

# Analytic continuation of overconvergent eigenforms.

Kevin Buzzard\*

October 21, 2004

## Abstract

Let  $f$  be an overconvergent  $p$ -adic eigenform of level  $Np^r$ ,  $r \geq 1$ , with non-zero  $U_p$ -eigenvalue. We show how  $f$  may be analytically continued to a subset of  $X_1(Np^r)^{\text{an}}$  containing, for example, all the supersingular locus. Using these results we extend the main theorem of [BT] to many ramified cases.

## Introduction

Firstly we give an overview of our main results and applications. Let  $r \geq 1$  be an integer, let  $p$  be a prime, and let  $N \geq 5$  be a positive integer prime to  $p$ . Let  $k$  be an integer and let  $f$  be an overconvergent  $p$ -adic modular form of weight  $k$  and level  $Np^r$ . That is,  $f$  is a section of  $\omega^{\otimes k}$  over a rigid analytic subspace of  $X_1(Np^r)$  strictly containing the connected component of the ordinary locus containing the cusp  $\infty$  (we shall be more precise about these notions in Section 1 below). We show that if  $U_p f = a_p f$  for some non-zero  $p$ -adic number  $a_p$  then  $f$  can be analytically continued to a section of  $\omega^{\otimes k}$  over a rather large analytic subspace of  $X_1(Np^r)$ . This subspace can be succinctly described as the complement of the component of the ordinary locus which contains the cusp 0. Note that this subspace contains the entire supersingular locus, and if  $r > 1$  then it contains a lot of the ordinary locus as well.

---

\*The author would like to thank the Miller Institute and UC Berkeley for the financial support and hospitality they offered him whilst he was obtaining the majority of these results. The write-up was done over a period of several years, in Rennes, the IHP in Paris, Cambridge UK, and Imperial College London, and the author would also like to thank these institutions for their hospitality. He would also like to thank the referee for several helpful remarks.

Descriptive title: Overconvergent Eigenforms

2000 AMS subject classifications:

Primary: 11F80, 11F33.

Secondary: 11G18, 14G22, 14G35.

Key words and phrases: Galois representations,  $p$ -adic modular forms.

As an application of this result, we explain how one can extend the results of [BT] to cover a larger class of representations. The main result of [BT] is a criterion for establishing that certain 2-dimensional continuous odd irreducible  $p$ -adic representations of  $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$  which are unramified at  $p$  are in fact modular, coming from classical weight 1 eigenforms. Because of the overconvergence results mentioned in the previous paragraph, one can now extend these results to cover certain cases where the  $p$ -adic representation is ramified at  $p$ .

More precisely, let  $p$  be a prime, and let  $L/\mathbf{Q}_p$  be a finite extension. Let  $\mathcal{O}$  denote the integers of  $L$  and let  $\lambda$  be the maximal ideal of  $\mathcal{O}$ . Let  $\rho : \text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \rightarrow \text{GL}_2(\mathcal{O})$  be a continuous representation, and let  $\bar{\rho}$  denote the mod  $\lambda$  reduction of  $\rho$ . Let  $D_p$  denote a decomposition group of  $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ . Let  $K$  denote the field  $\mathbf{Q}(\sqrt{(-1)^{(p-1)/2}p})$  if  $p$  is odd, and  $\mathbf{Q}(i)$  if  $p = 2$ .

**Theorem.** *Assume that*

- $\rho$  is ramified at only finitely many primes,
- $\bar{\rho}$  is modular,
- $\bar{\rho}$  is absolutely irreducible when restricted to  $\text{Gal}(\overline{\mathbf{Q}}/K)$ ,
- $\rho|_{D_p}$  is the direct sum of two 1-dimensional characters  $\alpha$  and  $\beta : D_p \rightarrow \mathcal{O}^\times$ , such that  $\alpha(I_p)$  and  $\beta(I_p)$  are finite, and  $(\alpha/\beta) \bmod \lambda$  is non-trivial.

*If furthermore  $p = 2$ , assume that*

- $\bar{\rho}(c) \neq 1$
- $\bar{\rho}(c)$  is both  $\alpha$ -modular and  $\beta$ -modular, in the sense that there are eigenforms  $f_\alpha$  with  $T_2$ -eigenvalue  $\bar{\alpha}(\text{Frob}_2)$  and  $f_\beta$  with  $T_2$ -eigenvalue  $\bar{\beta}(\text{Frob}_2)$  giving rise to  $\bar{\rho}$ .
- The projective image of  $\bar{\rho}$  is not dihedral.

*Then  $\rho$  is modular, in the sense that there exists an embedding  $i : L \hookrightarrow \mathbf{C}$  and a classical weight 1 cuspidal eigenform  $f$  such that the composite  $i \circ \rho$  is isomorphic to the representation associated to  $f$  by Deligne and Serre.*

*Remarks.* 1. The extra assumptions for  $p = 2$  are because we only know Dickinson's  $R = T$  theorem under these assumptions, and in particular the necessary "companion form" result has only been proved by Gross modulo unchecked compatibilities.

2. The theorem implies that the image of  $\rho$  is finite, and moreover that the  $L$ -function of  $i \circ \rho$  has an analytic continuation to all of  $\mathbf{C}$ .
3. The paper [BT] proves the theorem in the case that both  $\alpha$  and  $\beta$  are unramified. Moreover, the introduction to [BT] gives several reasons why we were motivated to look at this problem.

