LECTURE ON SHIMURA CURVES 6: SPECIAL POINTS AND CANONICAL MODELS

PETE L. CLARK

1. Introduction

As mentioned several times in class, the arithmetic of quaternionic Shimura varieties is strongly controlled by the behavior of the class of CM points (just as in the case of modular curves). Just as for elliptic curves, if $\mathfrak{o}$ is an order in an imaginary quadratic field K, one has a notion of a QM-surface with $\mathfrak{o}$-CM. Namely, let $\iota : \mathcal{O} \hookrightarrow A$ be a QM structure on an abelian surface (we are still working over $\mathbb{C}$). We then have the notion of the QM-endomorphism ring and endomorphism algebra of A: $\mathrm{End}_{\mathrm{QM}}(A)$ is equal to the set of $\mathcal{O}$-equivariant endomorphisms of A: i.e., the set of endomorphisms α of A such that for all $x \in \mathcal{O}$, $\iota(x)\alpha = \alpha\iota(x)$. In other words, $\mathrm{End}_{\mathrm{QM}}(A)$ is the centralizer of $\mathcal{O}$ in $\mathrm{End}(A)$. By our classification of endomorphism algebras of abelian surfaces, there are essentially two possibilities: either $\mathbb{Z}$ or an order in an imaginary quadratic field.

Conversely, if $\mathfrak{o}$ is an order in an imaginary quadratic field, we can ask for the set of all QM surfaces A in $S(\mathcal{O})$ whose QM-endomorphism ring is $\mathfrak{o}$. This is called the $\mathfrak{o}$-CM locus. In the elliptic curve case, the $\mathfrak{o}$-CM locus is always in bijection with the Picard group of $\mathfrak{o}$, and moreover the locus forms a single orbit under the Galois group of K, with Galois group isomorphic to $\mathrm{Pic}(\mathfrak{o})$ (i.e., the ring class field of $\mathfrak{o}$). Moreover the set of CM points is dense on the j-line, for the Zariski topology and even for the $\mathbb{C}$-analytic topology. [1]

Here we want to understand the $\mathfrak{o}$-CM locus on a quaternionic Shimura curve ($F = \mathbb{Q}$).

2. Embeddings of quadratic orders into quaternion orders

Let B/K be a quaternion algebra over a number field. Recall that a quadratic extension L/K can be embedded in B if and only if it is a splitting field for B, i.e., $B \otimes_K L \cong M_2(L)$.

Theorem 1. *(Hasse) The quadratic subfields L/K of B are precisely those for which for every prime v of K for which $B_v := B \otimes_K K_v$ is a division algebra (i.e., v is ramified in B), $L_v := L \otimes_K K_v$ is a field.*

Remark: When $K = \mathbb{Q}$ this says: if B is definite, L is quadratic imaginary, and every prime p dividing the discriminant D of B is nonsplit in L. In particular the set of quadratic splitting fields is infinite, of density (in a natural sense) equal to

[1] The CM points are *not* dense in the p-adic analytic topology: I have had occasion to exploit this fact in my own work.

2^{-N}, where N is the number of ramified places of B.

(Sketch proof: The criterion is obviously necessary. Its sufficiency follows from the functoriality of the exact sequence relating the Brauer group of a global field to the Brauer groups of its completions.)

This gives a necessary condition on embedding an order $\mathfrak{o}$ of an imaginary quadratic field K into a given order $\mathcal{O}$ of an indefinite rational quaternion algebra B of discriminant D: there must exist a field embedding $\iota : K \hookrightarrow B$.

Suppose we have an order $\mathfrak{o}$ of K and an order $\mathcal{O}$ of B and a field embedding ι as above. We say that this embedding *optimally embeds* $\mathfrak{o}$ into $\mathcal{O}$ if $\mathfrak{o} = \iota^{-1}(\mathcal{O} \cap \iota(L))$: in other words, it embeds $\mathfrak{o}$ into $\mathcal{O}$ and does not embed any strictly larger quadratic order.

