1995a

SHIMURA VARIETIES: THE GEOMETRIC SIDE OF THE ZETA FUNCTION

J.S. MILNE

March 6, 1995

ABSTRACT. These are my notes for four talks at the Institute for Advanced Study, February 21,23,28, and March 2, 1995. Appearances to the contrary, they are *rough* notes.

After giving a brief introduction to Shimura varieties, and in particular explaining how to realize them as moduli varieties, we give a heuristic derivation of a formula for the number of points on the reduction of a Shimura variety with coordinates in a finite field.

CONTENTS

Introduction

Shimura varieties generalize elliptic modular curves. They have canonical models over number fields, and the study of their Hasse-Weil zeta functions has guided much research in the theory of automorphic representations. For example, the problem of L-indistinguishability, now called endoscopy, first manifested itself in the study of Shimura varieties.

For a smooth projective variety Y over a number field E, the zeta function is defined as follows: for all but finitely many prime ideals $\mathfrak{p}_v$, reducing the equations for Y modulo $\mathfrak{p}_v$ will give a smooth projective variety over the residue field $k(v)$, and one defines

$$Z_v(Y,T) = \exp\left(\sum_{m=1}^{\infty} \frac{N_m T^m}{m}\right), \quad N_m = \#Y(k(v)_m), \quad [k(v)_m : k(v)] = m,$$

and

$$\zeta(Y,s) = \prod_{v \text{ good}} Z_v(Y, q_v^{-s}) \times \prod_{v \text{ bad}} \cdots, \quad q_v = \#k(v).$$

One should also add Γ-factors for the infinite primes.

By "understanding the zeta function" we mean understanding some regularity in the N_m's, but the regularity may be quite complicated. For example, a curve of genus 0 over $\mathbb{Q}$ with a rational point is just $\mathbb{P}^1$, and

$$\zeta(\mathbb{P}^1, s) = \zeta(s)\zeta(s-1).$$

A curve of genus 1 over $\mathbb{Q}$ with a rational point is an elliptic curve A, and

$$\zeta(A,s) = \frac{\zeta(s)\zeta(s-1)}{L(s)}$$

where $L(s)$ is the Mellin transform of a modular form of weight 2, at least if the curve has no worse than multipicative reduction at 3 and 5. For a Shimura variety, we hope that the zeta function is expressible in terms of the L-series of automorphic representations.

One can also define the zeta function in terms of the étale cohomology of Y, namely,

$$Z_v(Y,T) = \frac{P_1(T)P_3(T)\cdots}{P_0(T)P_2(T)\cdots P_{2\dim Y}(T)}, \quad P_i(T) = \det(1 - \mathrm{Frob}_v \cdot T | H^i(Y_{k(v)^{\mathrm{al}}}, \mathbb{Q}_\ell))$$

When the variety is not complete, for example, a Zariski open subset of a projective variety, one can define the zeta function in terms of the intersection cohomology (with middle perversity) of a good compactification of the variety—in the case of a Shimura variety, one takes the Baily-Borel compactification. The resulting function depends on the compactified variety, but there is a part of it that can be regarded as the contribution of the variety itself, not its boundary, and which can be defined in terms of the numbers N_m as above.

At the conference in Ann Arbor in 1988, Kottwitz began by writing down a conjectural formula for the number of points on a Shimura variety over a finite field, and then gave a heuristic stabilization of the formula (Kottwitz 1990). In these four expository lectures, I shall review the definition and basic properties of Shimura varieties, especially their interpretation as moduli varieties, and then I shall give a heuristic derivation of the conjecture of Langlands and Rapoport (1987) on the structure of the points of a Shimura variety modulo a prime; finally, I shall briefly indicate how one derives the formula in Kottwitz's talk from the conjecture of Langlands and Rapoport.

Thus, I'll ignore the contributions to the zeta function of the boundary, the bad primes, and the infinite primes.

1. Locally Symmetric Varieties

1.1. Symmetric Hermitian domains. A *bounded symmetric domain* X is a bounded open connected subset of $\mathbb{C}^m$, for some m, that is symmetric in the sense that, for each point $x \in X$ there is an automorphism s_x of X of order 2 having x as an isolated fixed point. The simplest examples of bounded symmetric domains are the unit balls:

$$B_m = \{x \in \mathbb{C}^m \mid |x| < 1\}.$$

A complex manifold isomorphic to a bounded symmetric domain is called a *symmetric Hermitian domain.* The simplest example of a symmetric Hermitian domain is the complex upper-half-plane $H^+ = \{z \in \mathbb{C} \mid \Im(z) > 0\}$, which is isomorphic to B_1 by the map

$$z \mapsto \frac{z-i}{z+i} : H^+ \to B_1.$$

The symmetric Hermitian domains were classified by E. Cartan and Harish-Chandra using the theory of semisimple groups. Let X^+ be a symmetric Hermitian domain. The group $A = \mathrm{Aut}(X^+)$ of automorphisms of X^+ (as a complex manifold) is a real semisimple Lie group with trivial centre; moreover, the identity component[1] A^+ of A acts transitively on X^+, and the stabilizer of any point is a maximal compact subgroup of A^+. Thus every symmetric Hermitian domain can be realized as a quotient

$$G(\mathbb{R})^+/K$$

where G is a real semisimple algebraic group with trivial centre. However, if G is a real semisimple algebraic group with maximal compact subgroup K, then $G(\mathbb{R})^+/K$ will not usually be a symmetric Hermitian domain.

Let H be a simple adjoint group over $\mathbb{C}$, and choose a maximal torus T and a simple set of roots B. Recall that there is a unique (highest) root $\tilde{\alpha} = \sum_{\alpha \in B} n(\alpha)\alpha$ such that, for any other root $\sum m(\alpha)\alpha$, $n(\alpha) \geq m(\alpha)$. A node of the Dynkin diagram s_α is *special* if $n(\alpha) = 1$.

Theorem 1.1. *The symmetric Hermitian domains X^+ such that $\mathrm{Aut}(X^+)^+$ is the identity component of a real form of H are in one-to-one correspondence with the special nodes of the Dynkin diagram of H.*

Given a special node s_{α_0}, there exists a unique $\mu \in X_*(T)$ such that $\alpha_0 \circ \mu = 1$ and $\alpha \circ \mu = 0$ for the other simple roots α. Then G is the real form of H corresponding to the Cartan involution $\mathrm{ad}\mu(-1)$.

An examination of the tables in (Helgason, 1978, pp477-478) reveals that: every node of the Dynkin diagram of type A_n is special; the Dynkin diagrams of type B_n, C_n, and E_7

[1] In general, a + denotes a connected component for the real topology.

each have a single special node; the Dynkin diagram of type D_n has three special nodes; the Dynkin diagram of type E_6 has two special nodes; the Dynkin diagrams of type E_8, F_4, and G_2 have no special nodes.

1.2. Locally symmetric varieties. Let X^+ be a symmetric Hermitian domain, and let G be a semisimple algebraic group over $\mathbb{Q}$ such that $X^+ = G(\mathbb{R})^+/K$ with K a maximal compact subgroup of $G(\mathbb{R})^+$. Recall that a subgroup of $G(\mathbb{Q})$ is said to be *arithmetic* if it is commensurable with $G(\mathbb{Q}) \cap \mathrm{GL}_n(\mathbb{Z})$ for one (hence all) embeddings $G \hookrightarrow \mathrm{GL}_n$. A sufficiently small arithmetic subgroup Γ will be torsion-free—we always assume this is so. Then $S = \Gamma\backslash X^+$ will again be a complex manifold, with X^+ as universal covering space and the image of Γ in $\mathrm{Aut}(X^+)$ as its fundamental group.

Theorem 1.2. *The complex manifold S has a canonical structure of an algebraic variety. With this structure, every holomorphic map $V^{\mathrm{an}} \to S$ from a complex algebraic variety V (viewed as an analytic space) to S is a morphism of algebraic varieties.*

The first statement is the theorem of Baily and Borel (1966), and the second is proved in (Borel 1972, 3.10).

The varieties arising as in the theorem are called *locally symmetric varieties.*

In fact, Baily and Borel define a canonical map $S \to \bar{S}$ realizing S as an open subvariety of projective algebraic variety $\bar{S}$, called the Baily-Borel compactification. If S has no factors of dimension 1, $\bar{S}$ can be described as follows: let Ω^1 be the sheaf of holomorphic differential forms, and let $\omega = \bigwedge^{\dim S} \Omega^1$; define $A = \oplus_{n \geq 0} \Gamma(S, \omega^{\otimes n})$; it is a finitely generated graded $\mathbb{C}$-algebra, and so defines a projective algebraic variety $\bar{S} = \mathrm{Proj} A$; there is a canonical map $S \to \bar{S} = \mathrm{Proj} A$.

The compactification $S \hookrightarrow \bar{S}$ is minimal in the sense that for any nonsingular algebraic variety S' containing S as an open subvariety and such that $S' - S$ has only normal crossings as singularities, there is a unique morphism $S' \to \bar{S}$ whose restriction to S is the identity map.

Let G be an algebraic group over $\mathbb{Q}$, and let $G(\mathbb{Z}) = G(\mathbb{Q}) \cap \mathrm{GL}_n(\mathbb{Z})$ for some embedding $G \hookrightarrow \mathrm{GL}_n$. A subgroup of $G(\mathbb{Q})$ is a *congruence subgroup* if it contains

$$\Gamma(N) =_{df} \mathrm{Ker}(G(\mathbb{Z}) \to G(\mathbb{Z}/N\mathbb{Z}))$$

for some $N \geq 1$. More canonically, let $\mathbb{A}_f$ be the ring of finite adéles, i.e., $\mathbb{A}_f = \widehat{\mathbb{Z}} \otimes \mathbb{Q}$ where $\widehat{\mathbb{Z}} = \varprojlim_N \mathbb{Z}/N\mathbb{Z}$; then the congruence subgroups of $G(\mathbb{Q})$ are the subgroups of the form $G(\mathbb{Q}) \cap K$ with K a compact open subgroup in $G(\mathbb{A}_f)$.