4. The strategy of this paper is broadly the same as that of [BT], in that one obtains the classical weight 1 form by glueing together two overconvergent forms and applying rigid GAGA. But the techniques used for glueing in this paper are rather different to those used in [BT]. Two new ideas seem to be needed, one to prove the result when  $\alpha/\beta$  is tamely ramified, and another for the case of wild ramification, and one of the motivations of this paper is to explain these ideas. Also we shall explain another construction of the glueing in the case where  $\alpha/\beta$  is unramified. These glueing results appear in sections 9–11.
5. In the arguments of this paper, we frequently pass without comment from the algebraic to the  $p$ -adic analytic theory. For example, it is “clear” that the analytification of a modular curve parametrises generalised elliptic curves plus level structure in the rigid-analytic category. Also, if one takes an algebraic modular form and considers the associated analytic section of the associated analytic sheaf, then it is “clear” that the algebraically-defined  $q$ -expansion of the form, defined in terms of the algebraic Tate curve, equals the analytic  $q$ -expansion of the form, defined in terms of rigid-analytic Tate curves. Although we felt that these compatibilities were always justifiable, it was not in the spirit of the paper to justify them, and we knew of no references in the literature to put the reader’s mind at rest. We are hence very grateful to Brian Conrad for developing the foundations of this theory in the paper [Con], and the reader who is concerned with these foundational matters is referred to this paper.
6. Edray Goins seems to have recently used the glueing results in this paper to give a new proof that “Buhler’s representation” ([Bu]) is modular.

## 1 Modular curves.

In this section we define arithmetic models of the modular curves that we shall be interested in. This section is just a collection of results that we shall need later; everything here has been known to the experts for about 15 years, and a reference for this entire section is [KM].

Let  $p$  be a prime, and let  $N \geq 5$  be an integer prime to  $p$ . In this section we recall the definitions and basic properties of certain modular curves. Firstly we remark that there are two conventions currently in use for defining the curve  $X_1(M)$ , where  $M$  is a positive integer. One possibility is to think of the non-cuspidal locus of  $X_1(M)$  as parameterising elliptic curves equipped with a point of order  $M$ , and the other is to think of it as parameterising elliptic curves equipped with embeddings of  $\mu_M$ . It turns out that we shall be doing computations involving the Tate curve, and that it is convenient to use embeddings of  $\mu_M$  when  $M$  is prime to  $p$ . However, because we will be considering certain models of these curves over bases in which  $p$  is not invertible, it is more convenient to stick with the notation of [KM] and use points of order  $M$  when

$M$  is a power of  $p$ . The resulting curves are sometimes therefore hybrid and rather non-standard, but only differ by twists from either of the usual definitions. In particular, if our base contains an  $M$ th root of unity then all the curves we use are isomorphic to the modular curves defined by either of the standard conventions.

By  $Y_1(N)$  we shall mean the smooth curve over  $\mathbf{Z}[1/N]$  which parameterises elliptic curves  $E$  over  $\mathbf{Z}[1/N]$ -schemes equipped with an embedding from  $\mu_N$  to  $E$ . By  $X_1(N)$  we mean the usual compactification of  $Y_1(N)$ , which is smooth and proper over  $\mathbf{Z}[1/N]$ . The curve  $X_1(N)$  comes equipped with an invertible sheaf  $\omega_{X_1(N)}$ , which when restricted to  $Y_1(N)$  is the pushforward  $\omega_{Y_1(N)}$  of the sheaf of differentials of the universal elliptic curve over  $Y_1(N)$ .

If  $R$  is a  $\mathbf{Z}[1/N]$ -algebra, then  $X_1(N)_R$  will denote the base change of  $X_1(N)$  to  $R$ , and we shall denote the pullback of  $\omega_{X_1(N)}$  to  $X_1(N)_R$  by  $\omega_{X_1(N)}$ , which is unlikely to cause confusion because the base we are working over will usually be clear.

By  $X_1(N; p)$  we shall mean the usual proper flat model over  $\mathbf{Z}[1/N]$  of the modular curve associated to the group  $\Gamma_1(N) \cap \Gamma_0(p)$ . More precisely,  $X_1(N; p)$  is the curve described in [KM] whose non-cuspidal points parameterise elliptic curves  $E$  over  $\mathbf{Z}[1/N]$ -schemes equipped with an embedding  $i : \mu_N \rightarrow E$  and an isogeny  $\alpha : E \rightarrow E'$  of degree  $p$ . The fibre of  $X_1(N; p)$  in characteristic  $p$  is not smooth—it is the union of two smooth curves both isomorphic to  $X_1(N)_{\mathbf{F}_p}$ , and these curves cross transversally at points corresponding to supersingular elliptic curves.

The curve  $X_1(N; p)$  also comes equipped with a sheaf, which we shall call  $\omega_{X_1(N; p)}$ , which again on the non-cuspidal locus is the pushforward of the sheaf of differentials on the universal elliptic curve. For  $R$  a  $\mathbf{Z}[1/N]$ -algebra, we shall denote by  $X_1(N; p)_R$  the base change of  $X_1(N; p)$  to  $R$ , and denote the pullback of  $\omega_{X_1(N; p)}$  also by  $\omega_{X_1(N; p)}$ .

There are two degeneracy maps  $X_1(N; p) \rightarrow X_1(N)$ . The first, which we shall denote  $\pi_1$ , is induced on the non-cuspidal locus by the morphism of moduli problems sending  $(E, i, \alpha : E \rightarrow E')$  to  $(E, i)$ . The second, denoted  $\pi_2$ , is defined by sending  $(E, i, \alpha : E \rightarrow E')$  to  $(E', \alpha i)$  (recall that  $N$  is prime to  $p$ ). Both these morphisms are finite and flat. If  $R$  is a  $\mathbf{Z}[1/N]$ -algebra we shall still use  $\pi_1$  and  $\pi_2$  to denote the base extension of these morphisms from  $X_1(N; p)_R$  to  $X_1(N)_R$ . We have that  $\pi_1^* \omega_{X_1(N)} = \omega_{X_1(N; p)}$ .