Our problem for this section is to classify the optimal embeddings of $\mathfrak{o}$ into $\mathcal{O}$. By the Noether-Skolem theorem, any two embeddings of K into B are conjugate by an element of B. Let $N(\mathcal{O})$ be the normalizer of $\mathcal{O}$ in $B^\times/\mathbb{Q}^\times$: then $N(\mathcal{O})$ acts on the set of optimal embeddings. For G a subgroup of $N(\mathcal{O})$, we write $v_G(\mathfrak{o}, \mathcal{O})$ for the number of G-orbits of optimal embeddings, and we will write

$$v(\mathfrak{o}, \mathcal{O}) = v_{\mathcal{O}^\times}(\mathfrak{o}, \mathcal{O}).$$

As we shall see, this number is finite and can be computed by a local and local-global calculation similar to the one performed for class numbers and type numbers.

This time we will just give a summary of the results. Detailed proofs can be found in Vignéras' book.

Definition: Let K be a local field with uniformizer π, and let L/K be an étale quadratic algebra (in other words, L/K is either $K \oplus K$ or a separable field extension). The Eichler symbol $(\frac{\mathfrak{o}}{\pi})$ is defined as follows: if $\mathfrak{o}$ is not the maximal quadratic order or if $L = K \oplus \mathrm{K}$, it is 1. Otherwise, it is -1 if L/K is unramified and 0 if L/K is ramified.

Theorem 2. *(Optimal embedding theorem, split local case) Let K be a local field, L/K an étale quadratic algebra, $\mathfrak{o}$ an order of L and $\mathcal{O} \subset M_2(K)$ an order.*
a) If $\mathcal{O}$ is maximal, $v(\mathfrak{o}, \mathcal{O}) = 1$.
b) If $\mathcal{O}$ is a level π-Eichler order, then $v(\mathfrak{o}, \mathcal{O}) = 1 + (\frac{\mathfrak{o}}{\pi})$. In particular, we can embed $\mathfrak{o}$ in $\mathcal{O}$ unless $\mathfrak{o}$ is maximal and L/K is unramified.

Theorem 3. *(Optimal embedding theorem, nonsplit local case) Let K be a local field, L/K an étale quadratic algbra, B/K the (unique) division quaternion algebra, $\mathcal{O}$ the maximal order. If $\mathfrak{o} = \mathcal{O}_L$ is the maximal order, $v(\mathfrak{o}, \mathcal{O}) = 1 - (\frac{\mathfrak{o}}{\pi})$. If $\mathfrak{o}$ is not maximal, $v(s, \mathcal{O}) = 0$.*

As one might expect, the global formula takes into account the local formulas, together with a contribution from the class number of $\mathfrak{o}$.

Theorem 4. *(Optimal embedding theorem, global case) Let B/K be a ntd quaternion algebra over a number field K, and assume there exists a unique conjugacy* *class of Eichler orders of level $\mathcal{N}$; let $\mathcal{O}$ be one such order. Then*

$$v(\mathfrak{o},\mathcal{O}) = h(\mathfrak{o})\prod_v v(\mathfrak{o}_v,\mathcal{O}_v),$$

where the product extends over all finite places of K. Here $h(\mathfrak{o}) = \#\operatorname{Pic}(\mathfrak{o})$ is the class number of $\mathfrak{o}$.

Corollary 5. *Take $K = \mathbb{Q}$ and let $\mathcal{O}$ be an Eichler order of squarefree level N in a quaternion algebra of discriminant D ($D = 1$ is permitted!). Let $\mathfrak{o}$ be an order in the imaginary quadratic field of discriminant δ. Then*

$$v(\mathfrak{o},\mathcal{O}) = h(\mathfrak{o})\prod_{d \mid D}(1-(\frac{-\delta}{p}))\prod_{q \mid N}(1+(\frac{-\delta}{p})).$$

3. Special points on Shimura curves

We return momentarily to a context not considered since the early part of the course. Namely, let F be a totally real field of degree g over $\mathbb{Q}$, and let B/F be a quaternion algebra of type $(1, g-1)$, i.e., split at exactly one infinite place of F. Let $\mathcal{D}$ be the discriminant of B, i.e., the product of all the finite ramified primes. Let $\mathcal{O}$ be a maximal order of B and let $\Gamma = \Gamma(B,\mathcal{O})$ be the group of units of $\mathcal{O}$ of reduced norm 1 Let $X = X(\mathcal{O}) = \Gamma\backslash\mathcal{H}$ be the corresponding Riemann surface. So X is either the j-line (if $F = \mathbb{Q}$, $B \cong M_2(\mathbb{Q})$) or a compact Riemann surface. Note that this construction does, in general, depend upon the choice of $\mathcal{O}$, however in a controllable way:

Exercise: Let h_B be class number appearing in Eichler's Theorem, i.e., the degree of the ray class extension of F corresponding to the modulus $\infty_2\cdots\infty_g$ (the ramified infinite places of B). Show that as $\mathcal{O}$ ranges over all maximal orders of B, one obtains at most h_B pairwise nonisomorphic $\mathbb{C}$-algebraic curves $X(\mathcal{O})$. Is it obvious that this number of isomorphism classes is exactly h_B?[2]

It will be slightly more convenient for us to work with discrete subgroups of $GL_2(\mathbb{R})^+$; we then get the corresponding Fuchsian group by taking the image in $PSL_2(\mathbb{R}) = PGL_2(\mathbb{R})^+$. One often denotes the former group by $\tilde{\Gamma}$ and the latter by Γ; for simplicity of notation, we will (except in the immediately following discussion) not distinguish between them. This is almost harmless, except that if $\widetilde{\Gamma'} \subset \widetilde{\Gamma} \subset GL_2(\mathbb{R})^+$, the index $[\widetilde{\Gamma} : \widetilde{\Gamma'}]$ will either be equal to $[\Gamma : \Gamma']$ or $2[\Gamma : \Gamma']$ (the latter occurs iff $-1 \in \Gamma \setminus \Gamma'$).

Now let $\mathcal{N}$ be an integral ideal of F. We define a the principal congruence subgroup $\Gamma(\mathcal{N}) \subset \Gamma$ as the set of all $\gamma \in \Gamma$ such that $\gamma - 1 \in \mathcal{N}\mathcal{O}$. Notice that we have *not* assumed that $\mathcal{N}$ and $\mathcal{D}$ are coprime, so that these groups are cofinal in the system

[2]It is not obvious to me, but on the other hand I believe it to be true (at least when there are no elliptic points), so I allow for the possibility that it is obvious to you! But the point of the question is: h_B is the number of $B^\times/\mathbb{Q}^\times$ conjugacy classes of $Gamma(B,\mathcal{O})$, whereas for isomorphism of the corresponding Riemann surfaces we should be looking at $PGL_2(\mathbb{R})^+$-conjugacy classes. Moreover, if Γ has elliptic elements, it is conceivable that the Riemann surfaces are (noncanonically) isomorphic even if the corresponding Fuchsian $\Gamma(\mathcal{O})$, $\Gamma(\mathcal{O}')$ are nonconjugate. A later draft of these notes may contain a more complete discussion.

of all congruence subgroups of Γ.

Exercise: Suppose that $\mathcal{N}$ is prime to the discriminant $\mathcal{D}$ of B.
a) Compute the index $[\Gamma : \Gamma(\mathcal{N})]$ (interpreted in either sense as above).
b) Show that $\Gamma(\mathcal{N})$ is normal in Γ and compute the quotient.

We put $X(\mathcal{N}) = \Gamma(\mathcal{N})\backslash\mathcal{H}$.

Now let K/F be an imaginary quadratic extension (hence a CM field of degree $2g$), and let $\mathfrak{o}$ be an order in K. Consider as before the set of optimal embeddings of $\mathfrak{o}$ into $\mathcal{O}$ modulo $\mathcal{O}^\times$-conjugacy, of cardinality $v(\mathfrak{o}, \mathcal{O})$. Since $\mathcal{O}$ is maximal, $v(\mathfrak{o}, \mathcal{O}) > 0$ if and only if L is a splitting field for B (i.e., if and only if L/K is nonsplit at every prime v dividing the discriminant $\mathcal{D}$ of B).

Now, through the completion at the unique split place ∞_1, we have an embedding $B^\times \hookrightarrow GL_2(\mathbb{R})$ hence a natural action of $B^\times$ on $\mathcal{H}^\pm$. In particular, we have an action of B^+ on $\mathcal{H}$. Under this embedding, an element $x \in F^\times$ maps to the scalar matrix $\mathrm{diag}(\infty_1(x), \infty_1(x))$, hence acts trivially on $\mathcal{H}$.

For $\tau \in \mathcal{H}$, let G_τ denote its stabilizer in $B^\times$ (which is necessarily a subgroup of B^+).

Lemma 6. *Let $\iota : L \rightarrow B$ be an embedding of a CM quadratic extension of K. Then $\iota(L^\times) \subset B^+$. There exists a unique $\tau \in \mathcal{H}$ which is the common fixed point of all elements of $\iota(L^\times)$.*

Exercise: Prove it! (Hint: $\iota(L^\times)$ consists of commuting elliptic elements of $PSL_2(\mathbb{R})$!)