Notes: For the material in Section 1.1, see (Helgason 1978, Chapter VIII).

2. Shimura Varieties

As everyone knows, the modular curve $\Gamma(1)\backslash H^+$ parametrizes isomorphism classes of elliptic curves over $\mathbb{C}$, but what parametrizes isomorphism classes of elliptic curves with level N structure, i.e., pairs $(A, (\mathbb{Z}/N\mathbb{Z})^2 \xrightarrow{\approx} A[N])$? One might guess $\Gamma(N)\backslash H^+$, but this can't be correct because such pairs have a discrete invariant, namely, the N^{th} root of 1 that is the image of 1 under the map

$$\mathbb{Z}/N\mathbb{Z} = \Lambda^2(\mathbb{Z}/N\mathbb{Z}) \xrightarrow{\Lambda^2\alpha} \Lambda^2 A[N] \approx \mu_N,$$

and so they don't form a *connected* family. The correct answer is that they are parametrized by a certain Shimura variety, which is a finite disjoint union of locally symmetric varieties.

We write $\mathbb{S}$ for $\mathbb{C}^\times$ regarded as a real algebraic group:

$$\mathbb{S} = \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\,\mathbb{G}_m.$$

Thus $\mathbb{S}_\mathbb{C} \approx \mathbb{G}_m \times \mathbb{G}_m$, and we normalize the isomorphism so that, on points, $\mathbb{S}(\mathbb{R}) \to \mathbb{S}(\mathbb{C})$ is $z \mapsto (z, \bar{z})$.

The data needed to define a Shimura variety are a connected reductive group G over $\mathbb{Q}$ and a $G(\mathbb{R})$-conjugacy class X of homomorphisms $\mathbb{S} \to G_\mathbb{R}$ satisfying the following axioms:

(SV1) for each $x \in X$, the Hodge structure on $\mathrm{Lie}\,G$ defined by h_x is of type $\{(-1,1),(0,0),(1,-1)\}$;
(SV2) for each $x \in X$, $\mathrm{ad}h(i)$ is a Cartan involution on $G^{\mathrm{ad}}_\mathbb{R}$;
(SV3) the adjoint group G^{ad} has no factor defined over $\mathbb{Q}$ whose real points form a compact group; the identity component of the centre Z of G splits over a CM-field (equivalently, the action of complex conjugation on $X^*(Z^0)$ commutes with the action of all other elements of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$).

The Shimura variety $\mathrm{Sh}(G,X)$ is then the family

$$\mathrm{Sh}_K(G,X) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$$

where K runs through the compact open subgroups of $G(\mathbb{A}_f)$. In forming the quotient, we let $G(\mathbb{Q})$ act on X and $G(\mathbb{A}_f)$ on the left, and K act on $G(\mathbb{A}_f)$ on the right.

Write $G^{\mathrm{ad}}_\mathbb{R} = \prod G_i$ with the G_i simple. For $x \in X$, h_x defines by projection a homomorphism $h_i : \mathbb{S} \to G_i$, and the connected component X^+ of X containing x is a product,

$$X^+ = \prod X_i, \quad X_i = G_i(\mathbb{R})^+/K_i,$$

where K_i is the stabilizer in $G_i(\mathbb{R})^+$ of h_i. The axioms imply that $z \mapsto h_i(z,1)$, if nontrivial, is the "μ" attached to a special node of the Dynkin diagram of $G_{i\mathbb{C}}$, and hence that X_i is a symmetric Hermitian domain. Therefore X^+ is a symmetric Hermitian domain, and X is a finite disjoint union of symmetric Hermitian domains.

Let $G(\mathbb{Q})_+$ be the subgroup of $G(\mathbb{Q})$ of elements mapping into the identity component of $G^{\mathrm{ad}}(\mathbb{R})^+$; it is the stabilizer in $G(\mathbb{Q})$ of any connected component X^+ of X. The strong approximation theorem implies that the double coset space

$$G(\mathbb{Q})_+\backslash G(\mathbb{A}_f)/K$$

is finite. Choose a set of representatives $\{g\}$ for this set, and let Γ_g^{ad} be the image of $\Gamma_g = gKg^{-1} \cap G(\mathbb{Q})_+$ in $G^{\mathrm{ad}}(\mathbb{Q})$—$\Gamma_g$ is a congruence subgroup of $G(\mathbb{Q})$ and Γ_g^{ad} is an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})$. The map

$$\coprod_g \Gamma_g^{\mathrm{ad}}\backslash X^+ \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K,$$

sending $[x] \in \Gamma_g^{\mathrm{ad}}\backslash X^+$ to [x,g] is a homeomorphism. For K sufficiently small, Γ_g^{ad} will be torsion free, and so $\mathrm{Sh}_K(G,X)$ is a finite union of locally symmetric varieties. In particular, it is an algebraic variety.

From now on, we always assume K to be sufficiently small that the groups Γ_g^{ad} will be torsion free.

For varying K, the varieties $\mathrm{Sh}_K(G,X)$ form a projective system, on which the group $G(\mathbb{A}_f)$ acts: an element g of $G(\mathbb{A}_f)$ acts by

$$[x,a]\mapsto[x,ag]:\mathrm{Sh}_K(G,X)\to\mathrm{Sh}_{g^{-1}Kg}(G,X).$$

Because of the second statement in Theorem 1.2, these maps are algebraic.

The inclusion $\mathbb{R}^\times\hookrightarrow\mathbb{C}^\times$ corresponds to an inclusion $\mathbb{G}_m\hookrightarrow\mathbb{S}$; for any homomorphism $h:\mathbb{S}\to G$, we let $w_h=h^{-1}|_{\mathbb{G}_m}$. Because of (SV1), w_h maps into the centre of $G_\mathbb{R}$, and hence is independent of h—we denote it w_X, and call it the *weight* of the Shimura variety. It is always defined over a totally real number field, and we shall be especially interested in those Shimura varieties for which it is defined over $\mathbb{Q}$.

2.1. Examples. Let F be a totally real number field, and let B be a central simple algebra over F of degree 4. Then

$$B\otimes_\mathbb{Q}\mathbb{R}=B\otimes_F(F\otimes_\mathbb{Q}\mathbb{R})=\prod_{\sigma:F\hookrightarrow\mathbb{R}}B\otimes_{F,\sigma}\mathbb{R}\approx M_2(\mathbb{R})^c\times\mathbb{H}^d.$$

Assume $c\geq1$. Let G be the reductive group over $\mathbb{Q}$ such that $G(\mathbb{Q})=B^\times$. Then

$$G_\mathbb{R}\approx\mathrm{GL}_2(\mathbb{R})^c\times(\mathbb{H}^\times)^d$$

and we can define a homomorphism $h:\mathbb{S}\to G_\mathbb{R}$ by requiring that the first c components of $h(a+ib)$ be $\begin{pmatrix}a&-b\\b&a\end{pmatrix}$, and that the remainder are 1. We define X to be the $G(\mathbb{R})$-conjugacy class containing h.

If $F=\mathbb{Q}$ and $B=M_2(\mathbb{Q})$, then $G=\mathrm{GL}_2$ and $X=H^\pm=\mathbb{C}\setminus\mathbb{R}$. For $K=K(N)$,

$$K(N)=_{df}\mathrm{Ker}(\mathrm{GL}_2(\widehat{\mathbb{Z}})\to\mathrm{GL}_2(\widehat{\mathbb{Z}}/N\widehat{\mathbb{Z}})),$$

the Shimura variety is the moduli variety for elliptic curves with level N-structure.

If $B=M_2(F)$, then $G=\mathrm{GL}_{2,F}$, and we get the Hilbert modular varieties. The Shimura variety is a moduli variety for abelian varieties with real multiplication and level structure.

If $F=\mathbb{Q}$ and B is division algebra, the Shimura variety is a family of curves, which, in contrast to the elliptic modular curves, are projective.

If $d\neq0$, so that factors $\mathbb{H}$ occur, then the Shimura variety is not a moduli variety in any useful sense.

Notes: See (Deligne 1971; 1979).

3. Shimura Varieties as Parameter Spaces for Hodge Structures

So far, we have only defined the Shimura variety as a variety over $\mathbb{C}$. In order to be able to talk about the zeta function of a Shimura variety, we need a "canonical" model over a number field. Now it is known that every Shimura variety $\mathrm{Sh}(G,X)$ does have a canonical model over a certain explicit number field $E(G,X)$, but, in general, both the characterization of the canonical model and its construction are somewhat indirect. In particular, neither gives a description of its points in fields containing E, much less on the reduction of the Shimura variety. As a first step toward providing such a description we give a description of the points of the Shimura variety over $\mathbb{C}$ in terms of Hodge structures with tensor and level structures.

3.1. Hodge structures. A *real Hodge structure* is a vector space V over $\mathbb{R}$ together with a homomorphism $h : \mathbb{S} \to \mathrm{GL}(V)$. One then gets a decomposition (the *Hodge decomposition)*

$$V \otimes \mathbb{C} = \oplus V^{p,q}, \quad \overline{V^{p,q}} = V^{q,p}, \qquad h(z) \text{ acts on } V^{p,q} \text{ as } z^{-p}\bar{z}^{-q},$$

and a filtration (the *Hodge filtration)*

$$\cdots \supset F^pV \supset F^{p+1}V \supset \cdots, \qquad F^pV = \oplus_{p' \geq p} V^{p',q'}.$$

The *weight gradation*

$$V = \oplus V_m, \qquad V_m = \oplus_{p+q=m} V^{p,q},$$

is that defined by the map $w_h = h^{-1}|_{\mathbb{G}_m}$. If $V = V_m$, then V is said to have weight m.