Let  $m > 0$  be an integer, and let  $X_1(Np^m)$  denote the proper flat curve over  $\mathbf{Z}[1/N]$  associated to  $\Gamma_1(Np^m)$ . More precisely, we shall think of the non-cuspidal locus of this curve as parameterising the slightly hybrid structure of elliptic curves  $E$  over  $\mathbf{Z}[1/N]$ -schemes equipped with an embedding of  $\mu_N$  and a point of exact order  $p^m$  in the sense of [KM]. For  $R$  a  $\mathbf{Z}[1/N]$ -algebra, we denote by  $X_1(Np^m)_R$  the base extension of  $X_1(Np^m)$  to  $R$ . If  $R$  is a  $\mathbf{Z}[1/Np][\zeta]$ -algebra, where  $\zeta$  is a fixed primitive  $p^m$ th root of unity, then  $X_1(Np^m)_R$  inherits an automorphism  $w = w_\zeta$ , which on non-cuspidal points is the following morphism of moduli problems: For an elliptic curve  $E/S/\mathbf{Z}[1/Np][\zeta]$  equipped with an

embedding  $i : \mu_N \rightarrow E$  and a point  $P \in E(S)$  of exact order  $p^m$ , we define  $\phi : E \rightarrow F := E/\langle P \rangle$  and choose  $Q \in \ker(\check{\phi})(S) \subset F(S)$  such that under the canonical pairing  $\langle \cdot, \cdot \rangle_\phi : \ker(\phi) \times \ker(\check{\phi}) \rightarrow \mathbf{G}_m$ , we have  $\langle P, Q \rangle_\phi = \zeta$  (We remark here that there are 2 choices of “canonical pairing”. To fix ideas we shall use the one defined in §2.8 of [KM]. We could also tie down this pairing by stating that on the  $n$ -torsion in a Tate curve,  $\langle \zeta_n, q^{1/n} \rangle = \zeta_n$ ). The morphism  $w$  sends  $(E, i, P)$  to  $(F, \bar{i}, Q)$ . This morphism extends to the cusps and induces an automorphism of  $X_1(Np^m)_R$ .

The most esoteric curves that we shall consider will only be needed in characteristic prime to  $Np$ . Let  $\Gamma^0(p)$  denote the matrices  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  in  $\mathrm{SL}_2(\mathbf{Z})$  such that  $b \equiv 0 \pmod{p}$ . Let  $m > 0$  be an integer, and let  $X_1(Np^m; p)$  denote the smooth proper curve over  $\mathbf{Z}[1/Np]$  associated to the group  $\Gamma_1(Np^m) \cap \Gamma^0(p)$ . In other words, the non-cuspidal locus of  $X_1(Np^m; p)$  parameterises elliptic curves equipped with an embedding of  $\mu_N$ , a point  $P$  of exact order  $p^m$  and a cyclic subgroup of order  $p$  which has trivial intersection with the subgroup generated by  $P$ . As ever, if  $R$  is a  $\mathbf{Z}[1/Np]$ -algebra then let  $X_1(Np^m; p)_R$  denote the base extension of  $X_1(Np^m; p)$  to  $R$ .

There are two natural morphisms from  $X_1(Np^m; p)$  to  $X_1(Np^m)_{\mathbf{Z}[1/Np]}$ , denoted  $\pi_1$  and  $\pi_2$ , where  $\pi_1$  is the morphism which forgets the subgroup of order  $p$ , and  $\pi_2$  quotients out by it. Both are finite and flat of degree  $p + 1$ .

Finally, observe that the non-cuspidal locus of  $X_1(Np; p)_{\mathbf{Z}[1/Np][\zeta_p]}$  can be thought of as parameterising elliptic curves equipped with an embedding of  $\mu_N$ , and two points  $P$  and  $Q$  of order  $p$  generating  $E[p]$  and such that  $\langle P, Q \rangle = \zeta_p$ . This is basically because every cyclic subgroup  $D$  as above will have a unique generator which pairs with  $P$  to  $\zeta_p$ .

## 2 The geometric description of modular forms and $q$ -expansions.

Let  $N \geq 5$  be an integer. The functor on  $\mathbf{Z}[1/N]$ -schemes, sending a  $\mathbf{Z}[1/N]$ -scheme  $S$  to the set of isomorphism classes of elliptic curves over  $S$  equipped with an embedding of  $\mu_N$ , is then representable. This functor is represented by a universal elliptic curve  $E_1(N)$  over  $Y_1(N)$ . This leads to a rather slick definition of the space of meromorphic modular forms of level  $N$  and weight  $k$  defined over a  $\mathbf{Z}[1/N]$ -algebra  $R$ . Namely, such a form is just an element of  $H^0(E_1(N)_R, (\Omega_{E_1(N)_R/Y_1(N)_R}^1)^{\otimes k}) = H^0(Y_1(N)_R, \omega_{Y_1(N)_R}^{\otimes k})$ .

If one wants to discuss holomorphic modular forms using this method, there are two ways to proceed. One way, explained in [DR], is to use the fact that  $X_1(N)$  parameterises *generalised* elliptic curves over  $\mathbf{Z}[1/N]$ -schemes equipped with an embedding of  $\mu_N$  meeting every irreducible component of every geometric fibre. Then one can extend  $\omega$  to  $X_1(N)$  in an appropriate way and take sections of tensor powers of this extended invertible sheaf, over  $X_1(N)$ .