The (unique) point τ corresponding to the embedding $\iota : L \rightarrow B$ is called an L-special point of $\mathcal{H}$. For our fixed $\mathcal{O}$, there is a unique order $\mathfrak{o}$ of L which is optimally embedded into $\mathcal{O}$ via L, namely $\mathfrak{o} = \iota^{-1}(\mathcal{O} \cap \iota(L^\times)) \cup \{0\}$. We say that the point τ is $\mathfrak{o}$-special.

Lemma 7. *A point $\tau \in \mathcal{H}$ is L-special if and only if $G_\tau \cup \{0\}$ is a CM field L inside B. If τ is not special, $G_\tau = F^\times$.*

Exercise: Prove it.

It follows that the set of special points is countably infinite.

Now we look at image of the special points on the Shimura curve X.

Lemma 8. *Let $\tau_1,\ \tau_2$ be two special points on $\mathcal{H}$, and let $\varphi : \mathcal{H} \rightarrow \Gamma\backslash\mathcal{H}$ be the uniformization map. Then $\varphi(\tau_1) = \varphi(\tau_2)$ if and only if they are $\mathfrak{o}$-special for the same $\mathfrak{o}$, and the embeddings $\iota_1,\ \iota_2$ are conjugate by an element of $\mathcal{O}^\times$.*

Exercise: Prove it.

Corollary 9. *For every order $\mathfrak{o}$ in a CM quadratic extension L/K, there exist $v(\mathfrak{o}, \mathcal{O})$ $\mathfrak{o}$-special points on X. This number is positive if and only if L splits B.*

Let us write $\infty = \infty_1 \cdots \infty_g$ viewed as a modulus for F (in the sense of classfield theory in the non-adelic formulation). For an ideal $\mathcal{N}$ of F, let $F(\mathcal{N})$ be the ray-class field corresponding to the modulus $\mathcal{N} \cdot \infty$, i.e., the group of fractional ideals prime to $\mathcal{N}$ modulo ideals whose generators are totally positive and congruent to 1 modulo $\mathcal{N}$. (To be sure, when $F = \mathbb{Q}$ and $\mathcal{N} = N$, $F(\mathcal{N}) = \mathbb{Q}(\zeta_N)$.)

We would be justified in calling the following result the "Fundamental Theorem of Shimura Curves"; certainly it explains the nomenclature.

Theorem 10. *(Shimura) Consider the Riemann surface* $\Gamma(\mathcal{N})\backslash\mathcal{H}$.
a) There exists a smooth projective curve V defined over $F(\mathcal{N})$ and an isomorphism $\varphi : \Gamma(\mathcal{N})\backslash\overline{\mathcal{H}} \to V(\mathbb{C})$.
b) Let $\tau \in \mathcal{H}$ be an $\mathfrak{o}$-special point, where $\mathfrak{o}$ is the maximal order in the CM quadratic extension L/F. Then the compositum of the field of moduli $F(\varphi(\tau))$ and L is the $\mathcal{N}$-ray class field of L, i.e., the group of fractional $\mathfrak{o}$-ideals prime to $\mathcal{N}$ modulo principal ideals whose generator is congruent to 1 *modulo $\mathcal{N}\mathfrak{o}$.*
c) The curve $V_{/F(\mathcal{N})}$ is uniquely determined by the conditions of a) and b).

Remark: More generally, let $\mathfrak{o}$ be any order of L whose conductor is prime to $\mathcal{N}$, and let τ be an $\mathfrak{o}$-special point. Then the field $L(\varphi(\tau))$ is (I believe) the $\mathcal{N}$-ring class field of L, whose definition is identical to the one given in part b) for the maximal order.

The following discussion is preparation for the explicit reciprocity law. In the following discussion we shall assume that the narrow class number of F is equal to 1. Let $\mathfrak{o}$ be a maximal order in a CM extension L/F. Put $G = \mathrm{Gal}(L^1/L) = \mathrm{Pic}(\mathfrak{o})$ be the Galois group of the Hilbert class field of L. We will give an explicit action of G on the set $v(\mathfrak{o}, \mathcal{O})$ of $\mathcal{O}^\times$-conjugacy classes of optimal embeddings $\iota : \mathfrak{o} \hookrightarrow \mathcal{O}$. Namely, for $\sigma \in G$, let $a \in L^\times(A)$ be an idele corresponding to σ under the Artin map, i.e., such that $(a, L^1/L) = \sigma$. As discussed before, applying a to $\mathcal{O}$ coordinatewise gives us a left $\mathcal{O}$-ideal I. But since the class number is 1, this ideal is principal, so of the form $\alpha_\sigma\mathcal{O}$ for some $\alpha_\sigma \in B^\times$ of positive norm. (In fact, because σ has finite order, necessarily $\alpha \in \mathcal{O}$.)

