Geometric modular forms and the cohomology of torsion automorphic sheaves

Kai-Wen Lan

Abstract. In this survey article, we explain how modular forms can be defined geometrically in higher dimensions and in mixed characteristics using smooth toroidal compactifications, and how this can be useful for studying the cohomology groups of PEL-type Shimura varieties valued in torsion automorphic coefficients.

Contents

1. Background and introduction 1
2. Geometric modular forms in higher dimensions 3
3. Cohomology of torsion automorphic sheaves and vanishing theorems 19
References 22

1. Background and introduction

Let us begin with the proto-example, namely the classical theory of modular curves and modular forms. (References for facts summarized here can be found in, for example, the survey articles in [CY97] and [CSS97], where one can find perhaps the most famous application of the theory of modular forms. I personally find the experience of reading [Shi71], [Del71a], [DR73], and [KM85] most helpful.)

Consider an integer $k \geq 1$ defining the “weight”, and a congruence subgroup $\Gamma$ of $\text{SL}_2(\mathbb{Z})$ defining the “level”. Let $M_k(\Gamma)$ (resp. $S_k(\Gamma)$) be the space of holomorphic modular (resp. cusp) forms of weight $k$. Let $Y_{\Gamma,\mathbb{C}} = \Gamma \backslash \mathbb{H}$ (“open modular curve”) be the quotient of the Poincaré upper half plane $\mathbb{H}$, and let $X_{\Gamma,\mathbb{C}}$ (“compactified modular curve”) be the compact Riemann surface containing $Y_{\Gamma,\mathbb{C}}$ (by adding the “cusps”). (More details will be given in (2.1).

When $k \geq 2$, there is the so-called Eichler-Shimura isomorphism

\begin{equation}
M_k(\Gamma) \oplus S_k(\Gamma)^c \cong H^1(Y_{\Gamma,\mathbb{C}}, \text{Sym}^{k-2}(\mathbb{C}^{\otimes 2})),
\end{equation}

where the superscript $c$ means the complex conjugation, and where $\text{Sym}^{k-2}(\mathbb{C}^{\otimes 2})$ is the local system attached to the representation $\text{Sym}^{k-2}(\mathbb{C}^{\otimes 2})$ of $\text{GL}_2$.

2010 Mathematics Subject Classification. Primary 11G18; Secondary 14F17, 14F30, 11F75.

For the official published version, see Fifth International Congress of Chinese Mathematicians (ICCM 2010), AMS/IP Studies in Advanced Mathematics, Volume 51, 2011.
By Serre duality (realized by integration as a $\mathbb{C}$-anti-linear isomorphism), we may rewrite (1.1) as

\[(1.2) \quad H^0(X_{\Gamma, \mathbb{C}}, \omega_C \otimes k) \oplus H^1(X_{\Gamma, \mathbb{C}}, \omega_C^{\otimes (-k+2)}) \cong H^1(Y_{\Gamma, \mathbb{C}}, \text{Sym}^{k-2}(\mathbb{C}^{\otimes 2})),\]

where $\omega_C$ is a (holomorphic) line bundle over $X_{\Gamma, \mathbb{C}}$ canonically extending the familiar Hodge line bundle $\omega_C$ over $Y_{\Gamma, \mathbb{C}}$, such that we can define geometrically

\[M_k(\Gamma) := H^0(X_{\Gamma, \mathbb{C}}, \omega_C \otimes k).\]

Then the Eichler-Shimura isomorphism can be reinterpreted as a degeneracy statement in (mixed) Hodge theory. (Note that here $X_{\Gamma, \mathbb{C}}$ is compact and $Y_{\Gamma, \mathbb{C}}$ is non-compact. The use of nice compactifications is standard in mixed Hodge theory. See [Del71b] and [PS08].)

By the Betti-étale comparison, the right-hand side of (1.2) can be replaced with the $(\ell$-adic) étale cohomology (after making the coefficient $\ell$-adic using some fixed choice of an isomorphism $\mathbb{C} \cong \mathbb{Q}_{\ell}^{ac}$, where $\mathbb{Q}_{\ell}^{ac}$ is a fixed algebraic closure of the $\ell$-adic numbers $\mathbb{Q}_{\ell}$), carrying interesting actions of both the Hecke algebra and the absolute Galois group (of the field of definition of $X_{\Gamma, \mathbb{C}}$). This leads to, among other things, the important association of Galois representations with (Hecke) eigenforms.

On the other hand, while one can ask many interesting arithmetic questions about $M_k(\Gamma)$ and $S_k(\Gamma)$, they do not carry interesting Galois actions. In fact, the geometric objects $Y_{\Gamma, \mathbb{C}}$, $X_{\Gamma, \mathbb{C}}$, $\omega_C$, and $\omega_C$ all have very natural models over the integers, defined by moduli problems of elliptic curves with level structures, and by their degenerations. Using these integral models, people have been able to define and study the modular forms algebro-geometrically in mixed characteristics, and study interesting phenomena such as congruences among modular forms of very different natures (e.g., among cusp forms and Eisenstein series).

Now the question is what we can do in higher dimensions.

First of all, over $\mathbb{C}$, the modular curves are generalized by the so-called Shimura varieties, together with a theory of compactifications and automorphic bundles. (The group $\text{GL}_2$ is replaced with larger reductive algebraic groups, whose irreducible representations are used in the definition of such automorphic bundles.) The Eichler-Shimura isomorphism (1.2) has precise analogues given by the dual BGG spectral sequences of Faltings, with degeneracy due to mixed Hodge theory, relating modular forms on the left-hand sides to the group cohomology on the right-hand sides. Thanks to the recent advances in trace formula and related techniques, it is probably fair to say that people have made more progresses on the side of group cohomology (with coefficients over $\mathbb{C}$) than on the side of modular forms.

On the other hand, if we consider group cohomology with torsion coefficients, then known transcendental methods no longer apply, except that the (torsion version of) $p$-adic Hodge theory allow us to compare such group cohomology with log crystalline and log de Rham cohomology in mixed characteristics $(0, p)$, and that we still have some analogues of the dual BGG spectral sequences, from modular forms in mixed characteristics on the left-hand side to the log de Rham cohomology on the right-hand side. Now we might hope to answer questions about group cohomology with torsion coefficients by studying modular forms or related geometric objects.

In what follows, we will explain the following two topics in more detail:

1. Firstly, we will explain how modular forms can be defined and studied in higher dimensions and in good mixed characteristics using good integral models of compactifications and automorphic bundles.
(2) Secondly, we will explain some effective conditions for the spaces of such modular forms to enjoy nice properties (such as vanishing or freeness over coefficients rings at particular cohomology degrees), and deduce corresponding results for group cohomology with torsion coefficients. (This is a joint work with Junecue Suh.)

We will provide references of both original and introductory natures. (For references to [Lan08], although we will use the numbering in the original version, the reader is advised to consult the errata and revision for corrections of typos and minor mistakes, and for improved exposition.)

2. Geometric modular forms in higher dimensions

2.1. Review of the theory in dimension one. Let us begin with the analytic definition of modular forms of one variable. (For more information, see [Shi71] Ch. 2.) The group $GL_2(\mathbb{R}) := \{ \gamma \in GL_2(\mathbb{R}) : \det(\gamma) > 0 \}$ acts on the Poincaré upper-half plane $\mathcal{H} := \{ z \in \mathbb{C} : \text{Im}(z) > 0 \}$ by the familiar Möbius transformation $z \mapsto \gamma z := \frac{az+b}{cz+d}$, defined for any $z \in \mathcal{H}$ and $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{R})$. Let $k \geq 1$ be an integer, and let $\Gamma$ be a congruence subgroup (namely, defined by congruence conditions) of $SL_2(\mathbb{Z})$.

**Definition 2.1.** A (holomorphic) **modular form** of weight $k \geq 1$ and level $\Gamma$ is a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ satisfying the following two conditions:

1. (automorphy condition) For any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, we have the functional equation $f(\gamma z) = (cz+d)^{-k} f(z)$.

2. (growth condition) For any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, the function $(cz+d)^{-k} f(\gamma z)$ stays bounded as $\text{Im}(z) \to \infty$. (If $(cz+d)^{-k} f(\gamma z) \to 0$ as $\text{Im}(z) \to \infty$, we say that $f$ is a **cusp form**.)

We shall denote by $M_k(\Gamma)$ (resp. $S_k(\Gamma)$) the $\mathbb{C}$-vector space of modular forms (resp. cusp forms) of weight $k$ and level $\Gamma$.

For many applications of modular forms to number theory, it is desirable to answer the following fundamental questions:

**Question 2.2.** Can we define $M_k(\Gamma; R)$ and $S_k(\Gamma; R)$ (i.e., modular forms and cusp forms “over $R$”) for rings $R$ other than $\mathbb{C}$? Even better, can we make the definition functorial and compatible with arbitrary flat base changes in $R$?

For example, we would like to define $M_k(\Gamma; “\mathbb{Z}”) \text{ and } M_k(\Gamma; “\mathbb{F}_p”),$ where “$\mathbb{Z}$” stands for some localization of the ring of integers in some number field at a prescribed set of primes, and where “$\mathbb{F}_p$” is a residue field of “$\mathbb{Z}$” of characteristic $p > 0$.

Then we want $M_k(\Gamma) \cong M_k(\Gamma; \mathbb{C}) \cong M_k(\Gamma; “\mathbb{Z}”) \otimes_{\mathbb{Z}} \mathbb{C}$ and $S_k(\Gamma) \cong S_k(\Gamma; \mathbb{C}) \cong S_k(\Gamma; “\mathbb{Z}”) \otimes_{\mathbb{Z}} \mathbb{C}$, with the reduction maps $M_k(\Gamma; “\mathbb{Z}”) \to M_k(\Gamma; “\mathbb{F}_p”) \text{ for each } p$.

Since modular forms are a priori defined by transcendental conditions, it is not clear how these questions should be answered. Thus the first step would be to revise the definition of modular forms, by introducing the so-called **modular curves**.

For each point $z \in \mathcal{H}$, we can define a lattice $L_z := \mathbb{Z}z + \mathbb{Z}$ in $\mathbb{C}$, and an elliptic curve $E_z := \mathbb{C}/L_z$. By varying $z$, we obtain a holomorphic family $E \to \mathcal{H}$ of elliptic curves. If $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, then $L_z = \mathbb{Z}z + \mathbb{Z} = \mathbb{Z}(az + b) + \mathbb{Z}(cz + d) = (cz+d)L_{\gamma z}$, and hence the multiplication $(cz+d)^{-1} : \mathbb{C} \to \mathbb{C}$ defines an isomorphism $E_z = \mathbb{C}/L_z \sim E_{\gamma z} = \mathbb{C}/L_{\gamma z}$. This shows that:
(1) If $\Gamma$ is a torsion-free congruence subgroup of $\text{SL}_2(\mathbb{Z})$, then $E \rightarrow \mathcal{H}$ descends to a holomorphic family $E_{\Gamma,C} \rightarrow Y_{\Gamma,C} := \Gamma \backslash \mathcal{H}$ of elliptic curves.

(2) If $\omega_{\Gamma,C} := \text{Lie}_{E_{\Gamma,C}/Y_{\Gamma,C}}^{\vee}$ is the relative cotangent bundle along the identity section, then sections of $\omega_{\Gamma,C}^{\otimes k}$ can be represented by holomorphic functions $f : \mathcal{H} \rightarrow \mathbb{C}$ satisfying $f(\gamma z) = (cz + d)^k f(z)$ for any $z \in \mathcal{H}$ and $\gamma \in \Gamma$.