Another way, as explained in [K], is to define the Tate curve “ $\mathbf{G}_m/q^{\mathbf{Z}}$ ” as an elliptic curve over a  $\mathbf{Z}((q)) \otimes_{\mathbf{Z}} \mathbf{Z}[1/N]$ -algebra  $R$ , and then to define various

embeddings of  $\mu_N$  into this curve, after a possible extension of  $R$ . The resulting elliptic curve with level structure gives rise to a map from  $R$  to  $Y_1(N)$  and one can demand that the sections of  $(\Omega_{E_1(N)/Y_1(N)}^1)^{\otimes k}$ , when pulled back to the Tate curve, are of the form  $(\sum_{n \geq 0} a_n q^n)(\omega_{\text{can}})^{\otimes k}$ , that is, such that the coefficient of  $q^n$  in this expansion is zero for all  $n < 0$ . This is the point of view that we shall be adopting. For more details on why these two concepts coincide, we refer the reader to [Con]. Given a cusp  $c$  of a modular curve, we shall refer to the Tate curve and embedding of  $\mu_N$  controlling holomorphicity at this point as the ‘‘Tate curve at  $c$ ’’. We shall normalise things so that the Tate curve at  $\infty$  is the curve  $\mathbf{G}_m/q^{\mathbf{Z}}$  equipped with the embedding of  $\mu_N$  induced from the natural inclusion  $\mu_N \rightarrow \mathbf{G}_m$ .

One important result in this theory (see for example §1.6 of [K]) is the  $q$ -expansion principle, which loosely stated, says that if one is working over an irreducible modular curve (as will be the case for all curves we shall consider), then a modular form of some fixed weight  $k$  is determined by its  $q$ -expansion at  $\infty$ , which is exactly the sum  $\sum_{n \geq 0} a_n q^n$  that one obtains above when pulling back the form to the Tate curve at  $\infty$ . Hence we can sometimes prove that two modular forms are equal if we know that they have equal  $q$ -expansions.

We shall in fact be considering some slightly more general situations, where we will be taking sections of slightly modified sheaves.

Recall that  $Y_1(Np; p)_{\mathbf{Z}[1/Np]}$  parameterises elliptic curves equipped with an embedding  $i$  of  $\mu_N$ , a point  $P$  of exact order  $p$ , and a cyclic subgroup  $C$  of order  $p$  not containing  $P$ . Let  $\pi_2 : Y_1(Np; p)_{\mathbf{Z}[1/Np]} \rightarrow Y_1(Np)_{\mathbf{Z}[1/Np]}$  be the map which quotients out by the subgroup  $C$ , and consider the pullback  $(\pi_2)^* \omega^{\otimes k}$  on  $Y_1(Np; p)_{\mathbf{Z}[1/Np]}$ , where we write  $\omega$  for  $\omega_{Y_1(N)}$ , the pushforward of the differentials on the universal elliptic curve. Now a calculation on universal curves shows that a section of  $\pi_2^* \omega^{\otimes k}$  can be thought of as an element of  $H^0(E_1(Np; p)/C, (\Omega_{(E_1(Np; p)/C)/Y_1(Np; p)}^1)^{\otimes k})$ , where  $C$  is the universal cyclic subgroup of  $E_1(Np; p)$  of order  $p$ . Moreover, one can check that a given section extends to the cusps by evaluating the corresponding rule on the Tate curves at various cusps, and moreover the  $q$ -expansion principle still holds, when correctly interpreted in this setting.

One further generalisation that we need to make is that we shall sometimes be working in the category of rigid-analytic spaces instead of the category of schemes. In short, the analytic analogue of everything above still remains true, and details of many arguments can be found in [Con]. We summarise some of the facts that we need. Consider the curve  $Y_1(N)_K$ , where  $K$  is any complete subfield of  $\mathbf{C}_p$ . A global section of  $(\omega^{\text{an}})^{\otimes k}$  can be thought of as an element of  $H^0(E_1(N)^{\text{an}}, ((\Omega_{E_1(N)/Y_1(N)}^1)^{\otimes k})^{\text{an}})$  where  $E_1(N)$  is the universal elliptic curve over  $Y_1(N)$ . In this optic, the Tate curve at infinity can be thought of as the pullback of  $E_1(N)$  to the punctured disc consisting of the points of  $Y_1(N)_K$  which reduce to the cusp  $\infty$  in  $X_1(N)_{\overline{\mathbf{F}}_p}$ , and the other Tate curves can be thought of in a similar way. Again, one can verify that a section of  $(\omega^{\text{an}})^{\otimes k}$  extends to the cusps by computing  $q$ -expansions on these Tate curves. Finally, we have a  $q$ -expansion principle, saying that a section of  $(\omega^{\text{an}})^{\otimes k}$  is determined

by its  $q$ -expansion at  $\infty$ . One simple way of seeing this latter fact is that if an analytic function vanishes on a small disc then it must vanish everywhere, and one can apply this to any small disc contained in the punctured disc at infinity.

Next, we must extend the above paragraphs to certain admissible subsets of  $Y_1(N)^{\text{an}}$ . So let  $\mathcal{U}$  be a connected admissible open subset of  $X_1(N)^{\text{an}}$ , containing the cusp  $\infty$ . Choose a small disc  $\mathcal{U}_\infty$  in  $\mathcal{U}$  which contains  $\infty$ , and then remove  $\infty$  from it. The Tate curve at infinity in this setting can be thought of as the pullback of  $E_1(N)^{\text{an}}$  to  $\mathcal{U}_\infty$ , and one can deal with the other cusps in  $\mathcal{U}$  in the same way. An element of  $(\omega^{\text{an}})^{\otimes k}(\mathcal{U} - \{\text{cusps}\})$  can be thought of as an element of  $H^0(E_1(N) \times_{Y_1(N)} (\mathcal{U} - \{\text{cusps}\}), ((\Omega_{E_1(N)/Y_1(N)}^1)^{\otimes k})^{\text{an}})$ , one can check to see that it extends to  $\mathcal{U}$  by a computation on Tate curves, and finally, because  $\mathcal{U}$  is connected, we see that a section is determined by its pullback to  $\mathcal{U}_\infty \setminus \{\infty\}$  and so the  $q$ -expansion principle holds.