Theorem 11. *(Shimura reciprocity law) The element $\sigma \in G =$ $\mathrm{Pic}(\mathfrak{o})$ *sends the optimal embedding $\iota : \mathfrak{o} \to \mathcal{O}$ with corresponding special point z to the optimal embedding $\iota^\alpha = \alpha^{-1} \circ \iota \circ \alpha$, with corresponding special point $\alpha^{-1}z$.*

The preceding description is *'a la* Shimura. Deligne's (essentially equivalent) adelic formulation is as follows: let $G = B^\times$ viewed as an algebraic group over $\mathbb{Q}$. Let U_f be a compact open subgroup of $G(A_f)$. A cofinal system of such groups is given by the $U(\mathcal{N})$, which at every place v prime to $\mathcal{N}$ is $\mathcal{O}_v^\times$, and for dividing $\mathcal{N}$ to power exactly n_v, we take the units u in $\mathcal{O}_v$ such that $u - 1 \in \pi_v^{n_v}\mathcal{O}_v$ (where π_v is a uniformizer of F_v). Then, as we discussed, the double coset space

$$\mathrm{Sh}(B^\times, U_f) = B^\times\backslash\mathcal{H}^\pm \cdot B^\times(A_f)/U_f$$

is isomorphic to a finite disjoint union of Riemann surfaces each isomorphic as schemes to $\Gamma(U_f)\backslash\mathcal{H}$, where $\Gamma(U_f) = B \cap U_f$ and the components are parameterized by the class group $G = \mathrm{Gal}(F(\mathcal{N} \cdot \infty)/F)$. Then we can say that $\mathrm{Sh}(B^\times, U_f)$ has a canonical model over F itself such that the Galois action on the components is given by G.

In particular, this means that each component gets a canonical model over the field trivializing this action, i.e., over $F(\mathcal{N} \cdot \infty)$. The various components need *not* be isomorphic to each other as complex algebraic curves. Each component $V_{i/F(\mathcal{N}\cdot\infty)}$ is obtained from a fixed component $V_{1/F(\mathcal{N}\cdot\infty)}$ by "extending scalars" by $\sigma_i \in G$: $V_i = V_1 \otimes_{\sigma_i} \mathbb{C}$.

Remark: This theorem is not the last word on the arithmetic of the component curves V_i. Namely, the field $F(\mathcal{N}\cdot\infty)$ may, or may not, be a minimal field of definition for V_1. For instance, when $F = \mathbb{Q}$, $\mathcal{N} = p$, the theorem gives a canonical model for each of the $(\mathbb{Z}/p\mathbb{Z})^\times$ components of the modular curve over $\mathbb{Q}(\zeta_p)$. However, it can be shown that the Riemann surface $X(p)$ can always be defined over $\mathbb{Q}$, and there are models which are (in a different sense!) canonical over $\mathbb{Q}(\sqrt{p^*})$. Indeed, the same holds for $X^D(p)$ (in fact, whenever $F = \mathbb{Q}$), but not for a general totally real field. As far as I know, the problem of computing the field of moduli of the curve $X^{\mathcal{D}}$ and its congruence coverings remains open in the general case.

We are not going to give a complete proof of the fundamental theorem, but we shall make some comments. First, we separate out the cases $F = \mathbb{Q}$ (easy case) and $F \neq \mathbb{Q}$ ((extremely) hard case).