A *rational Hodge structure* is a vector space V over $\mathbb{Q}$ together with a Hodge structure on $V \otimes \mathbb{R}$ such that the weight gradation is defined over $\mathbb{Q}$, or, equivalently, such that w_h is a cocharacter of $\mathrm{GL}(V)$ defined over $\mathbb{Q}$.

One similarly defines an integral Hodge structure.

The rational Hodge structure $\mathbb{Q}(m)$ has underlying vector space $(2\pi i)^m\mathbb{Q}$ with $h(z)$ acting as $(z\bar{z})^m$ (hence $\mathbb{Q}(m)$ is of weight $-2m$).

A polarization of a rational Hodge structure of weight m is a morphism of Hodge structures $\psi : V \times V \to \mathbb{Q}(-m)$ such that

$$(x, y) \mapsto (2\pi i)^m \psi(x, h(i)y) : V(\mathbb{R}) \times V(\mathbb{R}) \to \mathbb{R}$$

is symmetric and positive-definite.

If Z is a complete smooth variety over $\mathbb{C}$, then Hodge theory provides $H^m(Z, \mathbb{Q})$ with a canonical polarizable Hodge structure of weight m. An algebraic cycle of codimension m on Z defines a cohomology class which lies in

$$H^{2m}(Z, \mathbb{Q}) \cap H^{m,m},$$

and the Hodge conjecture predicts that each of these $\mathbb{Q}$-vector spaces is generated by algebraic classes. More canonically, the algebraic classes are conjectured to generate the space

$$(H^{2m}(Z, \mathbb{Q}) \otimes \mathbb{Q}(m)) \cap H^{0,0},$$

which can also be described as the subspace of $H^{2m}(Z, \mathbb{Q}(m))$ of vectors (whose image in $H^{2m}(Z, \mathbb{Q}(m)) \otimes \mathbb{R}$ is) fixed by $h(z)$ for all $z \in \mathbb{C}^\times$.

From now on, we consider only polarizable rational Hodge structures.

3.2. A reinterpretation of the notion of a level structure. Fix an N. By an elliptic curve with level N structure, one normally means a pair

$$(A, \alpha : (\mathbb{Z}/N\mathbb{Z})^2 \xrightarrow{\approx} E[N]).$$

We wish to give another interpretation of the set of isomorphism classes of such pairs.

For an elliptic curve (or abelian variety) A over $\mathbb{C}$, we define

$$T_fA = \varprojlim A[n] = \prod T_\ell A,$$

where $A[n]$ is the set of n-torsion points and $T_\ell A$ is the Tate module, and we define

$$V_fA = T_fA \otimes_{\mathbb{Z}} \mathbb{Q}.$$

It is a free $\mathbb{A}_f$-module of rank 2 (more generally, $2\dim A$ if A is an abelian variety).

The category of *abelian varieties up to isogeny* over a field k has as objects the abelian varieties over k, but the morphisms are $\mathrm{Hom}(A,B)\otimes\mathbb{Q}$. Note that $A\mapsto V_fA$ is a functor on the category of abelian varieties up to isogeny, but $A\mapsto T_fA$ is not.

I claim that the isomorphism classes of pairs (A,α) as above are in one-to-one correspondence with the isomorphism classes of pairs $(B,[\eta]:(\mathbb{A}_f)^2\to V_fB)$ where B is an elliptic curve up to isogeny and $[\eta]$ is a $K(N)$-orbit of isomorphisms. Indeed, $(B,[\eta])$ will be isomorphic to a pair $(A,[\eta])$ with $T_fA=\eta(\widehat{\mathbb{Z}}^2)$, and the map

$$\alpha:(\mathbb{Z}/N\mathbb{Z})^2=(\widehat{\mathbb{Z}}/N\widehat{\mathbb{Z}})^2\xrightarrow{\eta}T_fA/NT_fA=A[N]$$

is independent of the choice of η in $[\eta]$.

3.3. Shimura varieties as parameter spaces for Hodge structures. Consider a Shimura variety $\mathrm{Sh}(G,X)$ whose weight is defined over $\mathbb{Q}$. Choose a faithful representation $G\hookrightarrow\mathrm{GL}(V)$. Because G is reductive, there will be a set of tensors $\mathfrak{t}=(t_i)_{i\in I}$, which we may take to be finite, such that G can be characterized as the subgroup of $\mathrm{GL}(V)$ fixing the t_i; more precisely, for any $\mathbb{Q}$-algebra R,

$$G(R)=\{\alpha\in\mathrm{GL}(V(R))\mid \alpha t_i=t_i,\quad i\in I\},\quad V(R)=R\otimes_{\mathbb{Q}}V.$$

Definition 3.1. Consider triples $(W,\mathfrak{s},[\eta])$ consisting of a rational Hodge structure $W=(W,h)$, a family $\mathfrak{s}$ of Hodge cycles indexed by I, and a K-level structure $[\eta]$ on W, i.e., a K-orbit of isomorphisms $\eta:V(\mathbb{A}_f)\to W(\mathbb{A}_f)$, K acting on $V(\mathbb{A}_f)$. We define $\mathcal{H}_K(G,X)$ to be the set of such triples satisfying the following conditions:

(a) there exists an isomorphism of $\mathbb{Q}$-vector spaces $\beta:W\to V$ mapping each s_i to t_i and sending h to h_x, some $x\in X$;
(b) for one (hence every) η representing the level structure, η maps each t_i to s_i.

An isomorphism from one such triple $(W,\mathfrak{s},[\eta])$ to a second $(W',\mathfrak{s}',[\eta'])$ is an isomorphism $\gamma:W\to W'$ of rational Hodge structures mapping each s_i to s'_i and such that $[\gamma\circ\eta]=[\eta']$.

Let $(W,\mathfrak{s},[\eta])$ be an element of $\mathcal{H}_K(G,X)$. Choose an isomorphism $\beta:W\to V$ satisfying (3.1a), so that β sends h to h_x for some $x\in X$. The composite

$$V(\mathbb{A}_f)\xrightarrow{\eta}W(\mathbb{A}_f)\xrightarrow{\beta}V(\mathbb{A}_f),\qquad \eta\in[\eta],$$

sends each t_i to t_i, and is therefore multiplication by an element $g\in G(\mathbb{A}_f)$, well defined up to multiplication on the right by an element of K (corresponding to a different choice of a representative η of the level structure). Since any other choice of β is of the form $q\circ\beta$ for some $q\in G(\mathbb{Q})$, $[x,g]$ is a well-defined element of $G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)/K$.

Proposition 3.2. *The above construction defines a bijection*

$$\alpha_K:\mathcal{H}_K(G,X)/{\approx}\to\mathrm{Sh}_K(G,X)(\mathbb{C}).$$

Proof. Let $(W',\mathfrak{s}',[\eta'])$ be a second system. If

$$\gamma:(W',\mathfrak{s}',[\eta'])\to(W,\mathfrak{s},[\eta])$$

is an isomorphism of triples and $\beta:W(\mathbb{Q})\to V(\mathbb{Q})$ is an isomorphism of vector spaces satisfying (3.1a), then $\beta\circ\gamma$ satisfies (3.1a) for $(W',\mathfrak{s}',[\eta'])$, and it follows that $(W',\mathfrak{s}',[\eta'])$ maps to the same element of $\mathrm{Sh}_K(G,X)$ as $(W,\mathfrak{s},[\eta])$. Conversely, if $(W,\mathfrak{s},[\eta])$ and $(W',\mathfrak{s}',[\eta'])$

map to the same class $[x,g]$, we can choose maps β and β' so that the triples map to the same element of $X \times G(\mathbb{A}_f)$; now $\gamma =_{df} \beta^{-1} \circ \beta'$ is an isomorphism

$$(W', \mathfrak{s}', [\eta']) \to (W, \mathfrak{s}, [\eta]).$$

Finally, for any $(x,g) \in X \times G(\mathbb{A}_f)$, $((V, h_x), \mathfrak{t}, [g])$ maps to $[x,g]$. □

Remark 3.3. A $g \in G(\mathbb{A}_f)$ defines a map

$$(W, \mathfrak{s}, [\eta]) \mapsto (W, \mathfrak{s}, [\eta \circ g]) : \mathcal{H}_K(G,X) \to \mathcal{H}_{g^{-1}Kg}(G,X).$$

For varying K's, the maps α_K are compatible with the actions of $G(\mathbb{A}_f)$.

Remark 3.4. (a) Consider the constant vector bundle $\mathcal{V}$ over X with fibre $V(\mathbb{R})$. When we endow the fibre over x with the Hodge structure defined by h_x, then the Hodge filtrations vary holomorphically on the base X; moreover, the complex structure on X is the unique one for which this is true.

(b) If the centre Z of G is such that $Z(\mathbb{Q})$ is discrete in $Z(\mathbb{A}_f)$, then the family of Hodge structure over X descends to a family of Hodge structures over $\mathrm{Sh}(G,X)$. Axiom (SV1) then implies that the family is a variation of Hodge structures, i.e., Griffiths transversality holds.

(c) Axiom (SV2) implies that the Hodge structures in $\mathcal{H}_K(G,X)$ are polarizable.

(d) If we drop the assumption that the weight is defined over $\mathbb{Q}$, then the above Proposition fails because, although the vector spaces W are defined over $\mathbb{Q}$, their weights are defined only over a totally real field.

Notes: See (Deligne 1979) and (Milne 1990, II.3).