Finally, we observe that one can consider other sheaves such as  $(\pi_2^* \omega^{\otimes k})^{\text{an}}$  in this setting as well, and can think of sections of this sheaf over rigid analytic open subsets of  $(X_1(Np; p)_K)^{\text{an}}$  as rigid differentials on a curve isogenous to the universal curve, as in the algebraic case.

### 3 The theory of the canonical subgroup.

In this section we review and extend the theory of the canonical subgroup. Many of the results here appear in [K] (for the canonical subgroup of order  $p$ ) and [G] (for the extension to canonical subgroups of order  $p^n$ ). These latter references put mild assumptions on  $p$ , which we remove.

As before, let  $p$  be a prime and let  $N \geq 5$  be an integer prime to  $p$ . Let  $W$  denote the ring of Witt vectors of  $\mathbf{F}_p$ , and let  $K_0$  be its field of fractions. Let  $K$  be any finite field extension of  $K_0$ , and let  $\mathcal{O}_K$  denote the integers of  $K$ . We shall identify the residue fields of  $K$  and  $W$  with  $\mathbf{F}_p$ . Normalise the norm on  $K$  so that  $|p| = p^{-1}$ . We know that  $K$  is complete with respect to this norm, and hence there is a theory of rigid analytic spaces over  $K$ . We shall use the foundations of rigid analysis as set up in [BGR]. The only notation that we shall use here that is not defined in [BGR] is that of a *wide-open* subspace of a smooth curve, which is defined in [Col2] in the generality that we need. For the purposes of this paper, a wide-open space is a rigid-analytic space over  $K$  that is isomorphic to an admissible open in a smooth proper curve over  $K$  whose complement is, after a finite extension of  $K$  if necessary, isomorphic to a disjoint union of a non-zero finite number of affinoid discs.

The non-cuspidal points on  $X_1(N)_K$  parameterise elliptic curves (with extra structure), and in this section we shall explain something about the region of  $X_1(N)_K$  corresponding to the elliptic curves with supersingular reduction. This region is neither Zariski-open nor closed, but it is an admissible open subset of the rigid analytic space  $X_1(N)^{\text{an}}$  over  $K$  associated to  $X_1(N)_K$ , and we shall use rigid analysis to understand this region more clearly.

There are a finite non-zero number of supersingular points in the special fibre of  $X_1(N)_{\mathcal{O}_K}$ , that is, points which correspond to supersingular elliptic curves equipped with a point of order  $N$ . Let  $x \in X_1(N)_{\mathcal{O}_K}(\overline{\mathbf{F}}_p) = X_1(N)_W(\overline{\mathbf{F}}_p)$  be a supersingular point. Then the completion of the local ring  $\mathcal{O}_{X_1(N)_W, x}$  is a  $W$ -algebra which is non-canonically isomorphic to  $W[[T_x]]$  for some parameter  $T_x$ . Fix such an isomorphism. The subset  $B_x$  of  $X_1(N)^{\text{an}}$  consisting of points which reduce to  $x$  is a rigid space, isomorphic to the open unit disc, and the choice of  $T_x$  can be thought of as a choice of a parameter on this disc. Indeed, we can think of  $T_x$  as an isomorphism of rigid spaces from  $B_x$  to the open unit disc in  $K$ . The fact that  $B_x$  is indeed a disc seems to be known to the experts, although a lot of the literature on this subject appears to only consider the case when the base field  $K$  is algebraically closed (which is clearly not true here). One reference which proves everything in the generality we need is Propositions 0.3.5 and 0.2.7 of [Be].

If  $0 < r \leq 1$  is of the form  $p^q$  with  $q$  a rational number, then define  $B_{x, < r}$  to be the subdisc of  $B_x$  consisting of points  $P \in B_x$  such that  $|T_x(P)| < r$ . In general this disc will depend on the choice of  $T_x$ . However, the only  $W$ -algebra automorphisms of  $W[[T_x]]$  send  $T_x$  to  $ap + uT_x$ , where  $a \in W$  and  $u \in W[[T_x]]$  is a unit, and this implies that for  $r > 1/p$  the disc  $B_{x, < r}$  only depends on  $x$  and  $r$ .

For  $1/p < r = p^q \leq 1$  as above, define  $X_1(N)_{\geq r}$  to be the complement of  $\cup_{x \in SS} B_{x, < r}$  in  $X_1(N)^{\text{an}}$ , where  $SS$  denotes the (finite non-empty) set of supersingular points. Then  $X_1(N)_{\geq r}$  is the intersection of finitely many affinoids in a separated space and is hence affinoid. We shall refer to  $X_1(N)_{\geq 1} = X_1(N)^{\text{an}} \setminus \cup_{x \in SS} B_x$  as the *ordinary locus* of  $X_1(N)^{\text{an}}$  and to the  $X_1(N)_{\geq r}$  for  $r < 1$  as a system of *strict neighbourhoods* of  $X_1(N)_{\geq 1}$ . If  $\omega^{\text{an}}$  denotes the analytification of the sheaf  $\omega$  on  $X_1(N)_K$ , and if  $k$  is an integer, then the Banach space  $H^0(X_1(N)_{\geq 1}, (\omega^{\text{an}})^{\otimes k})$  is the space of  $p$ -adic modular forms of weight  $k$ , and the union of the subspaces  $H^0(X_1(N)_{\geq r}, (\omega^{\text{an}})^{\otimes k})$  for  $r < 1$  as above is the space of overconvergent  $p$ -adic forms of weight  $k$ .