Therefore, we see that the automorphy condition in Definition 2.1 can be redefined geometrically. To take care of the growth condition, let $X_{\Gamma,C} := \Gamma \backslash \mathbb{H}$, where $\mathbb{H} = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ (as subsets of $\mathbb{P}^1(\mathbb{C})$, with natural actions of $\text{GL}_2(\mathbb{R})_+$ induced by the canonical action of $\text{GL}_2(\mathbb{C})$ on $\mathbb{P}^1(\mathbb{C})$) is given a topology such that the quotient $X_{\Gamma,C}$ has the structure of a compact Riemann surface, and such that $\omega_{\Gamma,C}$ extends naturally to a line bundle $\omega_{\Gamma,C}$ over $X_{\Gamma,C}$. We shall call $Y_{\Gamma,C}$ (resp. $X_{\Gamma,C}$) the modular curve (resp. compactified modular curve) of level $\Gamma$. By its very construction, $Y_{\Gamma,C}$ parameterizes isomorphism classes of pairs $(E, \alpha)$ over $\mathbb{C}$, where $E$ is an elliptic curve over $\mathbb{C}$, and $\alpha$ is a level $\Gamma$ structure, namely a $\Gamma$-orbit of isomorphisms $\alpha : \mathbb{Z}^{\oplus 2} \cong H_1(E, \mathbb{Z})$. Then we have canonical isomorphisms $M_k(\Gamma; \mathbb{C}) \cong H^0(X_{\Gamma,C}, \omega_{\Gamma,C}^{\otimes k})$ and $S_k(\Gamma; \mathbb{C}) \cong H^0(X_{\Gamma,C}, \omega_{\Gamma,C}^{\otimes k}(\infty))$, where $(-\infty)$ means vanishing at the cusps $X_{\Gamma,C} - Y_{\Gamma,C}$.

In [DR73], Deligne and Rapoport defined the moduli problem of elliptic curves with level $\Gamma$ structures over “$\mathbb{Z}$”, which in this case can be $\mathbb{Z}[\frac{1}{N}]$ for any integer $N$ such that $\Gamma$ contains $\Gamma(N) := \ker(\text{SL}_2(\mathbb{Z}) \rightarrow \text{SL}_2(\mathbb{Z}/N\mathbb{Z}))$, and showed that the moduli problem is representable by a scheme $Y_{\Gamma'}$ (over “$\mathbb{Z}$”). (Later in [KM85], by introducing the so-called Drinfeld level structures, Katz and Mazur could ensure that “$\mathbb{Z}$” is indeed $\mathbb{Z}$.) Moreover, $Y_{\Gamma'}$ has a natural compactification $X_{\Gamma'}$ (over “$\mathbb{Z}$”) given by the moduli of generalized elliptic curves (which are certain degenerations of elliptic curves with ordinary double points as singularities) with additional structures. The universal elliptic curve $E_{\Gamma} \rightarrow Y_{\Gamma'}$ extends to a smooth group scheme $E_{\Gamma,C}^{\text{ext}} \rightarrow X_{\Gamma'}$ given by the smooth part of the universal generalized elliptic curve, and hence the line bundle $\omega_{\Gamma} := \text{Lie}_{E_{\Gamma,C}^{\text{ext}}/X_{\Gamma'}}^{\vee}$ over $Y_{\Gamma'}$ extends to the line bundle $\omega_{X_{\Gamma,C}^{\text{ext}}/X_{\Gamma'}}$.

According to [DR73] Ch. VII, §4, the base extensions of the geometric objects $Y_{\Gamma}, X_{\Gamma}, \omega_{\Gamma},$ and $\omega_{X_{\Gamma,C}}$ from “$\mathbb{Z}$” to $\mathbb{C}$ are canonically isomorphic to $Y_{\Gamma,C}$, $X_{\Gamma,C}$, $\omega_{\Gamma,C}$, and $\omega_{X_{\Gamma,C}}$, respectively, which induce canonical isomorphisms $M_k(\Gamma; \mathbb{C}) \cong H^0(X_{\Gamma,C}, \omega_{\Gamma,C}^{\otimes k})$ and $S_k(\Gamma; \mathbb{C}) \cong H^0(X_{\Gamma,C}, \omega_{\Gamma,C}^{\otimes k}(\infty))$.

For any “$\mathbb{Z}$”-algebra $R$, let us denote the base extensions to $R$ by subscripts (which is justified when $R = \mathbb{C}$ thanks to the previous paragraph). Then we can define modular forms and cusp forms of weight $k$ and level $\Gamma$ over $R$ by

$$M_k(\Gamma, R) := H^0(X_{\Gamma,R}, \omega_{\Gamma,R}^{\otimes k})$$

and

$$S_k(\Gamma, R) := H^0(X_{\Gamma,R}, \omega_{\Gamma,R}^{\otimes k}(\infty)).$$

Moreover, for any ring homomorphism $R \rightarrow R'$, we have the desired functoriality $M_k(\Gamma, R) \rightarrow M_k(\Gamma, R')$ and $S_k(\Gamma, R) \rightarrow S_k(\Gamma, R')$, and for flat ring extensions $R \rightarrow R'$, we have $M_k(\Gamma, R) \otimes R' \rightarrow M_k(\Gamma, R')$ and $S_k(\Gamma, R) \otimes R' \rightarrow S_k(\Gamma, R')$.

Hence Question 2.2 has been answered.

2.2. Summary of key ingredients. Before moving on, let us summarize the key ingredients in the geometric definition of modular forms:
(1) The analytic definition of $Y_{\Gamma, C}$ and $X_{\Gamma, C}$ over $\mathbb{C}$.
(2) The integral model $Y_{\Gamma}$ of $Y_{\Gamma, C}$ defined using the moduli of elliptic curves with level $\Gamma$ structures. (The precise ring of definition “$Z$” depends on the definition of level $\Gamma$ structures.)
(3) The integral model $X_{\Gamma}$ of $X_{\Gamma, C}$ defined using degenerations of elliptic curves into curves with ordinary double points as singularities.
(4) The definition of $\omega_{\Gamma, C}$ using the holomorphic family $E_{\Gamma, C} \to Y_{\Gamma, C}$.
(5) The definition of $\omega_{\Gamma}$ using the universal elliptic curve $E_{\Gamma} \to Y_{\Gamma}$.
(6) The definition of $\Xi_{\Gamma}$ using the extended object $E_{\Gamma}^{\text{ext}} \to X_{\Gamma}$. (This uses implicitly some compactified object $E_{\Gamma} \to X_{\Gamma}$.)
(7) The definitions of $M_k(\Gamma; R)$ and $S_k(\Gamma; R)$ (in (2.3) and (2.4)) using $H^0$ of tensor powers of $\Xi_{\Gamma}$ (with vanishing along the cusps in the case of $S_k$).

In the following subsections, we will discuss generalizations of these one by one.

### 2.3. Shimura varieties and their compactifications over the complex numbers

The modular curve $Y_{\Gamma, C}$ over $\mathbb{C}$ is generalized by the Shimura varieties.

One starts with a Shimura datum $(G, X)$, where $G$ is a reductive algebraic group over $\mathbb{Q}$ and $X$ is a finite union of Hermitian symmetric domains carrying an action of $G(\mathbb{R})$ and satisfying some conditions. (We will not make these conditions precise because we will not need them.) Let $H$ be an open compact subgroup of $G(\mathbb{A}^\infty)$. (Here $\mathbb{A}^\infty = \mathbb{Z} \otimes \mathbb{Q}$, where $\mathbb{Z} = \lim\limits_{\leftarrow} \mathbb{Z}/N\mathbb{Z}$.) Then the Shimura variety at level $H$ attached to this datum is the double quotient

\[ \text{Sh}_{H, C} := G(\mathbb{Q}) \backslash (X \times G(\mathbb{A}^\infty)/H). \]

A priori, this is only a complex analytic space, which is a finite union of quotients of $X$ by arithmetic subgroups of $G(\mathbb{Q})$. (The finiteness of the union is due to the finiteness of $G(\mathbb{Q}) \backslash G(\mathbb{A}^\infty)/H$, by [Bor63] Thm. 5.1.)

According to Satake and Baily-Borel [BB66 10.11], there is a minimal compactification $\text{Sh}^{\text{min}}_{H, C}$ of $\text{Sh}_{H, C}$, which is canonical and given by a finite union of complex normal projective varieties. According to Mumford and his coworkers [AMRT75], there is a collection of (non-canonical) nonsingular and/or projective compactifications $\text{Sh}^{\text{tor}}_{H, C}$ in the category of complex algebraic spaces, called toroidal compactifications, parameterized by certain combinatorial data of compatible choices of cone decompositions, which resolve the (generally very complicated) singularities of the minimal compactification $\text{Sh}^{\text{min}}_{H, C}$. Thus, $\text{Sh}^{\text{tor}}_{H, C}$ is a nice geometric object with nice compactifications (in the category of algebraic varieties, if we only consider projective toroidal compactifications). This is important for defining and studying, for example, the Hodge structure on the de Rham cohomology. (Later in mixed characteristics even the very existence of nice compactifications is unclear, and we can say very little about relations among different compactifications, because we do not have analogues of Hironaka’s embedded resolution of singularities [Hir64a, Hir64b].)

Moreover, by the theory of canonical models, $\text{Sh}^{\text{min}}_{H, C}$ has a canonical model $\text{Sh}_H$ over a number field $F_0$ called the reflex field. (It is then customary to also call $\text{Sh}_H$ the Shimura variety over $F_0$.) According to Harris [Har89] and Pink [Pin89], $\text{Sh}^{\text{min}}_{H, C}$ and $\text{Sh}^{\text{tor}}_{H, C}$ also have canonical models $\text{Sh}^{\text{min}}_H$ and $\text{Sh}^{\text{tor}}_H$, respectively, over (the same) $F_0$. 

In the following subsections, we will discuss generalizations of these one by one.
The upshot is that the reflex field depends on the Shimura datum \((G, X)\) but not on the level \(H\). As a result, the inductive limit of \((\ell\text{-adic}) \acute{e}tale cohomology groups of \(\mathcal{S}_{H}\) (as the level \(H\) varies) carries both the Hecke action (defined by an action of \(G(\mathbb{A}_{\infty})\) on the limit) and the Galois action (of \(\text{Gal}(\overline{\mathbb{Q}}^{\text{ac}}/\mathbb{F}_0)\), where \(\overline{\mathbb{Q}}^{\text{ac}}\) is the algebraic closure of \(\mathbb{Q}\) in \(\mathbb{C}\)), and these commute with each other. (Similar statements can be made about the intersection cohomology defined using the minimal compactifications.) This is why people look for relations among automorphic representations and Galois representations in the cohomology of Shimura varieties.

The theory of Shimura varieties has its origin in, as suggested by its name, the works of Shimura [Shi02]. (See also [Shi71] and [Shi98].) The prevailing formulation is due to Deligne, whose papers [Del71a] and [Del79] are the canonical references in this subject. A helpful introduction is [Mil05]. (For the purpose of this article, we will not need the theory of canonical models.)

The theory of compactifications of arithmetic quotients of Riemannian or Hermitian symmetric spaces is an important subject by itself. Apart from the minimal and toroidal compactifications, there are many other compactifications as topological spaces, often far beyond the category of algebraic varieties. (See Borel-Ji [BJ06] for a nice overall treatment on the subject.) We emphasize the minimal and toroidal compactifications because they are the only ones that turned out to have nice models over the reflex fields, and even over integers. (See §2.5 below.)

2.4. Integral models of Shimura varieties. Not all Shimura varieties are known to have nice integral models. (Here integral means defined over \(\text{Spec}(\mathcal{O}_{F_0}[\frac{1}{N}])\) for some explicit integer \(N\).) Although by abstract nonsense algebraic varieties defined over number fields have some integral model over the ring of integers, such abstract models are not useful when we need to have precise control on the local structures. To the best of our knowledge, all useful integral models of Shimura varieties involve (either directly or indirectly) moduli spaces of abelian varieties with certain additional structures.

Among Shimura varieties that do possess useful integral models, the PEL-type Shimura varieties are those that admit interpretations as moduli spaces of abelian varieties with the PEL structures, namely the polarizations ("P"), the endomorphism structures ("E"), and the level structures ("L"). (This certainly has its origin in works of Shimura.) One reason for these structures to be useful is because it is easy to define moduli problems with these structures over very general rings. Definitions of moduli problems over "\(\mathbb{Z}\)" in the good reduction case (namely cases where we know the moduli problem is smooth over a base scheme with prescribed residue characteristics) can be found in Zink [Zin82] §1, Langlands-Rapoport [LR87] §6, and in utmost generality in Kottwitz [Kot92] §5, following earlier ideas of Grothendieck and Deligne (using objects up to isogeny). (I recommend reading the definition given by Kottwitz.)