4. Shimura Varieties as Moduli Varieties for Motives.

4.1. Techniques for obtaining models of varieties over number fields. Suppose we are given an algebraic variety Y over $\mathbb{C}$. How do we construct a model of it over a number field E? If we know the equations for Y over $\mathbb{C}$, we can try to find equations for it with coefficients in E, but even when this is possible, it may not be the most useful description of the model.

Roughly speaking, if Y is the solution to a moduli problem over $\mathbb{C}$ (for example, if it represents a functor), and the moduli problem (for example, the functor) is defined over the subfield E of $\mathbb{C}$, then descent theory shows that Y will have a model over E that, in fact, will be a solution to the moduli problem over E.

4.2. Motivic Hodge structures. Consider a pair (Λ, J) where Λ is a free $\mathbb{Z}$-module of rank 2 and J is an automorphism of $\Lambda \otimes \mathbb{R}$ such that $J^2 = -1$. This may not seem to be a very interesting object until you notice that it is an elliptic curve over $\mathbb{C}$: J defines a complex structure on $\Lambda \otimes \mathbb{R}$, and the quotient $\Lambda \otimes \mathbb{R}/\Lambda$ has a unique structure as an elliptic curve. More precisely, $A \mapsto (H_1(A,\mathbb{Z}), J)$ is an equivalence from the category of elliptic curves over $\mathbb{C}$ to the category pairs (Λ, J). It makes sense to speak of an elliptic curve being defined over a subfield of $\mathbb{C}$, or of the conjugate of an elliptic curve by an automorphism of $\mathbb{C}$, whereas neither makes sense for an arbitrary rational Hodge structure. Now a pair (Λ, J) is just an integral Hodge structure of rank 2 and type $\{(-1,0),(0,-1)\}$—take $(\Lambda \otimes \mathbb{C})^{-1,0}$ and $(\Lambda \otimes \mathbb{C})^{0,-1}$ to be the $+i$ and $-i$ eigenspaces respectively.

More generally, we have the following theorem.

Theorem 4.1. *The functor $A \mapsto H_1(A, \mathbb{Q})$ defines an equivalence from the category of abelian varieties over $\mathbb{C}$ up to isogeny to the category of polarizable rational Hodge structures of type $\{(-1,0),(0,-1)\}$.*

Thus, if the Hodge structures in $\mathcal{H}_K(G,X)$ are of this type, then $\mathrm{Sh}(G,X)$ parametrizes abelian varieties with Hodge cycle and level structure.

Definition 4.2. A Hodge structure is *motivic* if it is in the smallest subcategory that contains the cohomology groups of all algebraic varieties and is closed under the formation of tensor products, direct sums, duals, and direct summands. It is *abelian-motivic* if it is in the smallest subcategory containing the cohomology groups of abelian varieties and is closed under the same operations.

4.3. Shimura varieties whose Hodge structures are motivic. For which Shimura varieties are the Hodge structures in $\mathcal{H}_K(G,X)$ motivic? A necessary condition is that the weight be defined over $\mathbb{Q}$, and one hopes that it is also sufficient. Henceforth, we assume the weight is rational.

The *Mumford-Tate group* of a rational Hodge structure (V,h) is the smallest algebraic subgroup H of $\mathrm{GL}(V)$ (in particular, rational over $\mathbb{Q}$) such that $H_{\mathbb{R}}$ contains the image of h. It is connected (because $\mathbb{S}$ is connected) and it is reductive (because I'm assuming Hodge structures to be polarizable). The Mumford-Tate group of an abelian variety A over $\mathbb{C}$ is the Mumford-Tate group of the rational Hodge structure $H_1(A,\mathbb{Q})$.

Satake gave a list of the almost-simple groups over $\mathbb{Q}$ that arise as the derived group of the Mumford-Tate group of an abelian variety, or, more generally as follows: it may happen that an almost-simple group H has finite subgroups N_1 and N_2 such that $N_1 \cap N_2 = 1$ and both H/N_1 and H/N_2 arise as the derived groups of Mumford-Tate groups but not H itself; such an H is also to be included on the list.

Groups not on Satake's list: Among the groups arising in the theory of Shimura varieties, the groups E_6, E_7, and those of mixed type D_n, $n > 4$, do not occur on Satake's list. Also for certain nonmixed types D_n, the simply connected group is not on Satake's list.

A simple group over $\mathbb{Q}$ is of mixed type D_n if its simple factors over $\mathbb{R}$ correspond to special nodes at opposite ends of the Dynkin diagram.

In order to show a group is on Satake's list, one has to find a faithful family of representations (over $\mathbb{Q}$) on symplectic spaces with certain properties. It turns out that the highest weight of each representation will be a fundamental weight, and hence will correspond to a node of Dynkin diagram. For groups of type D, it will be a node at the opposite end from the special node. Since there is no automorphism of the Dynkin diagram switching nodes at opposite ends of the diagram, there is no symplectic representation defined over $\mathbb{Q}$ that over $\mathbb{R}$ gives the correct representations.

Theorem 4.3. *The Hodge structures in $\mathcal{H}_K(G,X)$ are abelian-motivic if and only if G^{der} has a finite covering by a product of groups on Satake's list.*

The Shimura varieties satisfying the condition in the theorem are said to be of *abelian type.* As was mentioned above, it is hoped that the Hodge structures in $\mathcal{H}_K(G,X)$ will be motivic whenever the weight is defined over $\mathbb{Q}$, but this is not known for a single Shimura variety not of abelian type.

We can state things more canonically as follows: let G^{Hdg} be the pro-reductive group attached to the Tannakian category of polarizable rational Hodge structures, and let G^{Mot} be the quotient group attached to the Tannakian subcategory of motives (defined using Hodge cycles). If the weight is defined over $\mathbb{Q}$, each point x of X defines a homomorphism $\rho_x : G^{\mathrm{Hdg}} \to G$: it is the unique homomorphism defined over $\mathbb{Q}$ such that $\rho_x \circ h_{\mathrm{Hdg}} = h_x$. The "hope" is that each ρ_x will factor through G^{Mot}; the theorem is that ρ_x will factor through G^{Mot}, and even the quotient group attached to the category of abelian motives, provided G^{der} has a finite covering by groups on Satake's list.

We explain the theorem in more detail. Let A be an abelian variety, and let q be an endomorphism of $H^*(A, \mathbb{Q}) =_{df} \oplus_r H^r(A, \mathbb{Q})$ as a rational Hodge structure such that $q^2 = q$; then

$$H^*(A, \mathbb{Q}) = \mathrm{Ker}(q) \oplus \mathrm{Im}(q).$$

For an integer m, we define

$$H(A, q, m) = \mathrm{Im}(q) \otimes \mathbb{Q}(m).$$

Then the theorem says that each Hodge structure in the family $\mathcal{H}_K(G, X)$ is a sum of Hodge structures of the form $H(A, q, m)$.

It makes sense to talk of an abelian variety being defined over a subfield of $\mathbb{C}$, but what about Hodge tensors (including q)? The Hodge conjecture would say that the Hodge tensors are all the classes of algebraic cycles on powers of the abelian variety. The Hodge conjecture is not known for abelian varieties, but a theorem of Deligne shows that Hodge cycles do make good sense on abelian varieties over fields of characteristic zero (Deligne 1982). More precisely, as we explain in the next subsection, one can define a good theory of motives over such fields.

4.4. Abelian motives in characteristic zero. Let G be an algebraic group over a field k, or, more generally, a projective limit of such groups. The category $\mathrm{Rep}_k(G)$ of representations of G on finite-dimensional vector spaces over k is a k-linear abelian category with good notions of tensor product and duals. Conversely, every category possessing these properties and also an exact faithful functor ω to $\mathbf{Vec}_k$ preserving the structures can be realized as the category of representations of a pro-algebraic group G. Such a category is called a *neutral Tannakian category.*

For any field k of characteristic zero, it is possible to define a neutral Tannakian category $\mathbf{Mot}(k)$, the category of *abelian motives* over k. Each triple (A, q, m) as above defines a motive $h(A, q, m)$ over k, and if we ignore the Artin motives, each motive is a direct sum of such objects. Once an embedding of k into $\mathbb{C}$ has been chosen, one obtains a functor ω_B from $\mathbf{Mot}(k)$ to rational Hodge structures extending $A \mapsto H_1(A \otimes_k \mathbb{C}, \mathbb{Q})$. Also, once an algebraic closure $\bar{k}$ of k has been chosen, one obtains a functor ω_f from $\mathbf{Mot}(k)$ to $\mathbb{A}_f$-modules with a continuous action of $\mathrm{Gal}(\bar{k}/k)$ extending

$$A \mapsto V_f A, \quad V_f A = T_f A, \quad T_f A = \varprojlim_N A(\bar{k})[N].$$

When $k = \mathbb{C}$, $\omega_f(M) = \omega_B(M) \otimes_{\mathbb{Q}} \mathbb{A}_f$, and the functor ω_B defines an equivalence of $\mathbf{Mot}(\mathbb{C})$ with the category of abelian-motivic Hodge structures.

4.5. The canonical model. Before explaining how to realize the Shimura variety as a moduli variety over a number field, we must define the number field that is the "natural" field of definition for the Shimura variety.

