A is certainly simple, so Honda-Tate implies that A is determined up to k-isogeny by its frobenius root π. In this case, the characteristic polynomial of frobenius is $P_A(T) = T^2 - a_1 T + q$. If $N_1 = \#A(F_q)$, then using the 2×2- matrix identity $trace(\phi) = 1 + \det(\phi) - \det(1 - \phi)$ and the equalities $\det(\phi) = \deg(\phi) = q$, $\det(1 - \phi) = \deg(1 - \phi) = \#\ker(1\text{-}\phi) = \#A(F_q) =: N_1$, we get $a_1 = q + 1 - N_1$.

Using the classification of quadratic Weil numbers in Section 1 together with the constraint $ef = 2 \dim A = 2$, we conclude immediately that E_k is either an imaginary quadratic field or the quaternion algebra $B_{p,\infty}/Q$. In the latter case A is certainly supersingular, but the former is inconclusive as π may be unstable. Nevertheless, we have that A is supersingular if and only if its stable endomorphism algebra has center Q, and we conclude that any two supersingular elliptic curves become isogenous after a suitable base extension, so up to geometric isogeny there is a unique supersingular elliptic curve in every characteristic p. Next note that a_1 is playing the role of the totally real β of Section 1. This yields the inequality $|a_1| \leq 2\sqrt{q}$, i.e., the Weil bound for the number of points on an elliptic curve over F_q. Let us now consider an elliptic curve A/F_p and assume p is not 2 or 3 (the "especially nasty" primes for elliptic curves). We claim that in this case A is supersingular if and only if $a_1 = 0$. The sufficiency of $a_1 = 0$ is obvious, so assume that A is supersingular. We have seen that nevertheless the F_p-rational endomorphism algebra will be an imaginary quadratic field K of discriminant $4^{\epsilon}(a_1^2 - 4p), \epsilon \in \{0, 1\}$. We claim that p ramifies in K. Indeed, we see from Section 1 that the inert case requires q to be an *even* power

of p, whereas we have $q = p$. On the other hand, if p splits in K then it is impossible for K to inject into the quaternion algebra $B_{p,\infty}$ (tensor with Q_p to see a nontrivial product injecting into a division algebra). This establishes our claim. So p, which is not 2, divides the discriminant $4^\epsilon(a_1^2 - 4p)$ and hence also a_1. So if $a_1 \neq 0$, $p \leq |a_1| \leq 2\sqrt{p}$ forces $p \leq 4$, contradiction.

Applications to abelian surfaces

Let A/k be an abelian surface. We use the preceding sections to classify A up to isogeny and in particular to compute the possible endomorphism algebras. Let $P(T)$ be the characteristic polynomial of Frobenius; it has degree 4. We consider the various possibilities for its factorization.
Case 1: $P(T)$ is irreducible. Then any root π is a quartic Weil number, giving $f = 4$ and $e = \frac{2 \dim A}{f} = 1$. In this case A is k-simple and E_k is a quartic CM field.
Case 2: $P(T) = P_1(T)P_2(T)$ distinct irreducible quadratics. Then their respective roots π_1, π_2 are quadratic Weil numbers corresponding to non-k-isogenous elliptic curves A_1 and A_2, so $E_k = K_1 \times K_2$ the product of two imaginary quadratic fields. We note a technicality: K_1 and K_2 could be the same field, but not stably so: by the Deuring Lifting Theorem, lift A_1, A_2 to CM elliptic curves in characteristic zero. Then by the classical theory of CM, after a base extension, the lifted curves will become rationally isogenous, and we can reduce the isogeny to get an isogeny from A_1 to A_2 in characteristic p.
Case 3: $P(T) = P_1(T)^2$, with $P_1(T)$ an irreducible quadratic with real roots. This is exactly Case 2 of the analysis of real Weil numbers from Section 1; we conclude that A is k-simple with $E_k = B_{\infty_1,\infty_2}/Q(\sqrt{p})$. This case is inherently unstable.
Case 4: $P(T) = P_1(T)^2$ with $P_1(T)$ an irreducivble quadratic with imaginary roots. Let π be the associated Weil number and put $K = Q(\pi)$ an imaginary quadratic field. Then, either:
4a) π is a Weil number associated to an elliptic curve A_1. As we have seen, this occurs if and only if p is nonsplit in $Q(\pi)$ or p is split and one of $i_{\mathbf{p}1}, i_{\mathbf{p}2}$ is zero and the other is a. Then $A \sim_k A_1^2$, and $E_k = M_2(K)$ Or
4b) π is an exceptional quadratic Weil number associated to A, a k-simple abelian surface, and $E_k = B_{\mathbf{p}_1,\mathbf{p}_2}/K$.
Case 5: $P(T) = P_1(T)^2P_2(T)$ with $P_1(T)$ linear and $P_2(T)$ irreducible quadratic. Then $A \sim_k A_1 \times A_2$ with A_1, A_2 elliptic curves, and $E_k = B_{p,\infty}/Q \times K$ with K an imaginary quadratic field.
Case 6: $P(T) = P_1(T)^4$. Then $A \sim_k A_1^2$, A_1 a necessarily supersingular elliptic curve, and $E_k = M_2(B_{p,\infty})$.
Ruling out the unstable cases, we deduce the following result.