The bad reduction case (namely cases where the model can have singularities in fibers of positive characteristics) can be very difficult in general, and the justification for the theory (i.e., “How bad can we allow the models to be?”) may depend heavily on the applications. People might have the impression from low dimensions and a few very successful examples in arbitrary dimensions (such as [HT01]) that there should be some optimal theory, but we cannot rule out the possibility that different applications might deserve very different treatments. In what follows, we will leave the bad reduction case to specialists and focus mainly on the good reduction case.
For the purpose of compactifications later, we will prefer to define the moduli problems by parameterizing objects up to isomorphism. Since many readers today might be more familiar with the definition by isogeny classes, and since the definition by isomorphism classes is less canonical, we shall supply more detail. (The definition by isomorphism classes is certainly not new. This is the approach taken in many famous special cases, such as \cite{DR73, KM85, Rap78, DP94, MF94, FC90, Lar88, Lar92} and many later works.)

Let \((\mathcal{O}, \star, L, \langle \cdot, \cdot \rangle, h_0)\) be an integral PEL datum in the following sense:

1. \(\mathcal{O}\) is an order in a (nonzero) semisimple algebra, finite-dimensional over \(\mathbb{Q}\), together with a positive involution \(\star\).
2. \(L\) is an \(\mathcal{O}\)-lattice, namely a \(\mathbb{Z}\)-lattice with the structure of an \(\mathcal{O}\)-module.
3. \(\langle \cdot, \cdot \rangle : L \times L \to \mathbb{Z}(1)\) is an alternating pairing satisfying \(\langle bx, y \rangle = \langle x, b^*y \rangle\) for any \(x, y \in L\) and \(b \in \mathcal{O}\).
4. \(h_0 : \mathbb{C} \to \text{End}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes \mathbb{R})\) is an \(\mathbb{R}\)-algebra homomorphism satisfying:
   a) For any \(z \in \mathbb{C}\) and \(x, y \in L \otimes \mathbb{R}\), we have \(\langle h_0(z)x, y \rangle = \langle x, h_0(z^c)y \rangle\), where \(\mathbb{C} \to \mathbb{C} : z \mapsto z^c\) is the complex conjugation.
   b) For any choice of \(\sqrt{-1}\) in \(\mathbb{C}\), the pairing \(-\sqrt{-1} \langle \cdot, h_0(\sqrt{-1}\cdot) \rangle : (L \otimes \mathbb{R}) \times (L \otimes \mathbb{R}) \to \mathbb{R}\) is symmetric and positive definite. (This last condition forces \(\langle \cdot, \cdot \rangle\) to be nondegenerate.)

(In \cite{Lan08} Def. 1.2.1.3, \(h_0\) was denoted by \(h\).) The tuple \((\mathcal{O}, \star, L, \langle \cdot, \cdot \rangle, h_0)\) then gives us an integral version of the \((B, \star, V, \langle \cdot, \cdot \rangle, h_0)\) in \cite{Kot92} and related works. We shall denote the center of \(\mathcal{O} \otimes \mathbb{Q}\) by \(F\). (Then \(F\) is a product of number fields.)

**Definition 2.6** (cf. \cite{Lan08} Def. 1.2.1.5). Let \(\mathcal{O}\) and \((L, \langle \cdot, \cdot \rangle)\) be given as above. Then we define for any \(\mathbb{Z}\)-algebra \(R\)

\[
G(R) := \left\{ (g, r) \in \text{GL}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes R) \times \text{G}_m(R) : \begin{cases} (gx, gy) = r(x, y), & \forall x, y \in L \otimes \mathbb{R} \\ (g, r) \in \text{GL}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes R) \times \text{G}_m(R) \end{cases} \right\}.
\]

The assignment is functorial in \(R\) and defines a group functor \(G\) over \(\text{Spec}(\mathbb{Z})\).

The homomorphism \(h_0 : \mathbb{C} \to \text{End}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes \mathbb{R})\) defines a Hodge structure of weight \(-1\) on \(L\), with Hodge decomposition

\[(2.7) \quad L \otimes \mathbb{C} = V_0 \oplus V_0^c,\]

such that \(h_0(z)\) acts as \(1 \otimes z\) on \(V_0\), and as \(1 \otimes z^c\) on \(V_0^c\). Let \(F_0\) be the reflex field, namely the field of definition of the isomorphism class of the \(\mathcal{O} \otimes \mathbb{C}\)-module \(V_0\).

(See \cite{Kot92} §5 and \cite{Lan08} Def. 1.2.5.4. Note that this does not mean there exists an \(\mathcal{O} \otimes F_0\)-module whose base extension from \(F_0\) to \(\mathbb{C}\) is isomorphic to \(V_0\).)

We shall denote the ring of integers in \(F\) (resp. \(F_0\)) by \(\mathcal{O}_F\) (resp. \(\mathcal{O}_{F_0}\)).

We say that a rational prime number \(p > 0\) is good if it satisfies the following conditions (cf. \cite{Kot92} §5 and \cite{Lan08} Def. 1.4.1.1):

1. \(p\) is unramified in \(\mathcal{O}\) (as in \cite{Lan08} Def. 1.1.1.14)).
2. \(p \neq 2\) if \(\mathcal{O} \otimes \mathbb{Q}\) involves simple factors of type D (as in \cite{Lan08} Def. 1.2.1.15)).
3. The pairing \(\langle \cdot, \cdot \rangle\) is perfect after base change to \(\mathbb{Z}_p\).
Let us fix a choice of a good prime $p > 0$.

Let $\mathcal{H}$ be a neat open compact subgroup of $G(\hat{\mathbb{Z}}^p)$. (See [Pin89] 0.6 or [Lan08] Def. 1.4.1.8 for the definition of neatness. The assumption of neatness corresponds to the assumption that the congruence subgroup $\Gamma$ is torsion-free when defining modular curves of level $\Gamma$.)

By [Lan08] Def. 1.4.1.4 (with $\square = \{ p \}$ there), the data of $(L, (\cdot, \cdot), h_0)$ and $\mathcal{H}$ define a moduli problem $M_\mathcal{H}$ over $S_0 = \text{Spec}(\mathcal{O}_{F_0,(p)})$, parameterizing tuples $(A, \lambda, i, \alpha_\mathcal{H})$ over schemes $S$ over $S_0$ of the following form:

1. $A \to S$ is an abelian scheme.
2. $\lambda : A \to A'$ is a polarization of degree prime to $p$.
3. $i : \mathcal{O} \to \text{End}_S(A)$ is an $\mathcal{O}$-endomorphism structure as in [Lan08] Def. 1.3.3.1.
4. $\text{Lie}_{\mathcal{O}/S}$ with its $\mathcal{O} \otimes \mathbb{Z}(p)$-module structure given naturally by $i$ satisfies the determinantal condition in [Lan08] Def. 1.3.4.1 given by $(L \otimes \mathbb{R}, (\cdot, \cdot), h_0)$. (The idea and the formulation are due to [Kot92] §5 and [RZ96] 3.23(a), respectively.)
5. $\alpha_\mathcal{H}$ is an (integral) level-$\mathcal{H}$ structure of $(A, \lambda, i)$ of type $(L \otimes \hat{\mathbb{Z}}^p, (\cdot, \cdot))$ as in [Lan08] Def. 1.3.7.8. (In general this condition is nontrivial even if $\mathcal{H} = G(\hat{\mathbb{Z}}^p)$ is maximal, unlike in certain famous special cases.)

(The definition can be identified with the one in [Kot92] §5 by [Lan08] Prop. 1.4.3.3.) By [Lan08] Thm. 1.4.1.12 and Cor. 7.2.3.10, $M_\mathcal{H}$ is representable by a (smooth) quasi-projective scheme over $S_0$ (under the assumption that $\mathcal{H}$ is neat). (The proof of quasi-projectivity uses results of Moret-Bailly [MB85], but does not require Mumford’s geometric invariant theory [MFK94]. It is intriguing and somewhat mysterious, but certainly no surprise, that both involve algebraic theta functions.)

Consider the (real analytic) set $X = G(\mathbb{R})h_0$ of $G(\mathbb{R})$-conjugates $h : \mathbb{C} \to \text{End}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes \mathbb{R})$ of $h_0 : \mathbb{C} \to \text{End}_{\mathcal{O} \otimes \mathbb{R}}(L \otimes \mathbb{R})$. Let $H^p := \mathcal{H}$ and $H_p := G(\mathbb{Z}_p)$ be open compact subgroups of $G(\hat{\mathbb{Z}}^p)$ and $G(\mathbb{Q}_p)$, respectively, and let $H$ be the open compact subgroup $H^pH_p$ of $G(\hat{\mathbb{Z}})$. Define $\text{Sh}_{\mathcal{H},\mathbb{C}}$ by forming the double quotient as in [Lan08] 2.5, which we also view (by abuse of language) as a finite union of quasi-projective varieties (using Baily and Borel’s theorem [BB66] 10.11).

For any $h \in X$, the real analytic torus $A_h := (L \otimes \mathbb{Z})/L$ with complex structure given by $h$ has the structure of a polarized abelian variety with endomorphism and level structures defining a $\mathbb{C}$-valued point of $M_\mathcal{H}$. More generally, any element of $X \times G(\mathbb{A}^\infty)$ defines some $\mathbb{C}$-valued point of $M_\mathcal{H}$, and two elements define the same point if they lie in the same double orbit (for $G(\mathbb{Q})$ and $H$). In other words, $\text{Sh}_{\mathcal{H},\mathbb{C}}$ can be viewed as an analytic moduli space of abelian varieties with PEL structures defined by $(\mathcal{O}, *, L, (\cdot, \cdot), h_0)$. By arguing carefully as in [Kot92] §8 and [Lana] §2, one can show that there is a canonical isomorphism between $\text{Sh}_{\mathcal{H},\mathbb{C}}$ and an open and closed subscheme of $M_\mathcal{H} \otimes \mathbb{C}$. (In general, $\text{Sh}_{\mathcal{H},\mathbb{C}}$ might not be isomorphic to the whole of $M_\mathcal{H} \otimes \mathbb{C}$, due to the so-called failure of Hasse’s principle. See for example [Kot92] §8 and [Lan08] Rem. 1.4.3.11.)
2.5. Integral models of compactifications. Just as the compactification $X_\Gamma$ of $Y_\Gamma$ (see §2.1) is constructed using moduli of generalized elliptic curves (which are certain degenerations of elliptic curves), to compactify integral models of PEL-type Shimura varieties, we would like to study the degeneration of the objects parameterized by the corresponding moduli problems.

Based on the work of Mumford [Mum72], Faltings and Chai (see [Fal85], [Cha85], and especially [FC90]) studied the theory of semi-abelian degenerations for polarized abelian varieties over noetherian normal complete adic rings satisfying certain reasonable conditions, and constructed smooth integral models of (smooth) toroidal compactifications of the Siegel modular varieties (parameterizing principally polarized abelian schemes with principal level structures, over base schemes over which the primes dividing the level are invertible).

In their construction, they did not use any moduli problem of degenerating objects. Instead, they glue boundary charts (whose construction depends on choices of cone decompositions) to the moduli problem in the ´etale topology. Such a process is feasible because the sheaves of relative log differentials can be explicitly calculated and compared over the charts. As a byproduct, they obtained integral models of the minimal compactifications of Siegel modular varieties by taking the projective spectra of certain graded algebras of (scalar-valued) algebraic automorphic forms. (Unlike in the theory over $\mathbb{C}$, we do not know any direct construction of the minimal compactifications.) The theory is generalized in Larsen’s thesis [Lar88] (see also [Lar92]) for certain Picard modular varieties.