The reflex field. Associated with each $h \in X$, there is a cocharacter

$$\mu_h : \mathbb{G}_m \to G_{\mathbb{C}}, \quad \mu_h(z) = h_{\mathbb{C}}(z, 1),$$

of $G_{\mathbb{C}}$. The set of cocharacters $\{\mu_h\}$ lies in a single $G(\mathbb{C})$-conjugacy class, M_X say. Let T be a maximal torus in G (rational over $\mathbb{Q}$). Since all maximal tori in $G_{\mathbb{C}}$ are conjugate, some element of M_X will have image in $T_{\mathbb{C}}$. But T splits over $\mathbb{Q}^{\mathrm{al}}$, and so M_X has a representative defined over $\mathbb{Q}^{\mathrm{al}}$. If two cocharacters of $G_{\mathbb{Q}^{\mathrm{al}}}$ are $G(\mathbb{C})$-conjugate, then they are $G(\mathbb{Q}^{\mathrm{al}})$-conjugate. Hence any two elements of M_X that are defined over $\mathbb{Q}^{\mathrm{al}}$ lie in the same $G(\mathbb{Q}^{\mathrm{al}})$-conjugacy class, $M_X(\mathbb{Q}^{\mathrm{al}})$ say. We may define the *reflex field* $E(G, X)$ to be the field of definition of $M_X(\mathbb{Q}^{\mathrm{al}})$, i.e., to be the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ stabilizing $M_X(\mathbb{Q}^{\mathrm{al}})$.

Definition 4.4. Let k be a field containing $E(G, X)$. Consider triples $(M, \mathfrak{s}, [\eta])$ consisting of an abelian motive M over k, a family $\mathfrak{s}$ of Hodge cycles on M indexed by I, and a K-level structure $[\eta]$ on M, i.e., a K-orbit of isomorphisms $V(\mathbb{A}_f) \to \omega_f(M)$. We define $\mathcal{M}_K(G, X)(k)$ to be the set of such triples satisfying the following conditions:

(a) there exists an isomorphism of $\mathbb{Q}$-vector spaces $\beta : \omega_B(M) \to V$ mapping each s_i to t_i and sending h to h_x, some $x \in X$;
(b) for one (hence every) η representing the level structure, η maps each t_i to s_i; moreover, $[\eta]$ is stable under the action of $\mathrm{Gal}(k^{\mathrm{al}}/k)$.

An isomorphism from one such triple $(M, \mathfrak{s}, [\eta])$ to a second $(M', \mathfrak{s}', [\eta'])$ is an isomorphism $\gamma : M \to M'$ of motives mapping each s_i to s_i' and such that $[\omega_f(\gamma) \circ \eta] = [\eta']$.

Remark 4.5. The functor ω_B depends on the choice of an embedding of k into $\mathbb{C}$, but if we choose this to be an $E(G, X)$-embedding, then the condition (a) and the set $\mathcal{M}_K(G, X)$ do not depend on the choice of the embedding.

The functor ω_B defines a bijection

$$\mathcal{M}_K(G, X)(\mathbb{C})/{\approx} \to \mathcal{H}_K(G, X)/{\approx} .$$

On combining this with the bijection in Proposition 3.2, we obtain a bijection

$$\mathcal{M}_K(G, X)(\mathbb{C})/{\approx} \to \mathrm{Sh}_K(G, X)(\mathbb{C}).$$

This can be shown to define $\mathrm{Sh}_K(G, X)(\mathbb{C})$ as a moduli variety, and the moduli problem is defined over $E(G, X)$. Therefore, as was discussed briefly in Section 4.1, this defines $\mathrm{Sh}_K(G, X)$ as a moduli variety over $E(G, X)$. In particular, this implies the following theorem.

Theorem 4.6. *There is a canonical model of* $\mathrm{Sh}_K(G, X)$ *over* $E(G, X)$ *and a map*

$$\beta(k) : \mathcal{M}_K(G, X)(k) \to \mathrm{Sh}_K(G, X)(k)$$

for each field $k \supset E(G, X)$*; the maps* $\beta(k)$ *are functorial in* k *and commute with the action of* $G(\mathbb{A}_f)$*; each map* $\beta(k)$ *defines a map*

$$\mathcal{M}_K(G, X)(k)/{\approx} \to \mathrm{Sh}_K(G, X)(k),$$

which is bijective when k *is algebraically closed, and is bijective for all fields if* $Z(\mathbb{Q})$ *is discrete in* $Z(\mathbb{A}_f)$*. For* $k = \mathbb{C}$*, the map is that defined above.*

Example 4.7. (a) Let $G = \mathrm{GL}_2$ and $X = H^{\pm}$. Here the reflex field is $\mathbb{Q}$, and $\mathrm{Sh}_K(G, X)$ is realized as a moduli variety over $\mathbb{Q}$ for elliptic curves with level structures. For example, for $N \geq 3$, $\mathrm{Sh}_{K(N)}(G, X)$ is defined over $\mathbb{Q}$, and its points in any field (or ring) L containing $\mathbb{Q}$ are the isomorphism classes of pairs $(B, [\eta] : (\mathbb{A}_f)^2 \xrightarrow{\approx} V_f(B))$ discussed in Section 3.2.

(b) Let $G = \mathrm{GL}_{2,F}$ with X as above. Again the reflex field is $\mathbb{Q}$, but the moduli problem is coarse because $Z(\mathbb{Q}) = F^{\times}$ is not discrete in $Z(\mathbb{A}_f) = (\mathbb{A}_f \otimes F)^{\times}$. For any field k containing $\mathbb{Q}$, $\mathcal{M}_K(G, X)(k)$ is the set of triples $(A, \mathfrak{s}, [\eta])$ where A is an abelian variety up to isogeny over k such that $\mathrm{End}(A) \supset F$ and $\dim A = [F : \mathbb{Q}]$. The set $\mathfrak{s}$ can be taken to be any set of generators for F.

(c) Let ψ be a nondegenerate skew-symmetric form on a finite-dimensional vector space V over $\mathbb{Q}$, and let $G = \mathrm{GSp}(\psi)$, the group of symplectic similitudes (automorphisms of V fixing ψ up to a rational number). Let X be the set of Hodge structures of type $\{(-1, 0), (0, -1)\}$ on V for which $\pm 2\pi i\psi$ is a polarization. Then G is the subgroup of $\mathrm{GL}(V) \oplus \mathrm{GL}\,\mathbb{Q}(1)$ fixing $2\pi i\psi$. This is the subgroup of $\mathrm{GL}(V \oplus \mathbb{Q}(1))$ commuting with (i.e., fixing) the projections onto the factors and fixing ψ. Again the reflex field is $\mathbb{Q}$, and for any field containing $\mathbb{Q}$, $\mathcal{M}_K(G, X)(k)$ is the set of triples $(A, \lambda, [\eta])$ where A is an abelian variety up to isogeny of dimension $\frac{1}{2} \dim V$ and λ is a polarization of A.

(d) Let $G = \mathrm{PGL}_2$, and let X be the obvious conjugacy class of homomorphisms $\mathbb{S} \to G_{\mathbb{R}}$. Then G has a natural representation on a three-dimensional vector space, and the motives are of the form $\mathrm{Sym}^2(A)$, for A an elliptic curve.

To proceed further, we need to assume that there is a good theory of abelian motives in characteristic p and a good reduction functor. Thus the next two sections are heuristic: we make plausible assumptions in order to discover what the description of the points over the finite fields should be.

Notes: See (Deligne and Milne 1982) for the notion of a Tannakian category and of a motive. Theorem 4.3 is proved in (Milne 1994b).

5. Integral Models

5.1. A criterion for good reduction. When should $\mathrm{Sh}_K(G, X)$ have good reduction at a prime v of $E = E(G, X)$ lying over p? If $p|N$ then it is known that the moduli variety $\mathrm{Sh}_{K(N)}(\mathrm{GL}_2, H^{\pm})$ of elliptic curves with level N structure does not have good reduction at p (see, for example, Deligne and Rapoport 1973). Thus we should assume that K contains a maximal compact subgroup K_p at p; in fact, we may as well assume $K = K^p \cdot K_p$ where K^p is a compact open subgroup of $G(\mathbb{A}_f^p)$, $\mathbb{A}_f^p = (\prod_{\ell \neq p} \mathbb{Z}_\ell) \otimes_{\mathbb{Z}} \mathbb{Q}$. However, even this is not sufficient to ensure that $\mathrm{Sh}_K(G, X)$ has good reduction at p: if $\mathrm{Sh}(G, X)$ is the Shimura variety associated with a quaternion algebra over $\mathbb{Q}$, then $\mathrm{Sh}_{K^p \cdot K_p}(G, X)$, K_p maximal, will have good reduction at p only if p does not divide the discriminant of B. That the following should be true was suggested in (Langlands, 1976, p411).

Conjecture 5.1. *The variety* $\mathrm{Sh}_K(G, X)$, $K = K^p \cdot K_p$ *has good reduction at* $v|p$ *if* K_p *is a hyperspecial group of* $G(\mathbb{Q}_p)$.

The algebraic group $G_{\mathbb{Q}_p}$ will have a hyperspecial subgroup if and only if it has a smooth model G_p over $\mathbb{Z}_p$ whose reduction modulo p is again a connected reductive group; the hyperspecial subgroup is then $G_p(\mathbb{Z}_p)$. In order for G to have a hyperspecial subgroup, it is obviously necessary that G be quasi-split over $\mathbb{Q}_p$ and split over an unramified extension, and the Bruhat-Tits theory of buildings shows that this condition is also sufficient.

5.2. The points on $\mathrm{Sh}_p(G,X)$. From now on we assume that $K = K^p \cdot K_p$ with K_p hyperspecial, and we choose a lattice $V(\mathbb{Z}_p)$ in $V(\mathbb{Q}_p)$ whose stabilizer is G_p; this means that for any $\mathbb{Z}_p$-algebra R, $G_p(R)$ is the stabilizer of $V(R)$ in $V(R) \otimes \mathbb{Q}_p$.