If  $N \leq 4$  then there is still a theory, but one must be slightly more careful because the sheaf  $\omega$  may not exist. In this case, one may proceed by adding an auxiliary level structure, working at higher level, and then taking invariants. The structure of the supersingular discs might be slightly more complicated if  $N \leq 4$ , and we shall avoid the problem completely by working only at level  $N \geq 5$ , and using the trick above to deal with smaller levels when necessary.

We now briefly explain how this approach to  $p$ -adic modular forms relates to the one described by Katz in [K] and by Gouvêa in [G]. Katz and Gouvêa use the Eisenstein series  $E_{p-1}$  to give a parameter near each supersingular point. Because  $E_{p-1}$  is not a true modular form when  $p \leq 3$ , Katz has to put some extra conditions on  $N$  in these cases, and Gouvêa avoids small primes completely. If  $E_{p-1}$  exists then it is a section of  $\omega^{\otimes(p-1)}$  on  $X_1(N)_W$ . If  $x$  is a supersingular point then choose an open affine  $U \subseteq X_1(N)_W$  over which  $\omega^{\otimes(p-1)}$  becomes isomorphic to  $\mathcal{O}_U$ . Then  $E_{p-1}$  can be thought of as a section of  $\mathcal{O}_U(U)$  and it is known that  $E_{p-1} \bmod p$  is the Hasse invariant, which has a simple zero



at  $x$ . Hence the image of  $E_{p-1}$  in  $\mathcal{O}_{X_1(N)_{W,x}}$  can be used as a choice for  $T_x$ . Our approach of choosing  $T_x$  arbitrarily has the advantage that it avoids these problems when  $p \leq 3$ . As far as we know, this approach was first adopted in [BT] but presumably it was known before then.

We explain how some of the theorems on the canonical subgroup in chapter 3 of [K] can be applied to (integral models of) elliptic curves corresponding to points in  $X_1(N)_{\geq r}$  when  $r = p^{-p/(p+1)}$ . In fact the extra conditions imposed by Katz on the level structure when  $p \leq 3$  can be removed, in the cases that we are interested in. We firstly recall some of the results on the canonical subgroup proved in [K], results which Katz attributes to Lubin.

Let  $L$  be a finite extension of  $K$  and let  $E/L$  be an elliptic curve, equipped with an embedding  $i : \mu_N \rightarrow E$  defined over  $L$ . Then the pair  $(E, i)$  defines an  $L$ -valued point  $\alpha$  of  $X_1(N)_K$ . We define a rational number  $v(E, i)$  associated (non-canonically) to the pair  $(E, i)$  (the definition is non-canonical because it depends on our choices of parameters  $T_x$ ). Note first that  $N \geq 5$  and hence  $E$  cannot have additive reduction. If  $E$  has multiplicative or good ordinary reduction then define  $v(E, i) = 0$ . If however  $E$  has supersingular reduction, then let  $x \in X_1(N)(\mathbf{F}_p)$  be the point corresponding to the reduction of  $(E, P)$ . Recall that we have fixed a parameter  $T_x$  on the disc  $B_x$  of points above  $x$ . Define  $|(E, i)| = |T_x(\alpha)|$  and define  $v(E, i)$  by  $|(E, i)| = p^{-v(E, i)}$ . Note that if  $|(E, i)|$  is zero then technically  $v(E, i) = +\infty$  is not defined, but our arguments will only concern elliptic curves corresponding to points near the edge of the supersingular discs and in particular we shall never be concerned with the exact value  $v(E, i)$  if it is at least 1. We remark here that if  $v(E, i) < p/(p+1)$  then one can use Theorem 3.10.7 of [K] to verify that  $v(E, i)$  is determined by the formal group of  $E$  and is hence independent of  $i$  and of  $T$ . Note that the calculation at this point of [K] does not depend on the fact that  $p \geq 5$ . In fact, although Katz has to be slightly careful when  $p \leq 3$  in his definition of the canonical subgroup for an elliptic curve over a  $p$ -adic ring, Theorem 3.10.7 of [K] can be used to deduce that if we restrict ourselves to elliptic curves defined over the integers of a finite extension of  $K_0$ , everything is independent of the level structure that we chose. What seems to be happening is that Katz is actually defining a rigid-analytic function on the supersingular discs, such that  $v(E, i)$  is the valuation of the value of this function at  $(E, i)$ . When  $p$  is less than 5, the actual value of the function could depend on the choice of  $i$ , but the valuation of this value does not, at least near the boundary of the disc. See also the forthcoming definition of  $v'$ , which is in general perhaps a more canonical measure of the ‘‘supersingularity’’ of  $E$ , but which depends on the choice of an auxiliary subgroup of order  $p$ .

**Definition 3.1.** *Let  $L/K$  be a finite extension, and say  $E/L$  is an elliptic curve. Say that  $E$  is not too supersingular if, after a finite extension  $L'$  of  $L$  if necessary, there exists an embedding  $i : \mu_N \rightarrow E$  such that  $v(E, i) < p/(p+1)$  (If this is the case for one embedding then it will be true for all embeddings).*

Note that this condition is also independent of both the choice of  $i$  and of  $T$ . If  $E$  fails to be not too supersingular, then we shall say that  $E$  is *too*

*supersingular.* Note that elliptic curves with multiplicative or good ordinary reduction are not too supersingular.