In my thesis [Lan08], I constructed smooth integral models of (smooth) compactifications of all types of PEL-type Shimura varieties (as defined in §2.4 above), based on a generalization of the method of Faltings and Chai’s in [FC90] with a new emphasis on the degeneration of level structures. (As remarked above, the definition of level structures in general is nontrivial even if $\mathcal{H} = G(\hat{\mathbb{Z}}_p)$ is maximal.) This involves some calculation of Weil pairings that has been unnecessary in earlier works.

The case of Siegel modular varieties with bad reductions of parahoric levels at $p$ is treated in the thesis of Stroh [Str08] (see also [Str10a, Str10b]). It should be possible to treat similar parahoric levels for most symplectic and unitary cases by combining Stroh’s ideas with ours. (Maybe Stroh will carry this out.) It is less clear what one should expect in cases of deeper levels, because (without applications in mind) it is already unclear how the integral models should be defined.

There are also the important canonical compactifications constructed by Alexeev and Nakamura [AN99, Ale02] and by Olsson [Ols08], which indeed use some moduli problem of degenerating objects. However, while such compactifications are very interesting for algebraic geometers, they have no known applications to the study of automorphic representations and related topics in number theory. (Although it might seem a disadvantage that the old-fashioned toroidal compactifications are noncanonical and have to be constructed by gluing boundary charts, it is also an advantage that we have a very precise description of the boundary!)

In what follows we will focus on good reduction models of (smooth) toroidal compactifications (which will eventually be the only ones useful for the strategy in §3). Let us summarize some results in [Lan08] Ch. 6–7.

**Theorem 2.8** (see [Lan08] Thm. 6.4.1.1, 7.2.4.1, and 7.3.3.4] for more details). When $\mathcal{H}$ is neat, $M_\mathcal{H}$ admits a toroidal compactification $M^\text{tor}_\mathcal{H} = \overline{M}_\mathcal{H,\Sigma}$, a scheme
properties: Σ

called cone decompositions projective and smooth over $S$

$\exists (2.9)$

Let $\Sigma$ described in Theorem 2.8.

$(M)$

Let $\tilde{H}$

In what follows, we shall sometimes omit $\Sigma$ when the choice is clear.

There is an extended Kodaira-Spencer isomorphism from a quotient
of $\text{Lie}_{\lambda^\text{ext}}(A^\text{ext})/\text{Lie}_{\lambda^\text{ext}}(A^\text{ext})^\vee$ (by precise relations defined by $\lambda^\text{ext}$ and $i^\text{ext}$) to

$$\Omega^1_{M^\text{tor}/S_0} := \Omega^1_{M^\text{tor}/S_0}(\log D) = \Omega^1_{M^\text{tor}/S_0}[d \log D],$$

the sheaf of modules of log 1-differentials on $M^\text{tor}$ over $S_0$, with respect to the relative Cartier divisor $D$.

Let $\omega := \wedge_{\text{top}} \text{Lie}_{\lambda^\text{ext}}(A^\text{ext})/\text{Lie}_{\lambda^\text{ext}}(A^\text{ext})^\vee$.

The scheme $\text{Proj}(\oplus_{r \geq 0} \Gamma(M^\text{tor}, \omega^r \otimes r))$ is normal and projective over $S_0$, contains $M^\text{tor}$ as an open dense subscheme, and defines the minimal compactification $M^\text{min}$ of $M^\text{tor}$ (independent of the choice of $\Sigma$). Moreover, the line bundle $\omega$ descends to an ample line bundle over $M^\text{min}$.

$M^\text{tor}$ is the normalization of the blow-up of $M^\text{min}$ along a coherent sheaf of ideals $\mathcal{J}$ of $\mathcal{O}_{M^\text{min}}$ whose pullback $\mathcal{J}$ to $\mathcal{O}_{M^\text{tor}}$ is of the form $\mathcal{O}_{M^\text{tor}}(-D')$, for some relative Cartier divisor $D'$ with normal crossings on $M^\text{tor}$ such that $D'_\text{red} = D$. In particular:

$\exists r_0 > 0$ such that $\omega^r(-D')$ is ample for every $r \geq r_0$.

(Here we use [FC90] Thm. 7.3.3.4 and the assumption that $\Sigma$ is projective.)

In what follows, we shall sometimes omit $\Sigma$ when the choice is clear.

Let $\text{Sh}_H$ denote the schematic image of the canonical morphism $\text{Sh}_H \to M_H$. Let $M_{H,0}$ (resp. $M^\text{tor}_{H,\Sigma,0}$) denote the schematic closure of $\text{Sh}_H$ in $M_H$ (resp. $M^\text{tor}_{H,\Sigma}$). Then $M_{H,0}$ is smooth over $S_0$, and $M^\text{tor}_{H,\Sigma,0} \to S_0$ is proper and smooth and shares the properties of $M^\text{tor}_{H,\Sigma} \to S_0$ listed above. By abuse of notation, we denote the pullback of $D$ to $M^\text{tor}_{H,\Sigma,0}$ still by $D$. Similarly, let $M^\text{min}_{H,\Sigma,0}$ denote the schematic closure of $\text{Sh}_H$ in $M^\text{min}_{H,\Sigma}$. Then $M^\text{tor}_{H,\Sigma,0} \to M^\text{min}_{H,\Sigma,0}$ enjoys the same properties of $M^\text{tor}_{H,\Sigma} \to M^\text{min}_{H}$ described in Theorem 2.8.
Remark 2.10. Although not logically related, (2) of Theorem 2.8 asserts that our choices of cone decompositions are consistent with those called SNC in [Har89], and subsequent works such as [HZ94a], [HZ94b], and [HZ01].

In [Lana], by using comparison between spaces of analytic and algebraic theta functions, it is shown that $M_{\text{tor}}^{\text{hor}}_{\mathcal{H}, \Sigma, 0} \otimes \mathbb{C}$ is indeed a toroidal compactification $\mathcal{S}h_{\mathcal{H}, \mathcal{C}}^{\text{tor}}$ of $\mathcal{S}h_{\mathcal{H}, \mathcal{C}} \cong M_{\mathcal{H}, 0} \otimes \mathbb{C}$ (constructed by the method of [AMRT75], [Har89], and [Pin89]), and as a consequence that $M_{\mathcal{H}, 0}^{\text{min}} \otimes \mathbb{C}$ is the minimal compactification $\mathcal{S}h_{\mathcal{H}}^{\text{min}}$ (constructed by the method of [BSh66] and [Pin89]). (We have to point out that the claims in certain works that this is true because the local charts of the analytic and algebraic constructions look similar, is not justified, because the construction methods are not logically related to each other.)

Thanks to its stratification by smooth locally closed subschemes, the (smooth) toroidal compactification $M^{\text{tor}}_{\mathcal{H}, \Sigma, 0}$ (with its long list of nice properties inherited from $M^{\text{tor}}_{\mathcal{H}, \Sigma}$ in Theorem 2.8) is a nice geometric object suitable for the study of Hodge, de Rham, and crystalline cohomology. (We will see some applications in [3].) An intriguing (yet less noticed) fact is that the construction of $M^{\text{tor}}_{\mathcal{H}, \Sigma, 0}$ is also useful for compact Shimura varieties. (If a boundary stratum is empty in characteristic zero, it has to be empty in positive characteristic too.) This leads to:

Corollary 2.11. If $\mathcal{S}h_{\mathcal{H}, \mathcal{C}}$ (with its real analytic structure inherited from $X$) is compact, then $M_{\mathcal{H}, 0}$ is proper (and hence projective) over $\mathcal{S}_0$.

(It is possible to prove this using only part of Faltings and Chai [FC90], without the gluing construction of $M^{\text{tor}}_{\mathcal{H}, \Sigma}$. See [Lanb] §4.)

2.6. Automorphic bundles and canonical extensions over the complex numbers. For simplicity, let us maintain the PEL-type setup in this subsection.

Consider the subgroup $P_{0, \mathcal{C}}$ of $G_{\mathcal{C}}$ stabilizing the Hodge filtration given by the Hodge decomposition (2.7). (This $P_{0, \mathcal{C}}$ will be compatible with the $P_0$ to be defined in (2.8) below.) The Hodge decomposition (2.7) itself then induces a splitting of the Levi quotient $M_{0, \mathcal{C}}$ as a subgroup of $P_{0, \mathcal{C}}$. By varying $h$ in $X$, the varying Hodge filtration then defines an embedding of $X$ in the flag variety $G_{\mathcal{C}}/P_{0, \mathcal{C}}$. By abuse of language, we shall also identify this flag variety with its complex points $G(\mathbb{C})/P_{0}(\mathbb{C})$. Given any algebraic representation $W_{\mathcal{C}}$ of $P_{0, \mathcal{C}}$, the sheaf of holomorphic sections of

$$(2.12) \quad (G(\mathbb{C}) \times W_{\mathcal{C}})/P_{0}(\mathbb{C}) \to G(\mathbb{C})/P_{0}(\mathbb{C})$$

defines a $G(\mathbb{C})$-equivariant holomorphic vector bundle on $(G(\mathbb{C})/P_{0}(\mathbb{C}))$, whose restriction to $X$ descends to its (smooth) arithmetic quotients and defines a holomorphic vector bundle $\text{O}_{\mathcal{C}}$ on $\mathcal{S}h_{\mathcal{H}, \mathcal{C}}$. The (holomorphic) vector bundles like $\text{O}_{\mathcal{C}}$ are generalizations of the line bundles $\omega_{\mathcal{C}}^{\otimes k}$ in [1].

On the other hand, given any algebraic representation $V_{\mathcal{C}}$ of $G_{\mathcal{C}} := G \otimes \mathbb{C}$, one can consider the sheaf of locally constant (resp. holomorphic) sections of

$$(2.13) \quad G(\mathbb{Q})\backslash((X \times V_{\mathcal{C}}) \times G(\mathbb{A}^\infty)/H) \to \mathcal{S}h_{\mathcal{H}, \mathcal{C}} = G(\mathbb{Q})\backslash(X \times G(\mathbb{A}^\infty)/H),$$

which we denote by $\mathcal{B}V_{\mathcal{C}}$ (resp. $V_{\mathcal{T}}$). (Here “$\mathcal{B}$” means the Betti version of the automorphic sheaves we consider.) The cohomology of $\mathcal{B}V_{\mathcal{C}}$ can be computed by the
de Rham cohomology (and hence can be computed using the Lie algebra cohomology, see [BW00 Ch. VII]). Moreover, $V_C$ is naturally equipped with an integrable connection $\nabla : V_C \rightarrow V_C \otimes \Omega^1_{\omega H,C}$ (inducing the algebraic de Rham complex $(V_C \otimes \Omega^i_{\omega H,C}, \nabla)$), such that $bV_C$ is canonically isomorphic to the sheaf of horizontal sections of $(V_C, \nabla)$, and so that we have

$$H^i(\Omega^i_{\omega H,C}, bV_C) \cong H^i_{dR}(\Omega^i_{\omega H,C}, V_C) := H^i(\Omega^i_{\omega H,C}, (V_C \otimes \Omega^i_{\omega H,C}, \nabla)).$$

(The last term is a hypercohomology of complexes.) The sheaves $bV_C$ are generalizations of the sheaves $\text{Sym}^{k-2}(\mathbb{C}^2)$ in $\{1\}$. However, for obtaining mixed Hodge structures on $H^i_{dR}(\Omega^i_{\omega H,C}, V_C)$, so that an analogue of (1.2) is possible, it is desirable to introduce the so-called canonical extensions of both bundles like $V_C$ and $\omega H$ over toroidal compactifications of $\Omega^i_{\omega H,C}$. (See [Mum77, FC90 Ch. VI], [Har89], [Har90], and [Mil90]. See in particular [Har89 Thm. 4.2] for the relation between canonical extensions and the notion of regular singularities of algebraic differential equations in [Del70] and [Kat71].) We shall denote canonical extensions by the superscript “can”.