Since we shall need to know the points on the Shimura variety with coordinates in nonalgebraically closed fields, we shall assume that $Z(\mathbb{Q})$ is discrete in $Z(\mathbb{A}_f)$. Then we might as well pass to the limit over smaller and smaller groups K^p, and set

$$\mathrm{Sh}_p(G,X) = \varprojlim_{K^p} \mathrm{Sh}_{K^p \cdot K_p}(G,X).$$

An isomorphism $\eta : V(\mathbb{A}_f) \to \omega_f(M)$ can be decomposed into a product $\eta^p \times \eta_p$ where η^p is an isomorphism $V(\mathbb{A}_f^p) \to \omega_f^p(M)$ and η_p is an isomorphism $V(\mathbb{Q}_p) \to \omega_p(M)$. Here $\mathbb{A}_f^p$ is the ring of finite adèles away from p, i.e., the restricted product of the $\mathbb{Q}_\ell$ for $\ell \neq p, \infty$, and $\omega_f^p(M)$ and $\omega_p(M)$ are suitable étale realizations of M.

Definition 5.2. Let k be a field containing $E(G,X)$. Consider quadruples $(M, \mathfrak{s}, \eta^p, \Lambda_p)$ consisting of an abelian motive M over k, a family $\mathfrak{s}$ of Hodge cycles indexed by I, an isomorphism $V(\mathbb{A}_f^p) \to \omega_f^p(M)$, and a lattice Λ_p in $\omega_p(M)$. We define $\mathcal{M}_p(G,X)(k)$ to be the set of such triples satisfying the following conditions:

(a) there exists an isomorphism of $\mathbb{Q}$-vector spaces $\beta : \omega_B(M) \to V$ mapping each s_i to t_i and sending h to h_x, some $x \in X$;
(b) the isomorphism η^p maps each t_i to s_i and is invariant under the action of $\mathrm{Gal}(\bar{k}/k)$;
(c) Λ_p is a $\mathbb{Z}_p$-lattice in $\omega_p(M)$, stable under the action of $\mathrm{Gal}(\bar{k}/k)$, for which there exists an isomorphism
$$V(\mathbb{Q}_p) \to \omega_p(M)$$
mapping each t_i onto s_i and such that $V(\mathbb{Z}_p)$ maps onto Λ_p.

There is an obvious notion of an isomorphism from one such quadruple to a second.

The map $(M, \mathfrak{s}, [\eta]) \mapsto (M, \mathfrak{s}, \eta^p, \eta_p(V(\mathbb{Z}_p))$ defines a bijection

$$\varprojlim_{K^p} \mathcal{M}_{K^p K_p}(G,X)(k) \to \mathcal{M}_p(G,X)(k).$$

Hence, for any field k containing $E(G,X)$, there is a canonical bijection

$$\mathcal{M}_p(G,X)(k)/{\approx} \to \mathrm{Sh}_p(k).$$

5.3. The canonical integral model. Let v be a prime of E lying over p. Then v is unramified over p, and we let B be the completion of the maximal unramified extension of E_v. Then B is the field of fractions of the ring of Witt vectors over an algebraic closure $\mathbb{F}$ of the residue field at v. Let $\mathcal{O}_v$ be the ring of integers in E_v. We have the diagram:

$$\begin{array}{ccccc} B & --- & W & --- & \mathbb{F} \\ | & & | & & | \\ E_v & --- & \mathcal{O}_v & --- & k(v). \end{array}$$

We want to understand the points of $\mathrm{Sh}_p(G,X)$ with coordinates in $\mathbb{F}$, but first we need a model of $\mathrm{Sh}_p(G,X)$ over $\mathcal{O}_v$, and we need to have a description of the points on the model.

Consider $S_N = \mathrm{Sh}_{K(N)}(\mathrm{GL}_2, H^{\pm})$. Its points over B are isomorphism classes of pairs

$$(A, \alpha : (\mathbb{Z}/N\mathbb{Z})^2 \xrightarrow{\approx} A[N])$$

where A is an elliptic curve over B and α is a level N-structure on A. If p does not divide N, then S_N has good reduction, and, in fact extends canonically to a smooth curve over W, whose points are the pairs (A, α) over B such that A has good reduction. Thus $S_N(W)$ will be a proper subset of $S_N(B)$. But recall that the Néron-Ogg-Shafarevich criterion for good reduction says that any elliptic curve A over B having all its N-torsion points rational over B has good reduction. Therefore, when we pass to the limit,

$$\varprojlim_{p\nmid N} S_N(B) = \varprojlim_{p\nmid N} S_N(W).$$

This example suggests that $\mathrm{Sh}_p(G,X)$ should have a model over $\mathcal{O}_v$ with the property that

$$\mathrm{Sh}_p(G,X)(W) = \mathrm{Sh}_p(G,X)(B).$$

In fact, an extension of this property can be used to characterize an *canonical integral model* over W, namely, it should be a smooth model S of $\mathrm{Sh}_p(G,X)$ over $\mathcal{O}_v$ such that

$$S(Y) = \mathrm{Sh}_p(G,X)(Y \otimes_{\mathcal{O}_v} E_v)$$

for every regular $\mathcal{O}_v$-scheme Y such that $Y \otimes_{\mathcal{O}_v} E_v$ is dense in Y. A theorem of Chai and Faltings allows one to verify that Siegel modular variety satisfies this condition, at least for $p \neq 2$.

5.4. The points with coordinates in $\mathbb{F}$. In the following, we assume that a smooth integral model exists. We expect a commutative diagram:

$$\begin{array}{ccccc} \mathcal{M}_p(W)/\approx & \xrightarrow{1:1} & \mathrm{Sh}_p(W) & = & \mathrm{Sh}_p(B) \\ \downarrow & & \downarrow & & \\ \mathcal{M}_p(\mathbb{F})/\approx & \xrightarrow{1:1} & \mathrm{Sh}_p(\mathbb{F}) & & \end{array} \tag{5.1}$$

The vertical arrow at right is onto because $\mathrm{Sh}_p(W)$ is smooth. Unfortunately, the description of $\mathcal{M}_p$ given in Definition 5.2 does not make sense in characteristic p, because p-adic étale cohomology is pathological in characteristic p. We need to change the description so that it is in terms of the de Rham (or crystalline) cohomology.

Let Z be a smooth projective variety over B with good reduction. It has two p-adic cohomologies:

- the p-adic étale cohomology groups; these are finite-dimensional $\mathbb{Q}_p$-vector spaces with an action of $\mathrm{Gal}(B^{\mathrm{al}}/B)$;
- the de Rham cohomology groups; these are finite-dimensional B-vector spaces with a canonical filtration. Since we are assuming Z reduces to a smooth variety Z_0 over $\mathbb{F}$, the de Rham cohomology of Z will be equal to the crystalline cohomology of Z_0, and hence acquires a p-linear Frobenius operator ϕ.

The same should be true for our abelian motives. Grothendieck conjectured many years ago that there should be a canonical way of going from one cohomology theory to the other. Thanks to the work of Fontaine and others, this is now well understood in the above situation, but is less well understood on the level of lattices.

In attempting to translate the Definition 5.2 from étale cohomology to de Rham cohomology, I arrived at the following definition.

Definition 5.3. Consider quadruples $(M, \mathfrak{s}, \eta^p, \Lambda_{\mathrm{crys}})$ consisting of an abelian motive M over B, a family $\mathfrak{s}$ of algebraic classes indexed by I, an isomorphism $V(\mathbb{A}_f^p) \to \omega_f^p(M)$ and a lattice

Λ_{crys} in $\omega_{\text{dR}}(M)$. We define $\mathcal{M}'_p(G,X)(B)$ to be the set of such triples satisfying the following conditions:

(a) there exists an isomorphism of $\mathbb{Q}$-vector spaces $\beta : \omega_B(M) \to V$ mapping each s_i to t_i and sending h to h_x, some $x \in X$;
(b) the isomorphism η^p maps each t_i to s_i and is invariant under the action of $\text{Gal}(\bar{k}/k)$;
(c) Λ_{crys} is a W-lattice in $\omega_{\text{dR}}(M)$ that is strongly divisible, i.e., such that $\sum p^{-i}\phi(Fil^i\Lambda) = \Lambda$, and such that there exists an isomorphism
$$\eta_{\text{crys}} : V(B) \to \omega_{\text{dR}}M$$
sending each t_i to s_i and mapping $Filt(\mu_0^{-1})$ to the Hodge filtration.

Here μ_0 is a cocharacter of G_B representing M_X and well-adapted for K_p. More precisely, $\mu_0 \in X_*(T)$ where T is a maximal B-split torus in G_B containing the maximal $\mathbb{Q}_p$-split torus corresponding to appartment containing the hyperspecial point fixed by K_p.

I expect that Fontaine's theory provides a bijection
$$\mathcal{M}_p(B) \leftrightarrow \mathcal{M}'_p(B),$$
at least if p is not too small relative to the lengths of the filtrations. Furthermore, I expect that the diagram (5.1) exists with $\mathcal{M}'_p(\mathbb{F})$ defined as follows.

Definition 5.4. Consider quadruples $(M, \mathfrak{s}, \eta^p, \Lambda_{\text{crys}})$ consisting of an abelian motive M over $\mathbb{F}$, a family $\mathfrak{s}$ of algebraic cycles indexed by I, an isomorphism $V(\mathbb{A}_f^p) \to \omega_f^p(M)$, and a lattice Λ_{crys} in $\omega_{\text{crys}}(M)$. We define $\mathcal{M}'_p(G,X)(\mathbb{F})$ to be the set of such quadruples that lift to a quadruple in $\mathcal{M}'(G,X)(B)$.