Theorem 132 *Let A/k be an abelian surface. Then $E = End_{\overline{F_p}}(A)$ is one of the following:*

a) L, a quartic CM field.
b) $B_{\boldsymbol{p}_1,\boldsymbol{p}_2}/K$, an 8-dimensional division algebra with center an imaginary quadratic field.
c) $K_1 \times K_2$ a product of distinct imaginary quadratic fields.
d) $M_2(K)$, K an imaginary quadratic field.
e) $B_{p,\infty} \times K$, K an imaginary quadratic field.
f) $M_2(B_{p,\infty})$.
In the first two cases A is simple; otherwise it is isogenous to a product of two elliptic curves.

Bibliography

[Amitsur] S. A. Amitsur, Generic splitting fields of central simple algebras, *Ann. of Math.* (2), **62** 8-43, 1955.

[Baba] S. Baba, Shimura curve quotients with odd Jacobians, *Journal of Number Theory*, Vol. 87, No. 1, 2001.

[Bertolini-Darmon] M. Bertolini and H. Darmon. Heegner points on Mumford-Tate curves, *Inventiones Math.* **126** (1996) 413-456.

[Boutot-Carayol] J.F. Boutot and H. Carayol. Uniformisation p-adique des courbes de Shimura et théorèmes de Cerednik et de Drinfeld, in Courbes Modulaires et Courbes de Shimura, *Asterisque* **196-197** (1991).

[Buzzard] K. Buzzard, Integral models of certain Shimura curves, *Duke Math Journal*, Vol. 87, No. 3, 1997.

[CG] J.-P. Serre, *Cohomologie Galoisienne,* Lecture Notes in Mathematics 5, 5th revised edition, Springer-Verlag 1994.

[CL] J.-P. Serre, *Corps Locaux,* Hermann, Paris, 1962.

[Clark] P. L. Clark, Period-Index problems in Galois cohomology, submitted for publication.

[Clark-Elkies] P.L. Clark and N.D. Elkies. Pointless curves over finite fields, in preparation.

[Deligne-Rapoport] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques,* in: A. Dold and B. Eckmann (eds.) Modular Functions of One Variable II, Lecture Notes in Math. 349, Springer-Verlag, New York, (1973), 143-316.

[Diamond-Taylor] F. Diamond and R. Taylor, Non-optimal levels of mod ℓ modular representations, *Inventiones Math.* **115** (1994), 435-462.

[Ellenberg] J. Ellenberg, Serre's conjecture over F_9, preprint.

[Grothendieck] A. Grothendieck, Le groupe de Brauer I,II, III in *Dix exposes sur la cohomologie des schémas*, Amsterdam: North-Holland, 1968.

[Jordan I] B.W. Jordan, On the Diophantine arithmetic of Shimura curves, Harvard PhD. thesis, 1981.

[Jordan II] B.W. Jordan, Points on Shimura curves rational over number fields, *J. Reine Angew. Math.* **371** (1986), 92-114.

[Jordan-Livné I] B.W. Jordan and R. Livné, Local diophantine properties of Shimura curves, *Math. Ann.* **270** (1985), 235-248.

[Jordan-Livné II] B.W. Jordan and R. Livné, Divisor classes on Shimura curves rational over local fields, *J. Reine Angew. Math.* **378** (1985), 46-52.

[Jordan-Livné III] B.W. Jordan and R. Livné, On Atkin-Lehner quotients of Shimura curves, *Bull. London Math. Soc.* **31** (1999), 681-685.

[JLV] B.W. Jordan, R. Livné and Y. Varshavsky, Local points on p-adically uniformized Shimura varieties, to appear.

[Katok] S. Katok, *Fuchsian groups,* Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992.

[Katz-Mazur] N. Katz and B. Mazur, *Arithmetic Moduli of Elliptic Curves,* Annals of Mathematics Studies 108, Princeton University Press, 1985.