For simplicity, assume that $G_C$ has no type D factors (so that it is connected, and so that its irreducible representations are uniquely determined by their highest weights; otherwise, we need to group together several highest weights sharing the same irreducible representation, as in [LSb, LSc]). By choosing a suitable common maximal torus of $G_C$ and $M_{0,C}$, and by choosing suitable Borel subgroups of $G_C$ and $M_{0,C}$, we can compare the weights of $G_C$ and $M_{0,C}$ and assume that the dominant weights $X^+_w(G_C)$ for $G_C$ form a subset of the dominant weights $X^+_w(M_{0,C})$ for $M_{0,C}$. Let $W_{G_C}$ (resp. $W_{M_{0,C}}$) denote the Weyl group of $G_C$ (resp. $M_{0,C}$), and let $W_{M_{0,C}} := \{ w \in W_{G_C} : w(X^+_w(G_C)) \subset X^+_w(M_{0,C}) \}$.

Let $\rho$ be the half sum of positive roots of $G_C$, and let $w \cdot \mu = w(\mu + \rho) - \rho$ be the familiar dot action (respecting the infinitesimal weight of Harish-Chandra). Let $V_{\mu,C}$ (resp. $W_{\mu,C}$) denote the irreducible representation of $G_C$ (resp. $M_{0,C}$) of highest weight $\mu$ (resp. $\nu$).

In [Fal83], using older ideas of Bernstein-Gelfand-Gelfand [BGG75], Faltings showed that there is the dual BGG spectral sequence, which is of the form (2.14)

$$E^2_{a,i-a} := \bigoplus_{w \in W_{M_{0,C}}} H^i_{dR}(\Omega^a_{\omega H,C}, \text{Gr}^\omega_w((\mathrm{Sym}^\nu(\omega H,C))^\text{can})) \Rightarrow H^i_{dR}(\Omega^a_{\omega H,C}, \omega H,C).$$

The degeneracy of (2.14) is shown in certain special cases in [Fal83] (the anisotropic case) and [FC90] (the Siegel case), and in general in [HZ01 Cor. 4.2.3], all using some kinds of mixed Hodge theory.

We shall consider cohomology classes in $H^i(\Omega^a_{\omega H,C}, ((\omega H,C)^\text{can}))$, where $i \in \mathbb{Z}$ and $\nu$ is any dominant weight of $M_{0(C)}$, modular forms or rather automorphic forms of weight $\nu$ and level $H$ (for our PEL datum). (If need be, we can also specify the cohomology degree $i$ we are using.) Then (2.14) (with its degeneracy) asserts that there is a filtration on $H^i_{dR}(\Omega^a_{\omega H,C}, \omega H,C)$ with graded pieces given by automorphic forms of some specific weights (and level $H$). This is the desired generalization of (1.1), or rather (1.2). (We encourage the reader to work out the details and understand why this generalizes (1.2), and what this does generalize.)
2.7. Automorphic bundles in mixed characteristics. In mixed characteristics, the integral models of Shimura varieties are not defined by quotients of symmetric spaces, and hence the definition of automorphic bundles will require more algebraic considerations. The idea we learned from Milne [Mi90] is that it is better to consider the so-called principal bundles, which are geometric objects which take care of the needed twisting without having to go through the quotients as in (2.13) and (2.12) for each individual representations. (The idea is then successfully applied to the mixed characteristics setup in, for example, [MT02].)

Let \( p \) be the good prime chosen in (2.4)

**Lemma 2.15.** There exists a finite extension \( F_0' \) of \( F_0 \) in \( \mathbb{C} \), unramified at \( p \), together with an \( \mathcal{O} \otimes \mathcal{O}_{F_0'(p)} \)-module \( L_0 \) such that \( L_0 \otimes \mathbb{C} \cong V_0 \) as \( \mathcal{O} \otimes \mathbb{C} \)-modules.

See [Lan08] Lem. 1.2.5.9 in the revision for a proof. For each fixed \( F_0' \), the choice of \( L_0 \) is unique up to isomorphism.

Let us denote by \( \langle \cdot , \cdot \rangle_{\text{can.}} : (L_0 \oplus L_0^\vee(1)) \times (L_0 \oplus L_0^\vee(1)) \to \mathcal{O}_{F_0'(p)}(1) \) (cf. [Lan08] Lem. 1.1.4.16] the alternating pairing \( \langle x_1 , f_1 , (x_2 , f_2)_{\text{can.}} := f_2(x_1) - f_1(x_2) \). The natural right action of \( \mathcal{O} \) on \( L_0^\vee(1) \) defines a natural left action of \( \mathcal{O} \) by composition with the involution \( * : \mathcal{O} \to \mathcal{O} \). Then (2.7) canonically induces an isomorphism \( L_0^\vee(1) \otimes \mathbb{C} \cong V_0^\vee \) of \( \mathcal{O} \otimes \mathbb{C} \)-modules.

**Definition 2.16.** For any \( \mathcal{O}_{F_0'(p)} \)-algebra \( R \), set

\[
G_0(R) := \left\{ (g,r) \in \text{GL}_{\mathcal{O}_{F_0'(p)}}(L_0 \oplus L_0^\vee(1)) \otimes R \times \text{G}_m(R) : (gx,gy)_{\text{can.}} = r(x,y)_{\text{can.}}, \forall x,y \in (L_0 \oplus L_0^\vee(1)) \otimes \mathcal{O}_{F_0'(p)}(1) \right\},
\]

\[
P_0(R) := \left\{ (g,r) \in G_0(R) : g(L_0^\vee(1)) \otimes \mathcal{O}_{F_0'(p)} \cong L_0^\vee(1) \otimes \mathcal{O}_{F_0'(p)} \right\},
\]

\[
M_0(R) := \text{GL}_{\mathcal{O}_{F_0'(p)}}(L_0^\vee(1)) \otimes \mathcal{O}_{F_0'(p)} \times \text{G}_m(R),
\]

where we view \( M_0(R) \) canonically as a quotient of \( P_0(R) \) by \( P_0(R) \to M_0(R) : (g,r) \mapsto (g(\mathcal{L}_0^\vee(1)) \otimes R, r) \).

The assignments are functorial in \( R \), and define group functors \( G_0, P_0, \) and \( M_0 \) over \( \text{Spec}(\mathcal{O}_{F_0'(p)}) \).

By [Lan08] Prop. 1.1.1.17, Cor. 1.2.5.7, and Cor. 1.2.3.10], there exists a discrete valuation ring \( R_1 \) over \( \mathcal{O}_{F_0'(p)} \) satisfying the following conditions:

1. The maximal ideal of \( R_1 \) is generated by \( p \), and the residue field \( k_1 \) of \( R_1 \) is a finite field of characteristic \( p \). In this case, the \( p \)-adic completion of \( R_1 \) is isomorphic to the Witt vectors \( W(k_1) \) over \( k_1 \).

2. The ring \( \mathcal{O}_F \) is split over \( R_1 \), in the sense that \( \text{Hom}_{\mathbb{Z},\text{alg}}(\mathcal{O}_F , R_1) \) has cardinality \( [F : \mathbb{Q}] \).

3. There exists an isomorphism

\[
(\mathcal{L} \otimes R_1 , \langle \cdot , \cdot \rangle) \cong (L_0 \oplus L_0^\vee(1) , \langle \cdot , \cdot \rangle_{\text{can.}}) \otimes \mathcal{O}_{F_0'(p)} \otimes R_1
\]
inducing an isomorphism $G \otimes R_1 \cong G_0 \otimes R_1$ realizing $P_0 \otimes R_1$ as a subgroup of $G \otimes R_1$. (The existence of the isomorphism \[2.17\] follows from [Lan08, Cor. 1.2.3.10] by comparing multi-ranks.)

**Remark 2.18.** For the purpose of studying arithmetic questions, it is often harmless (and helpful) to enlarge the coefficient ring.

From now on, let us fix the choices of $R_1$ and the isomorphism \[2.17\], and set $O_{F,1} := O_F \otimes R_1$, $O_1 := O \otimes R_1$, $L_1 := L \otimes R_1$, $L_{0,1} := L_0 \otimes R_1$, $G_1 := G_0 \otimes R_1 \cong G \otimes R_1$, $P_1 := P_0 \otimes R_1$, and $M_1 := M_0 \otimes R_1$. We shall also denote base changes of geometric objects such as $M_{N,0}$ (from $O_{F,1(p)}$ to $R_1$ by replacing $0$ with $1$).

**Definition 2.19.** The principal $G_1$-bundle over $M_{N,1}$ is the $G_1$-torsor

$$\mathcal{E}_{G_1} := \text{Isom}_{\hat{O}, \hat{O}_{M_{N,1}}}((\hat{H}^1_1(A/M_{N,1}), \langle \cdot, \cdot \rangle_\lambda, \hat{\mathcal{O}}_{M_{N,1}}(1)), ((L_{0,1} \oplus L_{0,1}'(1)) \otimes \mathcal{O}_{M_{N,1}}(1), \langle \cdot, \cdot \rangle_{\text{can}}, \mathcal{O}_{M_{N,1}}(1)))$$

the sheaf of isomorphisms of $\hat{\mathcal{O}}_{M_{N,1}}$-sheaves of symplectic $\hat{O}$-modules. (The pairing $\langle \cdot, \cdot \rangle_\lambda$ on $\hat{H}^1_1(A/M_{N,1})$ is defined by the polarization $\lambda$ as in [DP94, 1.5]. The group $G_1$ acts as automorphisms on $(L \otimes \hat{\mathcal{O}}_{M_{N,1}}, \langle \cdot, \cdot \rangle_{\text{can}}, \hat{\mathcal{O}}_{M_{N,1}}(1))$ by definition. The third entries in the tuples represent the values of the pairings. We allow isomorphisms of symplectic modules to modify the pairings up to units.)

**Definition 2.20.** The principal $P_1$-bundle over $M_{N,1}$ is the $P_1$-torsor

$$\mathcal{E}_{P_1} := \text{Isom}_{\hat{O}, \hat{O}_{M_{N,1}}}((\hat{H}^1_1(A/M_{N,1}), \langle \cdot, \cdot \rangle_\lambda, \mathcal{O}_{M_{N,1}}(1), \mathcal{O}_{\text{Lie}^\vee_{\text{A}/M_{N,1}}}, ((L_{0,1} \oplus L_{0,1}'(1)) \otimes \mathcal{O}_{M_{N,1}}(1), \langle \cdot, \cdot \rangle_{\text{can}}, \mathcal{O}_{M_{N,1}}(1), L_{0,1}'(1) \otimes \mathcal{O}_{M_{N,1}}(1)))$$

the sheaf of isomorphisms of $\mathcal{O}_{M_{N,1}}$-sheaves of symplectic $\hat{O}$-modules with maximal totally isotropic $\hat{O} \otimes R_1$-submodules. (The sheaf $\mathcal{O}_{\text{Lie}^\vee_{\text{A}/M_{N,1}}}$ is a subsheaf of $\hat{H}^1_1(A/M_{N,1})$ totally isotropic under the pairing $\langle \cdot, \cdot \rangle_\lambda$. The group $P_1$ acts as automorphisms on $(L \otimes \mathcal{O}_{M_{N,1}}, \langle \cdot, \cdot \rangle_{\text{can}}, \mathcal{O}_{M_{N,1}}(1), L_{0,1}'(1) \otimes \mathcal{O}_{M_{N,1}}(1))$ by definition.)

These are torsors because $(\hat{H}^1_1(A/M_{N,1}), \langle \cdot, \cdot \rangle_\lambda, \mathcal{O}_{M_{N,1}}(1), \mathcal{O}_{\text{Lie}^\vee_{\text{A}/M_{N,1}}})$ and $(L_{0,1} \oplus L_{0,1}'(1)) \otimes \mathcal{O}_{M_{N,1}}(1), \langle \cdot, \cdot \rangle_{\text{can}}, \mathcal{O}_{M_{N,1}}(1), L_{0,1}'(1) \otimes \mathcal{O}_{M_{N,1}}(1))$ are étale locally isomorphic by the theory of infinitesimal deformations (cf. for example [Lan08, Ch. 2]) and the theory of Artin’s approximations (cf. [Art69, Thm. 1.10 and Cor. 2.5]).