In particular, we should have a bijection
$$\mathcal{M}'_p(\mathbb{F})/\!\approx \to \text{Sh}_p(\mathbb{F})$$
commuting with the actions of $G(\mathbb{A}_f^p)$ and the Frobenius automorphisms. Call a pair $N = (M, \mathfrak{s})$ *admissible* if there exists an η^p and a Λ_{crys} such that $(M, \mathfrak{s}, \eta^p, \Lambda_{\text{crys}}) \in \mathcal{M}'_p(\mathbb{F})$. Fix an admissible N and define $S(N)$ to be the set of isomorphism classes of quadruples $(M, \mathfrak{s}, \eta^p, \Lambda_{\text{crys}})$ in $\mathcal{M}'_p(\mathbb{F})$ with $(M, \mathfrak{s}) \approx N$. Then
$$\text{Sh}_p(\mathbb{F}) = \coprod_N S(N)$$
where the disjoint union is over a set of representatives for the isomorphism classes admissible N's. Moreover
$$S(N) = I(N)\backslash X^p(N) \times X_p(N)$$
where $I(N)$ is the set of automorphisms of N and $X^p(N)$ and $X_p(N)$ are the sets of η^p and Λ_{crys} such that $(N, \eta^p, \Lambda_{\text{crys}}) \in \mathcal{M}'_p(\mathbb{F})$.

Note that $X^p(N)$ is a principal homogeneous space for $G(\mathbb{A}_f^p)$—therefore the choice of an element of $X^p(N)$ determines a bijection $G(\mathbb{A}_f^p) \to X^p(N)$. Similarly, the choice of an isomorphism $\beta : \omega_{\text{crys}}(M) \to V(B)$ sending each s_i to t_i determines an explicit description of $X_p(N)$. There will a $b \in G(B)$ such that action of the Frobenius ϕ on $\omega_{\text{crys}}(M)$ corresponds to $x \mapsto b{\cdot}\sigma x$ on $V(B)$. Here σ is the map on $V(B) = B \otimes_{\mathbb{Q}_p} V$ by the Frobenius automorphism of B. Under β, Λ_{crys} will correspond to $gV(W)$ for some $g \in G(B)$ and the map $\Lambda_{\text{crys}} \mapsto gG(W)$ determines a bijection
$$X_p(N) \to \{g \cdot G(W) \in G(B)/G(W) \mid g^{-1} \cdot b \cdot \sigma g \in G(W) \cdot \mu_0(p^{-1}) \cdot G(W)\}.$$

In order for this description of $\mathrm{Sh}_p(\mathbb{F})$ to be useful, we need to understand the admissible pairs $N = (M, \mathfrak{s})$. Fortunately, this is possible (conjecturally).

Notes: See (Milne 1992, §2) and (Milne 1994b, §4).

6. Abelian Motives over $\mathbb{F}$

We need a down-to-earth description of the category of abelian motives over a $\mathbb{F}$. It should be a Tannakian category over $\mathbb{Q}$ but with no exact tensor functor to the category of $\mathbb{Q}$-vector spaces. According to the Tannakian philosophy, such a category should be equivalent to the category of representations of a groupoid.

A *groupoid in sets* is a small category in which every morphism has an inverse. Thus it consists of a set S of objects, a set $\mathfrak{G}$ of morphisms, maps $t, s : \mathfrak{G} \rightrightarrows S$ sending each object to its target and source, and a law of composition

$$\mathfrak{G} \underset{s,S,t}{\times} \mathfrak{G} \to \mathfrak{G} \text{ where } \mathfrak{G} \underset{s,S,t}{\times} \mathfrak{G} = \{(h,g) \in \mathfrak{G} \times \mathfrak{G} \mid s(h) = t(g)\}.$$

For example, if S has a single element, then $\mathfrak{G}$ is just a group. For each $a \in S$,

$$\mathfrak{G}_a =_{df} \mathrm{Aut}(a) = \{g \in \mathfrak{G} \mid s(g) = a = t(g)\}$$

is a group. If $\mathfrak{G}$ is *transitive,* i.e., $\mathrm{Hom}(a,b)$ is always nonempty, then these groups are all isomorphic, but not (quite) canonically so unless they are commutative.

Now consider a morphism of schemes $S \to S_0$, for example, $\mathrm{Spec}\,\mathbb{Q}^{\mathrm{al}} \to \mathrm{Spec}\,\mathbb{Q}$. An S/S_0-groupoid is a scheme $\mathfrak{G}$ over S_0 together with two S_0-morphisms $t, s : \mathfrak{G} \rightrightarrows S$ and a law of composition

$$\mathfrak{G} \underset{s,S,t}{\times} \mathfrak{G} \to \mathfrak{G}$$

such that, for all S_0-schemes T, $(S(T), \mathfrak{G}(T), (t,s), \circ)$ is a groupoid in sets. Thus a groupoid in schemes generalizes the notion of a group scheme, just as a groupoid in sets generalizes the notion of group. We shall always assume that our groupoids are *transitive*, i.e., that the map $(t,s) : \mathfrak{G} \to S \times_{S_0} S$ is surjective and flat. The kernel $G = \mathfrak{G}^{\Delta} =_{df} (t,s)^{-1}(\Delta)$ of $\mathfrak{G}$ is a group scheme over S, or over S_0 when G is commutative.

A group scheme G over S_0 defines a "trivial" S/S_0-groupoid, $\mathfrak{G}_G = G \times_{S_0} (S \times_{S_0} S)$.

A vector space V over $\mathbb{Q}^{\mathrm{al}}$ defines a $\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}$-groupoid $\mathfrak{G}_V$. There is an obvious notion of a morphism of two groupoids, and a *representation* of a groupoid is a morphism $\mathfrak{G} \to \mathfrak{G}_V$.

In general, the $\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}$-groupoids with a given kernel G are classified by a nonabelian cohomology group $H^2(\mathbb{Q}, G)$; when the kernel is commutative, then this becomes a more usual abelian cohomology group.

Let T be a torus over a field $\mathbb{Q}$. If T is split, then the irreducible representations are classified by the characters $X^*(T)$: the representation corresponding to χ is one-dimensional, and T acts via χ.

More generally, T will split over $\mathbb{Q}^{\mathrm{al}}$, and the irreducible representations are classified by the orbits of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ in $X^*(T)$. Given χ, let $E(\chi)$ be the subfield of $\mathbb{Q}^{\mathrm{al}}$ fixed by the stabilizer of χ. There is a homomorphism $T \to \operatorname{Res}_{E(\chi)/\mathbb{Q}} \mathbb{G}_m$ corresponding to the homomorphism of character groups

$$\mathbb{Z}^{\operatorname{Hom}(E(\chi),\mathbb{Q}^{\text{al}})} \to X^*(T), \quad \sum n_\sigma \sigma \mapsto \sum n_\sigma \sigma\chi.$$

The representation corresponding to the orbit of χ is the composite of this homomorphism with the obvious representation of $\operatorname{Res}_{E(\chi)/\mathbb{Q}} \mathbb{G}_m$ on $E(\chi)$ regarded as a $\mathbb{Q}$-vector space. The endomorphism algebra of the representation is $E(\chi)$.

For a groupoid $\mathcal{T}$ with kernel a torus T, the irreducible representations are again parametrized by the orbits of $\operatorname{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$ in $X^*(T)$, but now the representation has as endomorphism algebra a division algebra with centre $E(\chi)$; moreover, the class of the division algebra in the Brauer group of $E(\chi)$ is the image of the class of $\mathcal{T}$ under the homomorphism

$$H^2(k,T) \to H^2(k, \operatorname{Res}_{E(\chi)/k} \mathbb{G}_m) = H^2(E(\chi), \mathbb{G}_m) = \operatorname{Br}(E(\chi)).$$

What should the groupoid attached to the category of abelian motives over $\mathbb{F}$ be? Each abelian variety defines an abelian motive, and hence a representation of the groupoid. Thus, we seek a groupoid that contains among its represenations, one representation for each simple abelian variety over $\mathbb{F}$, and whose category of representations is generated by such representations.

Fortunately, Tate and Honda have classified the isogeny classes of simple abelian varieties over a finite field. Let A be a simple abelian variety over $\mathbb{F}_q$. Weil showed that Frobenius endomorphism π of A is an algebraic integer with the property that $|\rho\pi| = q^{\frac{1}{2}}$ for every $\rho : \mathbb{Q}[\pi] \hookrightarrow \mathbb{C}$. Call such a π a *Weil q-integer*. To each simple abelian variety over $\mathbb{F}_q$ we can attach the set of conjugates of π in $\mathbb{Q}^{\text{al}}$ (algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$), and the theorem of Honda and Tate says that the map $A \mapsto \{\pi\}$ gives a bijection from the set of isogeny classes of abelian varieties and to the set of Galois orbits of Weil q-integers. Moreover, $\operatorname{End} A \otimes \mathbb{Q}$ is a central division algebra over $\mathbb{Q}[\pi]$ whose invariant i_v at a prime v of $\mathbb{Q}[\pi]$ is determined by the rule,

$$||\pi||_v = q^{i_v}$$

where $||\cdot||_v$ denotes normalized valuation.

Let $W(q)$ be the subgroup of $\mathbb{Q}^{\text{al}\times}$ generated by the Weil q-integers—its elements will be called There is a map $\pi \mapsto \pi^N : W(q) \to W(q^N)$, and we let

$$W(p^\infty) = \varinjlim_N W(p^N).$$

An element π of $W(p^\infty)$ is represented by a Weil p^n-number π_n for some n, and π_n and π_m represent the same element of $W(p^\infty)$ if and only if $\pi_n^{mN} = \pi_m^{nN}$ for some N. Define $\mathbb{Q}\{\pi\} = \mathbb{Q}[\pi_n]$ where π_n is chosen so that $[\mathbb{Q}[\pi_n] : \mathbb{Q}]$ is as small as possible. With the same π_n, let $\delta(\pi)$ be the element of $\operatorname{Br}(\mathbb{Q}\{\pi\})$ whose invariant i_v satisfies

$$||\pi_n||_v = (p^n)^{i_v}.$$

The theorem of Honda and Tate then gives a bijection from the set of isogeny classes of abelian varieties over $\mathbb{F}$ and to the set of Galois orbits in $W(p^\infty)$ represented by a Weil p^n-integer for some n; moreover the endomorphism algebra of the abelian variety corresponding to π is a central division algebra over $\mathbb{Q}\{\pi\}$ with invariant $\delta(\pi)$.