In the easy case, we are much aided by the fact that $X(\mathcal{N})$ is a coarse moduli space (even a fine moduli space, for sufficiently large $\mathcal{N}$) for the moduli problem of QM-abelian surfaces with level $\mathcal{N}$-structure. In particular, since this is a moduli problem which can be formulated in the category of $\mathbb{Q}$-schemes (since $\mathbb{Q}$ has characteristic zero, there are no subtleties here, and we do not delve into the technical meaning of these terms), the uniqueness of the canonical model, if it exists, comes for free (or from the Yoneda Lemma). Moreover, nowadays there are similarly quite general techniques for showing the existence of moduli spaces (and in fact, we will soon enough be stating the existence of *integral* canonical models for Shimura curves without any justification), but let us say a little bit about Shimura's approach, which is valid for a much larger class of moduli varieties (namely for PEL-type Shimura varieties). Essentially, the problem reduces to finding a caonical $\mathbb{Q}$-model for Siegel moduli spaces, or, to having a good understanding of the field of moduli of a polarized abelian variety with level structure. Let (A, P) be a polarized abelian variety. Then (by a famous theorem of Lefschetz) $3P$ embeds A into a projective space, so gives a point on some Chow variety. (Shimura interprets the additional structure coming from the endomorphisms and the level structure as giving a variety in some larger projective space, hence a point on some Chow variety.) The assignment of a "Chow point" to a PEL-moduli space $X_{/\mathbb{C}}$ gives a morphism Φ from a PEL-type moduli space to a Chow variety V (which does come with a canonical $\mathbb{Q}$-rational model), in such a way that the field of moduli of a point $P \in X(\mathbb{C})$ is the field $\mathbb{Q}(\varphi(P))$ generated by the coordinates of the Chow point. If $\mathbb{Q}(V) = \mathbb{Q}(g_i)$ is the function field of V, then putting $f_i = \Phi^*(g_i) = g_i \circ \Phi$, we get a set of functions f_i on X such that the field of moduli of a point $P \in X(\mathbb{C})$ is $\mathbb{Q}(f_i(P))$. Thus we take $K = \mathbb{Q}(f_i)$ to be the canonical (birational) $\mathbb{Q}$-model for X. One still must check that K is a regular extension of $\mathbb{Q}$ (i.e., that $\mathbb{Q}$ is algebraically closed in K); Shimura does this, and in so doing constructs a specific $\mathbb{Q}$-rational model for a PEL-type moduli variety.

Obviously no such general argument will establish the explicit reciprocity law.

This is a generalization of the main results of the theory of complex multiplication for elliptic curves to abelian surfaces which are isogenous to the square of a CM elliptic curve. Thus, to do justice to this argument we would have to carefully revisit the theory of complex multiplication, which we shall not do here. (We remark however that there is a similar theory for abelian varieties with not-necessarily-isotypic complex multiplication, developed by Shimura and Taniyama. This more general theory is a prerequisite to the theory of special points on more general Shimura varieties.)

In summary, if you have an excellent grasp of the theory of complex multiplication and moduli of polarized abelian varieties, proving the fundamental theorem for $F = \mathbb{Q}$ is not wildly difficult. Since, indeed, it was Shimura (partially in collaboration with Taniyama) who developed both of these theories, it was rather easy for him to prove Theorem 10 when $F = \mathbb{Q}$. In contrast, the proof of the theorem for $F \neq \mathbb{Q}$ took Shimura many more years. The barest outline of the proof is as follows: for each CM field L with $v(\mathfrak{o}_L, \mathcal{O}) > 0$, one constructs a canonical model over L for a different Shimura variety $W(L)$: it corresponds to choosing a reductive group (of unitary type) whose associated semisimple subgroup is $SL_1(B)$ but is itself not $B^\times$. The variety $W(L)$ is itself of PEL-type so has a canonical model by a proof along the lines of above. In particular, the reciprocity law holds for the action of the Hilbert class group of L on a set of $\mathfrak{o}_L$-CM points of $W(L)$. In this way, $X(\mathcal{N})$ gets a canonical model over $L(\mathcal{N})$, the $\mathcal{N}$-ray class field of L (because the identity component of a Shimura variety depends only on the associated semisimple group, which is the same for V as for $W(L)$). This is not good enough: we want a canonical model over the $\mathcal{N}$-ray class field of F. What Shimura shows is that the infinitely many different canonical models over the various $L(\mathcal{N})$'s glue together to give a unique model over $F(\mathcal{N})$. In the proof of this, the explicit reciprocity law at the special points plays a critical role, giving us, roughly speaking, glueing data. It is not an easy argument: indeed, Deligne found the argument perplexing enough (he called these canonical models "modèles étranges") to warrant the development of a different approach, exploiting more systematically the "functoriality" between Shimura varieties (i.e., each Shimura variety is defined in terms of a (suitable) reductive group $G_{\mathbb{Q}}$; a homomorphism $G \to G'$ of reductive groups leads to a morphism of Shimura varieties). Deligne's work extends and simplifies Shimura's work, but it is still too difficult for us to discuss here.