*Remark 3.2.* Geometrically speaking, the supersingular but not too supersingular elliptic curves are the ones near the outer edges of the supersingular discs  $B_x$ .

If  $E$  is not too supersingular, then  $v(E, i)$  is independent of the finite field extension  $L'$  and the choice of  $i$ , and we call it  $v(E)$ . In [K], Katz explains how to associate a *canonical subgroup* to an elliptic curve which is not too supersingular. If  $p \leq 3$  then Katz must choose some auxiliary level structure before he can define a canonical subgroup. But because we are only working with elliptic curves over finite extensions of  $K_0$ , Theorem 3.10.7 of [K] shows us that the canonical subgroup will be independent of this choice. We remark also that Katz attributes this work to Lubin.

We now recall some of the main results about canonical subgroups in this language.

**Theorem 3.3 (Katz, Lubin).** *For a not too supersingular elliptic curve  $E/L$ , let  $H = H(E) \subset E$  be its canonical subgroup. Then  $H$  has order  $p$ . Moreover,*

- (i) *If  $v(E) = 0$  then  $H$  is the finite étale subgroup of  $E$  corresponding to the kernel of the reduction map from  $E[p](\bar{L})$  to the  $p$ -torsion in the Néron model of  $E$  over  $\mathcal{O}_L$ .*
- (ii) *If  $v(E) < 1/(p+1)$  then  $E/H$  is not too supersingular, and  $v(E/H) = pv(E)$ .*
- (iii) *If  $v(E) = 1/(p+1)$  then  $E/H$  is too supersingular.*
- (iv) *If  $v(E) \in (1/(p+1), p/(p+1))$ , then  $E/H$  is again not too supersingular,  $v(E/H) = 1 - v(E)$ , and furthermore the canonical subgroup of  $E/H$  is  $E[p]/H$ .*
- (v) *If  $v(E) < p/(p+1)$  and  $C \neq H$  is a subgroup of  $E$  of order  $p$ , then  $v(E/C) = v(E)/p$  and the canonical subgroup  $H(E/C)$  of  $E/C$  is  $E[p]/C$ .*
- (vi) *If  $v(E) \geq p/(p+1)$  and  $C \subset E$  is a subgroup of order  $p$ , then  $v(E/C) = 1/(p+1)$  and the canonical subgroup  $H(E/C)$  of  $E/C$  is  $E[p]/C$ .*

*Proof.* All these results other than (iii) are contained in either the beginning of §3.4 of [K], or Theorem 3.10.7 of [K]. We deduce (iii) as follows. If  $v(E) = 1/(p+1)$  then by the proof of [K] 3.10.7(2) we see that there are points on  $(E/H)[p]$  with ordinal  $1/(p^2 - 1)$ . But [K] 3.10.7(1) shows that this cannot happen if  $E/H$  is not too supersingular. □

We now extend the definition of a canonical subgroup to canonical subgroups of order  $p^n$  for  $n > 1$ , as in [G].

Let  $E/L$  be an elliptic curve, which is not too supersingular. If  $n \geq 1$  is an integer, and  $v(E) < p^{2-n}/(p+1)$ , then define a subgroup  $H_n = H_n(E)$  of  $E$

as follows. The definition is inductive. If  $n = 1$  then set  $H_1(E) = H(E)$ , the canonical subgroup of  $E$ . For  $n > 1$ , let  $H_n(E)$  be the preimage of  $H_{n-1}(E/H)$  under the natural projection  $E \rightarrow E/H$ . Note that we used part (ii) of Theorem 3.3 above to show that  $H_{n-1}(E/H)$  makes sense. Moreover, it is a simple exercise, using part (v) of Theorem 3.3, to show that  $H_n$  is cyclic of order  $p^n$ .

**Definition 3.4.** *If  $E$  and  $n$  are as above, with  $v(E) < p^{2-n}/(p+1)$ , then call  $H_n = H_n(E)$  the canonical subgroup of order  $p^n$  of  $E$ . More compactly, if  $E$  is not too supersingular and  $v(E) < p^{2-n}/(p+1)$ , we shall say that  $E$  has a canonical subgroup, or  $E$  has an  $H_n$  for brevity.*

The main result we need on these higher canonical subgroups is below. Before we state the result, we establish some notation. Say  $E/L$  is an elliptic curve such that  $E$  has an  $H_n$ . Let  $m \geq 0$  be an integer, and let  $C_m$  denote any cyclic subgroup of  $E$  of order  $p^m$ , such that  $C_m \cap H_n = 0$ . Finally, say  $D_{m+n}$  is a cyclic subgroup of order  $p^{m+n}$  in  $E$ , such that  $H_n \subseteq D_{m+n}$ .

**Proposition 3.5.** *With notation as above,  $v(E/C_m) < p^{2-(m+n)}/(p+1)$ , so  $E/C_m$  has an  $H_{m+n}$ . Moreover,  $H_{m+n}(E/C_m) = (D_{m+n} + C_m)/C_m$ .*

*Proof.* This can no doubt be proved in the same way as Katz proves Theorem 3.10.7 of [K], but we shall instead show how it can be deduced from Theorem 3.3.

Repeated applications of (v) of Theorem 3.3 gives us that  $E/C_m$  has an  $H_{m+n}$ . So it remains to verify that it is the image of  $D_{m+n}$ .