It is now evident that the groupoid $\mathfrak{P}$ we seek should have kernel a pro-torus P with $X^*(P) = W(p^\infty)$; moreover, the class δ of $\mathfrak{P}$ in $H^2(\mathbb{Q}, P)$ should map to $\delta(\pi)$ under the map

$H^2(\mathbb{Q}, P) \to \mathrm{Br}(\mathbb{Q}\{\pi\})$ defined by π. Happily[2], such a groupoid does exist, and, in fact, is uniquely determined up to a (nonunique) isomorphism by these conditions.

Conjecturally, the category of abelian motives over $\mathbb{F}$ is equivalent to the category of representations ϕ of $\mathfrak{P}$, and the category of pairs $N = (M, \mathfrak{s})$, motives with tensors, should be equivalent to the category of morphisms $\phi : \mathfrak{P} \to \mathfrak{G}_G$. Now, it is possible to attach to each such morphism ϕ a set

$$S(\phi) = I(\phi)\backslash X^p(\phi) \times X_p(\phi),$$

and the conjecture of Langlands and Rapoport takes the form

$$\mathrm{Sh}_p(\mathbb{F}) = \coprod_{\phi} S(\phi)$$

where[3] the disjoint union is over the set of isomorphism classes of "admissible" morphisms $\phi : \mathfrak{P} \to \mathfrak{G}_G$. The set $I(\phi)$ is the automorphism group of ϕ, $X^p(\phi)$ is a principal homogenous space for $G(\mathbb{A}_f^p)$, and $X_p(\phi)$ has a description similar to that of $X_p(N)$.

In order to have a completely down-to-earth conjecture, it remains to characterize the "admissible" homomorphisms, namely, those that conjecturally correspond to the admissible pairs $N = (M, \mathfrak{s})$.

A point $x \in X$ is *special* if the image of h_x is contained in $T_{\mathbb{R}}$ for some rational torus $T \subset G$. Langlands and Rapoport (1987) attach to such an x a special homomorphism $\phi_x : \mathfrak{P} \to \mathfrak{G}_G$. Every special homomorphism should be admissible. On the other hand, it is possible to give a necessary condition that ϕ be admissible, namely, a local condition for each prime ℓ (including p and ∞) and the condition that $\mathfrak{P} \to \mathfrak{G}_G \to \mathfrak{G}_{G^{\mathrm{ab}}}$ be special. Happily, the two conditions, one sufficient and the other necessary, are equivalent when G^{der} is simply connected. Thus, in this case, one can take either condition as the definition. When G^{der} is not connected, it is known that one must define a morphism to be admissible if it is isomorphic to a special homomorphism.

Notes: This section motivates the conjecture (Langlands and Rapoport 1987, p169). See also (Milne 1992, §3,4) and (Milne 1994a).

7. The Formula for the Number of Points in a Finite Field

The Shimura variety Sh_p has no points in a finite field, because the objects it parametrizes have infinite level structure—for example, there is no elliptic curve defined over a finite field and having its N-torsion points rational over the field for all N prime to p. To obtain a meaningful result, we must put the K^p back in, and look at

$$\mathrm{Sh}_K(G, X) = \mathrm{Sh}_p(G, X)/K^p, \quad K = K^p \cdot K_p.$$

Assume

- the weight is defined over $\mathbb{Q}$;
- $Z(\mathbb{Q})$ is discrete in $Z(\mathbb{A}_f)$, $Z = Z(G)$;
- G^{der} is simply connected.

[2] Otherwise the Tate conjecture would fail for abelian varieties over finite fields!

[3] More precisely, the conjecture says that there exists a one-to-one correspondence between the two sides, and the correspondence can be chosen to respect the actions of $G(\mathbb{A}_f^p)$ and the Frobenius automorphism.

Then it is possible to derive from the conjecture of Langlands and Rapoport a formula of the following shape:

$$\text{Card}\ \text{Sh}_K(\mathbb{F}_q) = \sum_{(\gamma_0;\gamma,\delta)} c(\gamma_0)\cdot I(\gamma_0;\gamma,\delta).$$

The sum is over a set of representatives for triples $(\gamma_0;\gamma,\delta)$ satisfying a certain cohomological condition,

$$c(\gamma_0) = \text{Card}(\text{Ker}(\text{Ker}^1(\mathbb{Q}, I_0) \to H^1(\mathbb{Q}, G)))$$

measures the failure of a Hasse principle, and

$$I(\gamma_0;\gamma,\delta) = \text{vol}(I(\mathbb{Q})\backslash I(\mathbb{A}_f))\cdot O_\gamma(f^p)\cdot TO_\delta(\phi_r)$$

is the product of a volume with an orbital integral and a twisted orbital integral. The triples $(\gamma_0;\gamma,\delta)$ are of the following form:

- γ_0 is a semisimple element of $G(\mathbb{Q})$ that is elliptic in $G(\mathbb{R})$;
- $\gamma = (\gamma(\ell))_{\ell\neq p,\infty}$ is an element of $G(\mathbb{A}_f^p)$ such that, for all ℓ, $\gamma(\ell)$ becomes conjugate to γ_0 in $G(\mathbb{Q}_\ell^{\text{al}})$;
- δ is an element of $G(B(\mathbb{F}_q))$ such that

$$\mathcal{N}\delta =_{df} \delta\cdot\sigma\delta\cdot\ldots\cdot\sigma^{n-1}\delta, \quad n = [\mathbb{F}_q : \mathbb{F}_p]$$

  becomes conjugate to γ_0 in $G(\mathbb{Q}_p^{\text{al}})$.

Moreover I_{γ_0} is the centralizer of γ_0 in G; I is a certain inner form of I_0 defined by local conditions involving the $\gamma(\ell)$ and δ.

Variants of the formula occur in the writings of Langlands, but it is stated most definitively in (Kottwitz 1990, §3). Historically, the form of the formula was suggested, not by studying the points on the Shimura variety, but by examining what one needs to prove that the zeta function takes the form conjectured by Langlands. The passage from the formula to the zeta function is now a problem in representation theory.

Notes: The derivation of the formula for the cardinality of $\text{Sh}_K(\mathbb{F}_q)$ from the conjecture of Langlands and Rapoport is carried out in (Milne 1992, §4,5).

References

Baily, W., and Borel, A., Compactification of arithmetic quotients of bounded symmetric domains, Ann. of Math. 84, **1966**, 442–528 (= Borel, A., Coll. Papers II, no. 69).

Borel, A., Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem, J. Differ. Geom. 6, **1972**, 543–560 (= Coll. Papers III, no. 95).

Deligne, P., Travaux de Shimura, Sém. Bourbaki, Exposé 389, Lecture Notes in Math. 244, Springer, Heidelberg, **1971**.

Deligne, P., Variétés de Shimura: Interprétation modulaire, et techniques de construction de modèles canoniques, Proc. Symp. Pure Math., 33, **1979**, part 2, 247–290.

Deligne, P., Hodge cycles on abelian varieties (Notes by J.S. Milne), in Hodge Cycles, Motives, and Shimura Varieties, Lecture Notes in Math, 900, Springer, **1982**, 9–100.

Deligne, P., and Milne, J., Tannakian categories, in Hodge Cycles, Motives, and Shimura Varieties, Lecture Notes in Math, 900, Springer, **1982**, 101–228.

Deligne, P., and Rapoport, M., Les schémas de modules de courbes elliptiques, in Modular Functions of One Variable II, Lecture Notes in Math., 349, Springer, Heidelberg, **1973**, 143–316.

Helgason, S., Differential Geometry, Lie Groups, and Symmetric Spaces, Academic Press, **1978**.

Kottwitz, R., Shimura varieties and λ-adic representations, in Automorphic Forms, Shimura Varieties, and L-functions, Perspectives in Math. 10, **1990**, 161–209.

Langlands, R., Some contemporary problems with origin in the Jugendtraum, Mathematical developments arising from Hilbert's problems, Amer. Math. Soc., **1976**, 401-418.

Langlands, R., and Rapoport, M., Shimuravarietäten und Gerben, J. reine angew. Math. 378, **1987**, 113–220.

Milne, J., Canonical models of (mixed) Shimura varieties and automorphic vector bundles, in Automorphic Forms, Shimura Varieties, and L-functions, Perspectives in Math. 10, **1990**, pp283–414.

Milne, J., The points on a Shimura variety modulo a prime of good reduction, in The Zeta Function of Picard Modular Surfaces, (Langlands and Ramakrishnan Eds.), Les Publications CRM, Montréal, **1992**, 153-255.

Milne, J., Motives over finite fields, in Motives (Eds. U. Jannsen, S. Kleiman, J.-P. Serre), Proc. Symp. Pure Math., AMS, 55, Part 1, pp401–459, **1994a**.

Milne, J., Shimura varieties and motives, in Motives (Eds. U. Jannsen, S. Kleiman, J.-P. Serre), Proc. Symp. Pure Math., AMS, 55, Part 2, pp447–523, **1994b**.

Mathematics Department, University of Michigan, Ann Arbor, MI 48109

E-mail address: jmilne@math.lsa.umich.edu