4. A bit of Atkin-Lehner theory

In the case $F = \mathbb{Q}$, one can also give an explicit action of the Atkin-Lehner group $N\mathcal{O}/\mathcal{O}^\times \cong \prod_{p \mid D} \mathbb{Z}/2\mathbb{Z}$ of X^D (this is the group of modular automorphisms of X^D; it plays an extremely important role in the arithmetic of the Shimura curves and the corresponding abelian surfaces, but for lack of time we have not yet found an opportunity to discuss it formally). It can be described as follows: for every positive integer $d \mid D$, there exists an element $w_d \in \mathcal{O}$ of reduced norm d. Conjugation by w_d normalizes $\mathcal{O}$, and w_d^2 centralizes $\mathcal{O}^\times$, so we get an involutory automorphism of $\Gamma^D(1)\backslash\mathcal{H}$. When $d = D$, we can explicitly take w_D to be an element whose square is $-D$, since the quadratic order $\mathbb{Z}[\sqrt{-D}]$ can always be embedded in B.
Since W is trivial iff $D = 1$, its action is exactly what distinguishes the nonsplit case from the classical case, and explains the fact that the $\mathfrak{o}$-CM points need not form a single orbit. Indeed, fix an imaginary quadratic field K which splits B, and

let W' be the subgroup of W generated by the w_p's for primes p which are inert in K. (Since K splits B, the only other possibility is that p ramifies in K.) Then it can be shown that $W' \times \mathrm{Pic}(\mathfrak{o}_K)$ acts *simply transitively* on $v(\mathfrak{o}, \mathcal{O})$.

Exercise: Fix a $d \mid D$. Since w_d has positive reduced norm, it acts on $\mathcal{H}$.
a) Show that the abelian surfaces A_τ and $A_{w_d \tau}$ are isomorphic as $\mathbb{C}$-tori. (Hint: w_d can be viewed as an element of $GL_2(\mathbb{C})$ which carries one lattice to the other.)
b) Nevertheless, if $d \neq 1$, w_d does not act trivially on X^D: why not? (Hint: w_d acts also on the quaternionic structure $\iota : \mathcal{O} \hookrightarrow A_\tau$ via its conjugation action on $\mathcal{O}$.)

It follows that a given abelian surface A admitting an $\mathcal{O}$-QM structure, will in general admit several nonisomorphic QM structures corresponding to its conjugates under the action of the Atkin-Lehner group. Recall that the QM structure determined a principal polarization on A, so that by conjugating the QM structure, the associated polarization may, or may not, change. It turns out that, except for a finite set of points, the subgroup H of W preserving the principal polarization on $A_\tau \in X^D$ is independent of τ. Important work of V. Rotger describes this subgroup H. It turns out that $\#H = 2$ or 4: $w_D \in H$ always, and whether or not there is an additional element depends on a rather subtle invariant of $\mathcal{O}$ together with the element μ such that $\mu^2 = -D$ that we chose long ago in order to get a positive involution on B. We do not state the precise condition here. However, we can already derive many important consequences:

Define $X^{D+} = X^D/w_D$, and $X^D_H = X^D/H$. There is a natural morphism $X^D \to \mathcal{A}_{2,1}$ from the Shimura curve into the two-dimensional Siegel modular variety obtained by taking a triple (A, ι, P) and forgetting the quaternionic struture: $(A, \iota, P) \mapsto (A, P)$.

Theorem 12. *(Rotger) The forgetful map $X^D \to \mathcal{A}_{2,1}$ factors through to an embedding $X^D/H \to \mathcal{A}_{2,1}$.*

Since $\#H \geq 2$, Shimura curves are *never* canonically embedded inside the Siegel moduli space. (In fact the same holds for Hilbert modular varieties: the canonical forgetful map factors through a finite group of modular automorphisms.)

Because $\#H \leq 4$ and $\#W$ can be arbitrarily large, we immediately get abelian surfaces with many distinct principal polarizations. However, this is not the right quantity to count: in general, there can be infinitely many distinct principal polarizations on an abelian variety. But this can only occur if A has an infinite automorphism group (as our QM surfaces do). More precisely, say that two polarizations P, P' on an abelian variety are **conjugate** if there exists an automorphism α of P such that $\alpha^* P' = P$.[3] An important theorem of Narasimhan-Nori says that the number of conjugacy classes of principal polarizations (or indeed polarizations of any given type $D = (d_1, \ldots, d_g)$) on an abelian variety A is finite. To say more is a very interesting problem.