**Definition 2.21.** The principal $M_1$-bundle over $M_{N,1}$ is the $M_1$-torsor

$$\mathcal{E}_{M_1} := \text{Isom}_{\hat{O}, \hat{O}_{M_{N,1}}}((\mathcal{L}_{\text{Lie}^\vee_{\text{A}/M_{N,1}}}, \mathcal{O}_{M_{N,1}}(1)), (L_{0,1}'(1) \otimes \mathcal{O}_{M_{N,1}}(1), \mathcal{O}_{M_{N,1}}(1)))$$

the sheaf of isomorphisms of $\mathcal{O}_{M_{N,1}}$-sheaves of $\hat{O} \otimes R_1$-modules. (We view the second entries in the pairs as an additional structure, inherited from
the corresponding objects for $P_1$. The group $M_1$ acts as automorphisms on $(L_{\mathcal{O}_{H_1}}^0(1) \otimes \mathcal{O}_{M_1}, \mathcal{O}_{M_1}(1))$ by definition.)

**Definition 2.22.** For any $R_1$-algebra $R$, we denote by $\text{Rep}_R(G_1)$ (resp. $\text{Rep}_R(P_1)$, resp. $\text{Rep}_R(M_1)$) the category of finite $R$-modules with algebraic actions of $G_1 \otimes R$ (resp. $P_1 \otimes R$, resp. $M_1 \otimes R$).

**Definition 2.23.** Let $R$ be any $R_1$-algebra. For any $W \in \text{Rep}_R(G_1)$, we define

\[\mathcal{E}_{G_1,R}(W) := (\mathcal{E}_{G_1,G_1} \otimes R) \times W,\]

and call it the **automorphic sheaf** over $M_{\mathcal{O}_{H_1}} \otimes R$ associated with $W$. It is called an **automorphic bundle** if $W$ is locally free as an $R$-module. We define similarly $\mathcal{E}_{P_1,R}(W)$ (resp. $\mathcal{E}_{M_1,R}(W)$) for $W \in \text{Rep}_R(P_1)$ (resp. $W \in \text{Rep}_R(M_1)$) by replacing $G_1$ with $P_1$ (resp. with $M_1$) in the above expression [2.24].

**Example 2.25.** We have $\mathcal{E}_{G_1,R_1}(L_1) \cong \mathcal{E}_{P_1,R_1}(L_1) \cong H_1^dR(A/M_{\mathcal{O}_{H_1}})$, with Hodge filtration defined by the submodule $\mathcal{E}_{P_1,R_1}(L_{\mathcal{O}_{H_1}}^0(1)) \cong \mathcal{E}_{M_{\mathcal{O}_{H_1}},R_1}(L_{\mathcal{O}_{H_1}}^0(1)) \cong \text{Lie}^dV/M_{\mathcal{O}_{H_1}}$, and with top graded piece $\mathcal{E}_{P_1,R_1}(L_{\mathcal{O}_{H_1}}) \cong \mathcal{E}_{M_{\mathcal{O}_{H_1}},R_1}(L_{\mathcal{O}_{H_1}}) \cong \text{Lie}^dV/M_{\mathcal{O}_{H_1}}$.

The Hodge filtration on $H_1^dR(A/M_{\mathcal{O}_{H_1}})$ can be (compatibly) generalized by defining a Hodge filtration $\mathcal{F}$ on any object $W \in \text{Rep}_R(P_1)$, which induces the Hodge filtration on $\mathcal{E}_{P_1,R}(W)$, still denoted by $\mathcal{F}$. (For $W \in \text{Rep}_R(G_1)$ one considers $\mathcal{E}_{G_1,R}(W) \cong \mathcal{E}_{P_1,R}(W|_{P_1})$; for $W \in \text{Rep}_R(M_1)$, the Hodge filtration on $\mathcal{E}_{M_1,R}(W)$ is always canonically split.)

There is a canonical way to define the **Gauss-Manin connections**

\[\nabla : \mathcal{E}_{G_1,R}(W) \to \mathcal{E}_{G_1,R}(W) \otimes \Omega^1_{M_{\mathcal{O}_{H_1}}/S_R} \]

using the Gauss-Manin connection of $H_1^dR(A/M_{\mathcal{O}_{H_1}})$. The complex

\[(\mathcal{E}_{G_1,R}(W) \otimes \Omega^*_{M_{\mathcal{O}_{H_1}}/S_R}, \nabla)\]

induces is called the **de Rham complex** attached to $\mathcal{E}_{G_1,R}(W)$. (Good places to learn about connections and de Rham complexes in the algebraic setup are [KO68, Kat71, Kat70, Kat72].)

### 2.8. Canonical extensions in mixed characteristics; compactifications of Kuga families

Our treatment of canonical extensions in mixed characteristics follow mainly [FC90, Ch. VI and MT02] (although we do have a different construction of (good toroidal) compactifications of Kuga families in [Lanc], using toroidal boundary strata of larger Shimura varieties).

Let $m \geq 0$ be any integer, and let $N_m := A^m$ be the $m$-fold fiber product of $A \to M_{\mathcal{O}}$. By [Lanc, Thm. 2.15], by taking $Q := \mathcal{O}^{\oplus m}$ there (cf. [Lanc, Ex. 2.2]), the abelian scheme $N_m \to M_{\mathcal{O}}$ (which we call a Kuga family) admits a collection of (non-canonical) toroidal compactifications $N_{\mathcal{O}}^{\text{tor}}_m$, indexed by a directed partially ordered set $K_{m,\mathcal{O}}$ of $\kappa$’s, such that the (smooth) structural morphism $f_m : N_m \to M_{\mathcal{O}}$ extends to a proper log smooth morphism $f^{\text{tor}}_{m,\kappa} : N_{\mathcal{O}}^{\text{tor}}_m \to M_{\mathcal{O},\Sigma}$ for each $\kappa \in K_{m,\mathcal{O}}$. This collection $\{N^{\text{tor}}_{m,\kappa}\}_{\kappa \in K_{m,\mathcal{O}}}$ enjoys a long list of nice
properties (see the statements of [Lanc, Thm. 2.15]); we will give precise references to them when needed.

By abuse of notation, we shall denote the pullbacks of \( f_{m,\kappa}^{\text{tor}} : \mathbb{N}_{m,\kappa} \to M_{H,\Sigma}^{\text{tor}} \) to \( M_{H,\Sigma,0}^{\text{tor}} \) and \( M_{H,\Sigma,1}^{\text{tor}} \) by \( f_{m,\kappa}^{\text{tor}} : \mathbb{N}_{m,\kappa} \to M_{H,\Sigma,1}^{\text{tor}} \) and \( f_{m,\kappa}^{\text{tor}} : \mathbb{N}_{m,\kappa} \to M_{H,\Sigma,1}^{\text{tor}} \), respectively (and similarly over other base schemes).

**Proposition 2.26.** The locally free sheaf \( H_1^{\text{dR}}(A/M_{H,1}) \) extends to a unique locally free sheaf \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \) over \( M_{H,\Sigma,1}^{\text{tor}} \), satisfying the following properties:

1. The sheaf \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \), canonically identified as a subsheaf of the quasi-coherent sheaf \( (M_{H,1} \hookrightarrow M_{H,\Sigma,1}^{\text{tor}})_{\ast}(H_1^{\text{dR}}(A/M_{H,1})) \), is self-dual under the pairing \( (M_{H,1} \hookrightarrow M_{H,\Sigma,1}^{\text{tor}})_{\ast}((\cdot, \cdot))_\lambda \). We shall denote the induced pairing by \( (\cdot, \cdot)_{\lambda}^{\text{can}} \).
2. \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \) contains \( \text{Lie}_{(A^{ext})^{\nu}/M_{H,\Sigma,1}^{\text{tor}}} \) as a subsheaf that is totally isotropic under \( (\cdot, \cdot)_{\lambda}^{\text{can}} \).
3. The quotient sheaf \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}}/\text{Lie}_{(A^{ext})^{\nu}/M_{H,\Sigma,1}^{\text{tor}}} \) can be canonically identified with the subsheaf \( \text{Lie}_{A^{ext}/M_{H,\Sigma,1}^{\text{tor}}} \) of \( (M_{H,1} \hookrightarrow M_{H,\Sigma,1}^{\text{tor}})_{\ast}\text{Lie}_{A/M_{H,1}} \).
4. The pairing \( (\cdot, \cdot)_{\lambda}^{\text{can}} \) induces canonical an isomorphism \( \text{Lie}_{A^{ext}/M_{H,\Sigma,1}^{\text{tor}}} \cong \text{Lie}_{(A^{ext})^{\nu}/M_{H,\Sigma,1}^{\text{tor}}} \) which coincides with \( d\lambda^{\text{ext}} \).
5. Let

\[
H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} := \text{Hom}_{M_{H,\Sigma,1}^{\text{tor}}}((H_1^{\text{dR}}(A/M_{H,1})^{\text{can}}, \mathcal{O}_{M_{H,\Sigma,1}^{\text{tor}}}).
\]

The Gauss-Manin connection of \( H_1^{\text{dR}}(A/M_{H,1}) \) extends to an integrable connection

\[
\nabla : H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \to H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \otimes \Omega^1_{M_{H,\Sigma,1}^{\text{tor}}/S_1}
\]

with log poles along \( D \), called the extended Gauss-Manin connection, compatible with the extended Kodaira-Spencer morphism in Theorem 2.8.

With these properties, we say that \((H_1^{\text{dR}}(A/M_{H,1})^{\text{can}}, \nabla)\) is the canonical extension of \((H_1^{\text{dR}}(A/M_{H,1}), \nabla)\).

The locally free sheaf \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \) (with all these stated properties) is unique once it exists. To show the existence, we use the morphism \( f_1^{\text{tor}} : \mathbb{N}_{1,\kappa} \to M_{H,\Sigma}^{\text{tor}} \) for some \( \kappa \in K_{1,\Sigma} \), and show that the \( \mathcal{O}_{M_{H,\Sigma,1}^{\text{tor}}} \)-dual of the locally free sheaf \( H_1^{\text{dR}}(N_{1,\kappa}/M_{H,\Sigma}^{\text{tor}}) := R^1(f_1^{\text{tor}})_{\ast}(\mathcal{O}_{M_{H,\Sigma,1}^{\text{tor}}}^{\text{dR}}) \) satisfies all the properties of \( H_1^{\text{dR}}(A/M_{H,1})^{\text{can}} \) stated in Proposition 2.26 (See [Lanc, proof of Prop. 6.9, based on Thm. 2.15] for details). The compactifications of Kuga families with \( m \geq 1 \) will be useful for other purposes as well. See [3.4] below.

**Remark 2.27.** We formulated Proposition 2.26 in this somewhat axiomatic way to emphasize that any construction achieving these properties would serve the same purpose for the construction of canonical extensions (of automorphic bundles). Therefore, one can refer to [FC90, Ch. VI] and related works in special cases, without having to explain the consistency with [Lanc]. (This is desirable because the methods in [FC90, Ch. VI] and [Lanc] are different.)
Then the principal bundle $\mathcal{E}_{G_1}$ (defined above in §2.7) extends canonically to a principal bundle $\mathcal{E}_{G_1}^{\text{can}}$ over $\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}$ by setting

$$
(2.28) \quad \mathcal{E}_{G_1}^{\text{can}} := \text{Isom}_G \otimes \mathcal{O}_{\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}} ((H^1_{dR}(A/M_{\mathbb{H}, 1})_{\text{can}}, \mathcal{O}_{\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}}(1)),
((L_{0,1} \oplus L_{G_1}^{\text{can}}(1)) \otimes \mathcal{O}_{\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}}, G_{\text{can}})_{\text{can}}).$$

Similarly, the principal bundle $\mathcal{E}_{P_1}$ (resp. $\mathcal{E}_{M_1}$) extends canonically to a principal bundle $\mathcal{E}_{P_1}^{\text{can}}$ (resp. $\mathcal{E}_{M_1}^{\text{can}}$) over $\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}$. As before, these are torsors by Artin’s theory of approximations (cf. [Art69] Thm. 1.10 and Cor. 2.5]), because they have sections over completions of strict local rings.