[Knutson] D. Knutson, *Algebraic spaces,* LNM 203, Springer Verlag, 1971.

[Kurihara] A. Kurihara, On some exapmles of equations defining Shimura curves and the Mumford Uniformization, *J. Fac. Soc. Univ. Tokyo, Sec. 1A* **25** (1979), 277-301.

[Ling-Oesterlé] S. Ling and J. Oesterlé, The Shimura subgroup of $J_0(N)$, *Astérisque 196-197*, (1991).

[Mazur] B. Mazur, Modular elliptic curves and the Eisenstein ideal, *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, No. 47 (1978), 33-186.

[Merel] L. Merel, Bornes pour la torsion des courbes elliptiques sur les corps de nombres, *Inventiones Math.* **124** (1996), 437-449.

[Michon] J.F. Michon, Courbes de Shimura hyperelliptiques, *Bull. Soc. math. France* **109** (1981), 217-225.

[Milne] J. Milne, Points on Shimura Varieties mod p, Proc. Symp. Pure Math. 33 (1979), part 2, 165-184.

[Mumford I] D. Mumford, An analytic construction of degenerating curves over complete local rings, *Comp. Math.* **24** (1972), 129-174.

[Mumford II] D. Mumford, *Abelian varieties,* Oxford University Press, 1970.

[Ogg I] A. Ogg, Real Points on Shimura Curves, *Birkhauser PM* **35** (1983) 277-307.

[Ogg II] A. Ogg, Mauvaise réduction des courbes de Shimura, *Birkhauser PM* **59** (1983-84), 199-217. Bad reduction of Shimura curves.

[Ohta] M. Ohta, On ℓ-adic representations of Galois groups obtain from certain two-dimensional abelian varieties, *J. Fac. Soc. Univ. Tokyo Sect. 1A Math.* **21** (1974), 299-308.

[Pierce] R. Pierce, *Associative algebras,* Springer-Verlag, 1982.

[Poonen] B. Poonen, Computational aspects of curves of genus at least 2, available at `http://math.berkeley.edu/ poonen/papers/ants2.pdf`

[Poonen-Stoll] B. Poonen and M. Stoll, The Cassels-Tate pairing on principally polarized abelian varieties, *Ann. of Math.* 150 (1999), 1109-1149.

[Raynaud] M. Raynaud, Schémas en groupes de types $(p, p, \ldots, p)$, *Bull. Soc. math. France,* **102**, 1974, 241-280.

[Ribet] K. Ribet, On modular representations of $\mathrm{Gal}(\overline{Q}/Q)$ arising from modular forms, *Inventiones Math.,* **100** (1990), 431-476.

[RI] B. Mazur, Rational isogenies of prime degree, *Inventiones Math.* **44** (1978), 129-162.

[Rohrlich] D. Rohrlich, Modular curves, Hecke correspondences, and l-functions, in *Modular forms and Fermat's Last thoerem*, 41-100, Springer, 1997.

[Rotger I] V. Rotger, On the group of automorphisms of Shimura curves and applications, *Comp. Math.* **131** (2002) 1-13.

[Rotger II-IV] V. Rotger, preprints:
Quaternions, polarizations and class numbers.
Modular Shimura Varieties and Forgetful maps.
On the field of moduli of quaternionic multiplication on abelian varieties.

[Runge] B. Runge, Endomorphism rings of abelian surfaces and projective models of their moduli spaces, *Tohoku Math. J.* **51** (1999) 283-303.

[Sadykov] M. Sadykov, Parity of genera of Shimura curves over a real quaratic field, available at `http://www.math.uiuc.edu/Algebraic-Number-Theory/0336/`.

[Serre-Tate] J.-P. Serre and J. Tate, Good reduction of abelian varieties, *Ann. of Math.* (2) **88** (1968), 492-517.

[Schmecta] T. Schmecta, Mumford-Tate curves, *Birkhauser PM* **187** (2000), 111-119.

[Silverman] J.H. Silverman, *Advanced topics in the arithmetic of elliptic curves,* Springer, New York, 1994.

[VFL] *Variétiés de Shimura et fonctions L*, Publications Mathématiques de l'Univiersité Paris 7 (1979), 73-81.

[Vignéras] M.-F. Vignéras, *Arithmétique des algèbres de quaternions,* Lecture Notes in Mathematics, 800, Springer, 1980.

[Waterhouse] W. Waterhouse, Abelian varieties over finite fields, *Ann. Sci. Ecole Norm. Sup.* (4) **2** (1969), 521-560.