Coming back to our situation, our first thought may be that the W-orbit of a

[3]It is more standard to call such polarizations "isomorphic", but this terminology is unappealing to me.

point $\tau \in X^D$ consists of a single conjugacy class of polarizations, but this is not necessarily the case: the automorphisms w_d lie in the normalizer of $\mathcal{O}^\times$, not $\mathcal{O}^\times$ itself. Moreover, there may be other polarizations on A_τ. The precise answer is as follows:

Theorem 13. *(Rotger) Let $A_{/\mathbb{C}}$ be an abelian surface with* $\mathrm{End}(A) \cong \mathcal{O}$*, a maximal order in a quaternion algebra of discriminant D. Then the number of conjugacy classes of principal polarizations of A is $\frac{h'(-D)+h'(-4D)}{2}$, where by $h'(b)$ denotes the class number of the quadratic order of discriminant b or 0 if there is no such order.*

It can be shown that if K is a field of characteristic 0 and $(A,P)_{/K}$ is a principally polarized abelian surface such that $\mathrm{End}_{\mathbb{C}}(A) \cong \mathcal{O}$ is a maximal order in an indefinite rational quaternion algebra B, then there exists an $\mathcal{O}$-QM structure on A inducing the polarization P. In other words, every principal polarization on a QM surface is compatible with some QM structure. Because of this, we can make the following definition. Definition: A **potentially quaternionic** abelian surface $A_{/K}$ (K a field of characteristic 0) is a principally polarized abelian surface (A,P) together with an embedding $\mathcal{O} \hookrightarrow \mathrm{End}_{\mathbb{C}}(A)$.

Every potentially quaternionic abelian surface A/K induces a point $x \in X_H^D(K)$. In particular, every PQM abelian surface acquires its QM over an abelian extension of K isomorphic to $\mathbb{Z}/2\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Warning: Not every K-rational point on $X_H^D(K)$ is induced by a PQM abelian surface. There is an obstruction, in general nontrivial, for such an abelian surface to be defined over its field of moduli. However, one could show that there do not exist any $\mathcal{O}$-PQM abelian surfaces $A_{/K}$ by showing that $X_H^D(K) = \emptyset$.

Rotger and Dieulefait have also showed that if $A_{/\mathbb{Q}}$ is a PQM abelian surface of GL_2-type, then $\#H = 4$ and the QM is defined over an imaginary quadratic extension $K = \mathbb{Q}(\sqrt{-m})/\mathbb{Q}$. Assuming Serre's conjecture (again, this seems like a good idea), A corresponds to a modular form f with Fourier field $F = \mathbb{Q}(\sqrt{d})\,\mathrm{End}^0_{\mathbb{Q}}(A)$; taking σ to be the nontrivial automorphism of F, there must exist an extra twist: $L(f^\sigma) = L(\chi)L(f)$, where χ is a quadratic Dirichlet character cutting out the field extension K, and B is the quaternion algebra $(d,-m)$.

In particular, it is very interesting to study the locus $X^{D+}(\mathbb{Q})$, because if it is empty, then there are no modular $\mathcal{O}$-PQM abelian surfaces $A/\mathbb{Q}$. It is natural to conjecture that the latter holds for all but finitely many D.

However, determing the set $X^{D+}(\mathbb{Q})$ is very difficult:

Proposition 14. *Whenever there exists a class number one CM field K splitting D, there exists some $\mathfrak{o}_K$-CM point on X^D which becomes $\mathbb{Q}$-rational on X^{D+}.*

Conversely, if we choose D so that none of the (finitely many!) CM quadratic fields of class number at most 2 split B, then Shimura's reciprocity law immediately implies that there are no $\mathbb{Q}$-rational CM points on X^{D+}. It is reasonable to conjecture that, except possibly for finitely many counterexamples, the only $\mathbb{Q}$-rational points on X^{D+} are the CM points. However, this is a very difficult problem. For instance:

Theorem 15. *(Clark) For all* $p \leq \infty$, $X^{D+}(\mathbb{Q}_p) \neq \emptyset$.