**Definition 2.29.** Let $R$ be any $R_1$-algebra. For any $W \in \text{Rep}_R(G_1)$, we define

$$
(2.30) \quad \mathcal{E}_{G_1, R}(W) := (\mathcal{E}_{G_1}^{\text{can}} \otimes R) \times W,
$$

called the canonical extension of $\mathcal{E}_{G_1, R}(W)$, and define accordingly $\mathcal{E}_{P_1, R}(W) := \mathcal{E}_{P_1}^{\text{can}}(W) \otimes \mathcal{O}_D$, called the subcanonical extension of $\mathcal{E}_{G_1, R}(W)$, where $\mathcal{O}_D$ is the $\mathcal{O}_{\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 1}}$-ideal defining the relative Cartier divisor $D$ (in (2) of Theorem 2.8).

Using the extended Gauss-Manin connection in Proposition 2.26, the Gauss-Manin connection of $\mathcal{E}_{G_1, R}(W)$ extends (uniquely) to integral connections of $\mathcal{E}_{G_1, R}(W)$ and $\mathcal{E}_{P_1, R}(W)$ (with log poles along the boundary divisor $D$). We define similarly $\mathcal{E}_{P_1, R}(W), \mathcal{E}_{M_1, R}(W), \mathcal{E}_{G_1, R}(W)$, and $\mathcal{E}_{P_1, R}(W)$ with $G_1$ (and its principal bundle) replaced accordingly with $P_1$ and $M_1$ (and their respective principal bundles).

### 2.9. Geometric modular forms in higher dimensions

Now we are ready to give definitions of modular forms in higher dimensions and in mixed characteristics. We will call them algebraic automorphic forms.

Let $R$ be any $R_1$-algebra. For any $W \in \text{Rep}_R(G_1)$, we can then define the graded modules of $R$-valued algebraic automorphic forms of weight $W$ using cohomology of the coherent sheaf $\mathcal{E}_{G_1, R}(W)$ over $\mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, R} := \mathcal{M}^{\text{tor}}_{\mathbb{H}, \Sigma, 0} \otimes R$.

The theory is more useful if we have parameters for the weight modules $W$ to be used. In mixed characteristics $(0, p)$, it is helpful to at least introduce the notion of $p$-small weights: (In what follows, we will use the obvious notations such as the weights $X_G$ and roots $\Phi_G$ for $G_1$, although we have not defined them formally. The dominant weights $X^+_G$ and $X^+_M$ will be chosen compatibly with each other, as in the case over $\mathbb{C}$.)

**Definition 2.31.** We say $\mu \in X^+_G$ is $p$-small if $(\mu + \rho, \alpha^\vee) \leq p$ for every $\alpha \in \Phi_G^+$. We say $\mu \in X^+_M$ is $p$-small if $(\mu + \rho, \alpha^\vee) \leq p$ for every $\alpha \in \Phi_M^+$. We denote the subset of $X^+_G$ (resp. $X^+_M$) that are $p$-small by $X^{<p}_G$ (resp. $X^{<p}_M$).

Since $G_1$ (resp. $M_1$) is split over $R_1$, there exists a split reductive algebraic group $G_{\text{split}}$ (resp. $M_{\text{split}}$) over $\mathbb{Z}$ such that $G_1 \cong G_{\text{split}} \otimes \mathbb{R}_1$ (resp. $M_1 \cong M_{\text{split}} \otimes \mathbb{R}_1$).

Using the Weyl modules (over $\mathbb{Z}$), namely the span of a highest weight vector under the action of the group scheme and the distribution algebra over $\mathbb{Z}$, we can define canonically (by base extension from $\mathbb{Z}$ to $R_1$) the $R_1$-module $V_{\mu}$ (resp. $W_{\nu}$) of highest weight $\mu \in X^{<p}_G$ (resp. $\nu \in X^{<p}_M$). We set $V_{\mu, R} := V_{\mu} \otimes R$ (resp. $W_{\nu, R} := W_{\nu} \otimes R$).
\(W_{\nu,R} := W_\nu \otimes R\) for any \(R\)-algebra \(R\), and we set \(V_{\mu,R} := \mathcal{E}_{G_1,R}(V_{\nu,R}), \mathcal{E}^\text{can}_{G_1,R}(V_{\nu,R}), \mathcal{E}_{G_1,R}(V_{\nu,R}), \mathcal{E}^\text{can}_{G_1,R}(V_{\nu,R})\), and \(W^\text{can}_{\nu,R} := \mathcal{E}^\text{can}_{M_1,R}(V_{\nu,R})\). (For more details concerning the Weyl modules, see \[\text{LSb} \] §2.6, its references to \[\text{PT02}\], and the references of \[\text{PT02}\] to other works. Here for ease of exposition we assumed that \(G_1\) has no type D factors, but this assumption is not necessary.)

**Definition 2.32.** Let \(\nu \in X^{+,<\nu}_M\). Let \(R\) be any \(R\)-algebra. Consider the following graded modules of \(R\)-valued algebraic automorphic forms of weight \(\nu\):

1. \(A^\text{can}_{\nu,R}(\mathcal{H}; R) := H^\text{tor}_{\nu,R}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{can}_{\nu,R}, R).\) We call these forms \textit{canonical}.
2. \(A^\text{sub}_{\nu,R}(\mathcal{H}; R) := H^\text{tor}_{\nu,R}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{sub}_{\nu,R}, R).\) We call these forms \textit{subcanonical}.
3. \(A^\text{int}_{\nu,R}(\mathcal{H}; R) := \text{image}(H_{\text{tor}}^{\nu,R}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{can}_{\nu,R}) \rightarrow H^\text{tor}_{\nu,R}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{can}_{\nu,R})).\) We call these forms \textit{interior}.

(The modifier “canonical” will often be suppressed.) In all three cases, the choice of \(\Sigma\) is immaterial (cf. \[\text{Lan} \] (4) of Thm. 2.15 or rather \[\text{Lan08} \] proof of Lem. 7.1.1.3).

However, our terminology (canonical, subcanonical, and interior) are \textit{not} standard; there do not seem to be standard names for these spaces, except when the cohomology degree is 0 or \(d\) (thanks to the prototypical case \(R = \mathbb{C}\)). In degree 0, forms in \(A^0_{\nu,R}(\mathcal{H}; R)\) can be called holomorphic, while forms in \(A^0_{\nu,R}(\mathcal{H}; R)\) can be called cuspidal holomorphic. In degree \(d\), forms in \(A^d_{\nu,R}(\mathcal{H}; R)\) can be called anti-holomorphic (thanks to Hodge theory over \(\mathbb{C}\)), while forms in \(A^d_{\nu,R}(\mathcal{H}; R)\) can be called cuspidal anti-holomorphic (thanks to Serre duality, following the case of degree 0). We refrain from calling \(A^d_{\nu,R}(\mathcal{H}; R)\) cuspidal because this is not justified in degrees higher than 0. In general, \(A^d_{\nu,R}(\mathcal{H}; R)\) is not a submodule of \(A^d_{\nu,R}(\mathcal{H}; R)\).

The main justification we have for these terminologies is that these algebraic automorphic forms can be used to define filtrations on the algebraic log de Rham cohomology in mixed characteristics. More precisely, the usual \textit{Hodge spectral sequence} for the log de Rham cohomology groups

\[H^{a+b}_{\text{log-dR}}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{can}_{\nu,R}) := H^a_{\text{log-dR}}(M^{\text{tor}}_{\mathcal{H}, \Sigma,R}; V^\text{can}_{\nu,R}) \otimes_{\kappa^{\text{tor}}_{\mu,R}} \mathfrak{m}^\text{can}_{\mathcal{H},R}/\mathfrak{m}^\text{can}_{\mathcal{H},R}
\]

can be replaced with the dual \textit{BGG spectral sequence}

\[E_1^{a,b} := \oplus_{w \in W^{b1}} H^{a+b-l(w)}(M^{\text{tor}}_{\mathcal{H},R}; \text{Gr}^a_{\nu,R}((W^\text{can}_{w-R}))) \Rightarrow H^a_{\text{log-dR}}(M^{\text{tor}}_{\mathcal{H},R}; V^{\text{can}}_{\nu,R}).\]

There is also a “compactly supported” analogue with \((W^\text{can}_{w-R})\) replaced with \((W^\text{can}_{w-R})^{\text{sub}}\). (See Faltings \[\text{Fal83}\], Faltings-Chai \[\text{FC90}\], Mokrane-Tilouine-Polo \[\text{MPT02}\], and my article with Polo \[\text{LP} \] for details in the dual BGG construction.)

**Remark 2.35.** If \(R\) is a field of characteristic zero, then the log de Rham cohomology \(H^a_{\text{log-dR}}(M^{\text{tor}}_{\mathcal{H},R}; V^{\text{can}}_{\nu,R})\) calculates the usual de Rham cohomology \(H^a_{\text{dR}}(M_{\mathcal{H},R}; V^{\text{can}}_{\nu,R}).\) However, this is in general not true when \(R\) has a residue field of characteristic \(p > 0\). Moreover, because we do not have Hironaka’s embedded resolution of singularities \[\text{[Hir64a] [Hir64b]} \] in mixed characteristics, we do not know if any non-toroidal smooth compactifications of \(M_{\mathcal{H},R}/S_R\) with a simple normal crossings boundary divisor would yield the same cohomology groups.
The obvious advantage of (2.34) is that its left-hand side is given by a direct sum of spaces of algebraic automorphic forms, for which we might have better methods, rather than by some abstract hypercohomology.

3. Cohomology of torsion automorphic sheaves and vanishing theorems

Since Taylor’s thesis (see [Tay88, Thm. 4.2]), people know that the torsion in the cohomology of (unions of) locally symmetric spaces (including Shimura varieties) can be abundant and can have number-theoretic significance. Recent works such as Bergeron-Venkatesh [BV] have even shown examples where the torsion grows exponentially with levels. On the other hand, although it is speculated that there is little or no torsion in the case of Shimura varieties, not much seemed to be known.

In Mokrane-Tilouine [MT02] and Dimitrov [Dim05], they explored the idea in Faltings-Chai [FC90] of studying the \( \mathbb{Z}_p \)-valued Betti or étale cohomology using log crystalline and log de Rham cohomology in characteristic \( p \), and proved some vanishing and freeness results for the special cases of Siegel modular threefolds and Hilbert modular varieties, after localization at a prime of the Hecke algebra with non-effective conditions on the image of the associated mod \( p \) Galois representation. Their work is based on combinatorial comparisons of patterns of Hodge-Tate weights, and the non-effective conditions they used are to guarantee that the desired patterns can only appear in certain preferred cohomology degrees. A fundamental question is whether there is a method which does not rely on such non-effective conditions.

In my joint work with Junecue Suh [LSa, LSb, LSc], we discovered a way to translate the (geometric) “Kodaira type conditions” in vanishing theorems of Deligne-Illusie [DI87], Illusie [Ill90], Esnault-Viehweg [EV92], and Ogus [Ogu94] to the (representation-theoretic) “sufficient regularity conditions” in vanishing theorems of Faltings [Fal83], Vogan-Zuckerman [VZ84], Li-Schwermer [LS04], Saper [Sap05], and others, and proved new vanishing theorems with torsion coefficients. As a byproduct, we obtained freeness for the interior cohomology, namely the canonical image of \( H_c \) in \( H \), with coefficients in \( \mathbb{Z}_p \)-valued automorphic sheaves, under a list of mild and effective conditions, for all PEL-type Shimura varieties.

Our approach does not require any non-effective assumptions (such as those in Mokrane-Tilouine [MT02]), and gives after base change to \( \mathbb{C} \) a purely algebraic proof of several vanishing results so far only proved by transcendental methods.

3.1. Setup. Let us maintain the setup in §2 (and in particular the choices of \( p \) and \( R_1 \)). An important running assumption is that there is no level at \( p \).

The structural homomorphism \( O_{F_0} \to R_1 \) determines a \( p \)-adic place of \( F_0 \), and we will denote the completion of \( O_{F_0} \) at this place by \( W \). Since \( p \) is unramified in \( O \) and hence in \( O_{F_0} \), we can identify \( W \) with the ring of Witt vectors of its residue field. By passing to the completions, \( W \) embeds canonically into the \( p \)-adic completion of \( R_1 \).

3.2. Vanishing theorem for torsion Betti cohomology.

**Definition 3.1.** We say that \( \mu \in X_{G_\mathbb{C}}^+ \) is sufficiently regular if, for any positive root \( \alpha \in \Phi_{G_\mathbb{C}}^+ \), we have \( (\mu, \alpha^\vee) \geq 1 \) for any \( \alpha \) coming from a type A factor of \( G_\mathbb{C} \), and \( (\mu, \alpha^\vee) \geq 2 \) otherwise. (We define similarly for \( \mu \in X_{G_1}^+ \) when we work in mixed characteristics later.)
For any cohomology theory with the notion of compactly supported cohomology, we define the **interior cohomology** by setting $H_{\text{int}} := \text{image}(H_c \to H)$. Note that $H_{\text{int}}$ is *not* defined as a derived functor. Therefore, for example, there are no long exact sequences attached to short exact sequences.

**Theorem 3.2.** There exists an explicit function $C(\mu)$ (depending on $\mu$ and also on the PEL datum, but not on the level) such that, for any **sufficiently regular** weight $\mu \in X_{\mathbb{C}}$ as above, for any prime $p$ good for the PEL datum (defining $\text{Sh}_{H,\mathbb{C}}$) such that $p > C(\mu)$, and for any $W$-algebra $R$, we have:

1. $H^i(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R}) = 0$ for every $i < d := \dim_{\mathbb{C}}(\text{Sh}_{H,\mathbb{C}})$.
2. $H^i(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R}) = 0$ for every $i > d$.
3. $H^i_{\text{int}}(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R}) = 0$ for every $i \neq d$.
4. ("liftability") For $? = c$ or int, the canonical "reduction mod $p$" morphism $H^?_!(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R}) \to H^?_!(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R})$ is surjective.
5. ("freeness") For $? = \text{c}$ or int, if $\text{Sh}_{H,\mathbb{C}}$ is compact, or if $R$ is flat over $W$, then $H^?_!(\text{Sh}_{H,\mathbb{C}}, B_{\mu,R})$ is a free $R$-module of finite rank.

If $\text{Sh}_{H,\mathbb{C}}$ is compact, or if one only cares about $\mathbb{C}$, then the case where $\mathbb{C}$ can be proved by transcendental methods (harmonic forms, $L_2$ methods, etc) as in, for example, Faltings [Fal83]. In the non-compact case, to the best of our knowledge, the first analytic proofs of 1 and 2 were given by Li and Schwermer’s work on the Eisenstein cohomology of arithmetic groups (see [LS04, Cor. 5.6]), and roughly at the same time by Saper’s work on $L$-modules (see [Sap05, §11, Thm. 5]). Before their works, the important special case of symplectic groups with factors of rank two was treated using Franke’s method in [TU99, Appendix A]. However, there is no known transcendental proof for the torsion case $R = k_1$ (or rather $R = \mathbb{F}_{p^\infty}$).

In fact, it is an elementary exercise in homological algebra that the special case $R = k_1$ implies all other cases. Therefore it suffices to focus on this special case.

### 3.3. Vanishing theorem for torsion log de Rham and log Hodge cohomology

Based on a series of reduction steps using (torsion) comparison theorems among Betti, étale, log crystalline, and log de Rham cohomology (see [Del77, Aracata, V, Cor. 3.3] and [AGV73, XI, Thm. 4.4] for the Betti–étale comparison; see [LSb, §5] and [LSc, §9] for the explanation of the étale–log crystalline comparison based on [BM02, FM87, Fal89, and Fal02; see [Ber74, BO78], and [Kat89 Thm. 6.2] for the log crystalline–log de Rham comparison; for all of these we use the Kuga families and their good toroidal compactifications), the proof of Theorem 3.3 can be reduced to the following:

**Theorem 3.3.** For any $R_1$-algebra $R$, the analogues of Theorem 3.2 for the log Hodge and log de Rham cohomology over $M_{\log,R}$ are true.

### 3.4. Idea of the proof

It suffices to show that $H^i_{\log,\text{dR}}(M_{H,R}^{\text{tor}}; (\omega_{\mu,R})^{\text{can}}) = 0$ for $i < d$. By (2.34) (with $R = k_1$), it suffices to show that, for any $w \in W_{M_1}$ and any $i < d$, we have

$$H^{d-i}(w)(M_{H,k_1}^{\text{tor}}; (W_{w,R})^{\text{can}}) = 0.$$  

(3.4)

Without some condition on $\mu$, this cannot be true. (See [Suh08] for counter-examples in the context of compact Picard modular surfaces. There $\mu$ is trivial and hence violates the sufficient regularity condition in Definition 3.1.)
Now that the main point is to show that the cohomology groups of certain coherent sheaves vanish, it is natural for us to resort to generalizations of Kodaira vanishing in characteristic $p$, which is essentially the only vanishing we know other than the cruder Serre’s vanishing. (I would recommend starting with Deligne-Illusie [DI87] and Esnault-Viehweg [EV92].)

For any such vanishing theorem, the coherent sheaves are tensored with line bundles with certain positivity condition, and we prefer such line bundles to be canonical extensions of automorphic line bundles, i.e., of the form $W^\text{can}_{\nu_0,k_1}$ for some $\nu_0 \in X^{+1}_{M_1}$. On integral models of PEL-type Shimura varieties, there is essentially only one source of such line bundles, namely those constructed using variants of the $\omega$ in Theorem [2.8] and we can find choices of $\nu_0$ such that, for any $w \in W_{G_1}$, we have $(w(\nu_0), \alpha') \leq 1$ for any root $\alpha$ coming from a type $A$ factor of $G_1$, and $(w(\nu_0), \alpha') \leq 2$ otherwise (cf. Definition [3.1]).

In the compact case, we can use Deligne-Illusie [DI87] and Illusie [Ill90], together with (2.34), and show that

$$H^{1-\ell(w)}(M_{H,k_1}, W^\vee_{\nu_0,k_1}) = 0$$

for $\mu \in X^{+1}_{M_1}$ satisfying some condition $p > C_0(\mu)$ and for any $w \in W^{M_1}$. Here $W^\vee_{\nu_0,k_1} \cong W^\vee_{\nu_0,k_1} \otimes W^\vee_{w \cdot \mu,k_1}$, where $W^\vee_{w \cdot \mu,k_1}$ is an ample line bundle, and where $W^\vee_{w \cdot \mu,k_1}$ is a vector bundle with an integrable connection of a geometric origin, because up to Tate twist it can be constructed (via geometric plethysm) as a summand of the relative de Rham cohomology of some $N_{m,k_1} = A^m_{k_1} \rightarrow M_{H,k_1}$ for some $m \geq 0$. The results of Deligne-Illusie and Illusie apply because the whole setup lifts to $W_2(k_1)$ (and in fact even $W(k_1)$).

By changing our perspective a little bit, we can fix $\mu \in X^{+1}_{G_1}$ and ask for each $w \in W^{M_1}$ whether there exists some $\mu' \in X^{+1}_{G_1}$ such that $p > C_0(\mu')$ and $w \cdot \mu = w \cdot \mu' + \nu_0$. Equivalently, this is asking whether $\mu - w(\nu_0)$ is an element of $X^{+1}_{G_1}$ satisfying $p > C_0(\mu')$. If $\mu$ is sufficiently regular as in Definition [3.1], then $(\mu - w(\nu_0), \alpha') \geq 0$, and the condition is verified if we define $C(\mu)$ to be slightly larger than $C_0(\mu)$. This proves the key vanishing (3.4), as desired. (One can picture this as having enough room in the Weyl chamber for shifting $\mu$ towards the wall.)

To generalize the argument to the non-compact case, we need the following:

1. We need to use the fact that the relative log de Rham cohomology $\mathcal{H}$ of $f^\text{tor}_{m,n} : N^\text{tor}_{m,n} \rightarrow H^\text{tor}_{H^1}$ enjoys a long list of (unusually) nice properties. Note that since the morphism $f^\text{tor}_{m,n}$ is not semistable in general, most of the properties we need (in mixed characteristic) are impossible to prove by abstract methods. Fortunately, in [Lanc], we can prove everything we need by using explicit boundary charts of toroidal compactifications. (For experts working on varieties in general, it is perhaps surprising that such good compactifications exists at all in mixed characteristics.)

2. By applying Ogus’s result [Ogu94] to the log crystal attached to $\mathcal{H}$ (over $W(k_1)$), we obtain the so-called decomposition theorem (generalizing those of Deligne-Illusie’s and Illusie’s) for the push-forward of the de Rham complex (of the reduction over $k_1$) under the relative Frobenius morphism. (Here the main properties we need about $\mathcal{H}$ are that it is locally free, self-dual, with degenerate (relative) Hodge spectral sequence, and that the crystalline Frobenius of the attached log crystal is an isogeny.)
(3) In Esnault-Viehweg [EV92], they proved a Kodaira type vanishing theorem for line bundles that satisfy some positivity condition weaker than ampleness along the boundary, which quite luckily is satisfied in our context thanks to (2.9) in Theorem [LSa]. (This allowed us to give a simple proof of a liftability theorem in [LSa], which nevertheless is a special case of what we are proving now.) Combining their techniques with the decomposition theorem provided by Ogus, we can show that

$$H^{1-i(w)}(\mathcal{M}^\text{tor}_{H,k_1}, (W^\vee_{w+\mu+v_0,k_1})^\text{ran}) = 0.$$  

(Here the main property we need about $$\mathcal{H}$$ is that the Gauss-Manin connection has nilpotent residues along the irreducible components of the boundary divisor $$D$$ of $$\mathcal{M}^\text{tor}_{H,k_1}$$.)

(4) Then we can conclude the proof of (3.4) as in the compact case. This gives a sketch of the proof of Theorem 3.3. As a byproduct, we also obtain a suitable analogue of Theorem 3.3 for cohomological algebraic automorphic forms, namely those that appear in the left-hand side of (2.34) for some $$\mu$$.

### 3.5. Concluding remark.

**Remark 3.5 (sources of torsion).** The results we have obtained suggest that the only possible sources of $$p$$-torsion in the cohomology of Shimura varieties, or more ambitiously cohomology of arithmetic groups (i.e. of arithmetic quotients of Riemannian symmetric spaces, often without Hermitian structures), are as follows:

1. Lack of Hermitian structure (not even Shimura varieties).
2. Boundary cohomology (the cone of the canonical morphism from $$H_c$$ to $$H$$, complementing the interior cohomology).
3. Ramification at $$p$$ (including levels); i.e., bad reductions.
4. Large weights (compared with $$p$$).
5. Irregular weights.

We believe (1) and (2) are related. We are inclined to believe that (3), (4), (5) are difficult for intrinsic reasons.

There are many concurrent works on torsion in the cohomology of arithmetic groups, and we have to stop our survey here due to the limitation of our knowledge. We hope that this article is at least useful for understanding one aspect of the theory.

### References


[BGG75] I. N. Bernstein, I. M. Gelfand, and S. I. Gelfand, *Differential operators on the base affine space and a study of $g$-modules*, in Gelfand [Gel75], pp. 21–64.


GEOMETRIC MODULAR FORMS AND COHOMOLOGY


[LSc] , Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties, preprint.


[Mil05] , Introduction to Shimura varieties, in Arthur et al. [AEK05], pp. 265–378.


[MT02] A. Mokrane and J. Tilouine, Cohomology of Siegel varieties with p-adic integral coefficients and applications, in Cohomology of Siegel varieties [MPT02], pp. 1–95.


[PT02] P. Polo and J. Tilouine, Bernstein-Gelfand-Gelfand complex and cohomology of nilpotent groups over $\mathbb{Z}(p)$ for representations with $p$-small weights, in Cohomology of Siegel varieties [MP02], pp. 97–135.