The Proposition is vacuous for  $m = 0$ . Say we have proved it for  $m = 1$ . Then by induction on  $m$ , the Proposition is true. Hence it suffices to show the result for  $m = 1$ . In this case we have to show that the  $H_{1+n}(E/C_1)$  is the image of  $D_{1+n}$ . By definition,  $H_{1+n}(E/C_1)$  is the preimage in  $E/C_1$  of the  $H_n$  of the quotient of this curve by its canonical subgroup. Hence by Theorem 3.3 (v),  $H_{1+n}$  is the preimage of  $H_n(E)$  under the map  $E/C_1 \rightarrow E/E[p] \cong E$ . So  $H_{1+n}(E/C_1)$  is the preimage of  $(D_{1+n} + E[p])/E[p]$ , which equals  $(D_{1+n} + E[p])/C_1 = (D_{1+n} \oplus C_1)/C_1 = D_{1+n}$ . □

*Geometric remarks.* If  $E$  has an  $H_n$  then this means geometrically that either  $E$  has good ordinary or multiplicative reduction, or that  $E$  is supersingular but is very close to the boundary of the supersingular disc, in some sense. The proposition above shows that if one quotients out a not too supersingular elliptic curve by a “non-canonical” subgroup of order  $p^m$ , that is, one which has trivial intersection with the canonical subgroup, then it becomes even less supersingular. Conversely, by Theorem 3.3, if one quotients out a not too supersingular elliptic curve by its canonical subgroup, one usually ends up with an elliptic curve which is “more supersingular”, and if one continues to do this then it could happen that after a while the curve becomes too supersingular and hence does not have a canonical subgroup. The geometry of this situation will be clarified in the next section.

## 4 Admissible subspaces of modular curves.

Let  $X_1(N; p)^{\text{an}}$  denote the rigid space associated to  $X_1(N; p)_K$ . Let  $\infty$  denote the cusp of  $X_1(N; p)$  corresponding to the Tate curve  $E_q := \mathbf{G}_m/q^{\mathbf{Z}}$  equipped with the natural embedding of  $\mu_N$  into  $\mathbf{G}_m$  and the subgroup of  $p$ th roots of unity in  $\mathbf{G}_m$ . The reduction mod  $p$  of this cusp is a smooth point on one of the components of  $X_1(N; p)_{\mathbf{F}_p}$ ; call this component  $X_\infty$ . The non-cuspidal  $\overline{\mathbf{F}_p}$ -valued points on  $X_\infty$  correspond to elliptic curves over  $\overline{\mathbf{F}_p}$  equipped with a point of order  $N$  and a finite flat subgroup scheme of order  $p$  which only has one geometric point.

Let  $\mathcal{W}_0(p) \subset X_1(N; p)^{\text{an}}$  denote the set of points whose reduction lies in  $X_\infty$ . Then  $\mathcal{W}_0(p)$  is an admissible open subset of  $X_1(N; p)^{\text{an}}$ , and is in fact a wide-open space, as can be seen for example by checking that the proof of Proposition 3.3 of [Coll] does not use the fact that the base field is  $\mathbf{C}_p$ . We shall now define a function  $v' : \mathcal{W}_0(p) \rightarrow \mathbf{Q}$  which is related to  $v$  but has the advantage that it does not depend on the choice of uniformisers at each supersingular point.

Let  $x \in \mathcal{W}_0(p)$  be a point. If  $x$  is a cusp then define  $v'(x) = 0$ . Otherwise,  $x$  corresponds to a closed point of  $Y_1(N; p)_K$  whose reduction lies in  $X_\infty$ . In particular, there is a finite extension  $L/K$ , and an elliptic curve  $E/L$  equipped with an embedding  $i$  of  $\mu_N$  and a subgroup  $C$  of order  $p$ , such that the data  $(E, i, C)$  corresponds to  $x$ . Now define  $v'(x)$  as follows. If  $E$  is ordinary then define  $v'(x) = 0$ . If  $E$  has supersingular reduction but  $E$  is not too supersingular, and  $C$  happens to be the canonical subgroup of  $E$ , then define  $v'(x) = v(E)$ . If  $v(E, i) \geq p/(p+1)$  and hence  $E$  has no canonical subgroup, then define  $v'(x) = p/(p+1)$ . Finally, if  $E$  has supersingular reduction, and  $E$  has a canonical subgroup but it is not  $C$ , then by part (v) of Theorem 3.3, we have  $v(E/C) < 1/(p+1)$ , and we set  $v'(x) = 1 - v(E/C, i) \in (p/(p+1), 1)$ . A case by case check shows that  $v'$  does not depend on the choice of uniformising parameters at the supersingular points (we used these choices to define  $v$  but our definitions were independent of choice near the edges of the discs). Moreover,  $v'$  does not depend on  $i$ , and hence we can speak about  $v'(E, C)$  when we mean  $v'(x)$  for some point  $x$  corresponding to  $(E, i, C)$  for some embedding  $i$  of  $\mu_N$ .

A more concise definition of  $v'$  is the following: away from the supersingular region,  $v' = 0$ . In the supersingular region,  $v'(E, C) = v(E)$  if  $C$  is the canonical subgroup of  $E$ , and in general  $v'(E, C) + v'(E/C, E[p]/C) = 1$ , where we note that either  $C$  or  $E[p]/C$  are canonical by part (vi) of Theorem 3.3.

We explain another way of thinking about the function  $v'$ . On a supersingular annulus,  $v'$  looks rather like the valuation of an isomorphism between the annulus and the subset  $\{1/p < |z| < 1\}$  of affine 1-space. In fact, things are slightly more complicated, but not overly so. Let  $q$  be a rational number such that  $1/2 < q < p/(p+1)$ , and let  $r = p^{-q}$ . Consider the preimage of  $X_1(N)_{\geq r}$  under  $\pi_1$  in  $X_1(N, p)^{\text{an}}$ . This preimage is two disjoint affinoids. Moreover, the component  $X_1(N, p)_{\geq r}$  of this preimage containing the cusp  $\infty$  maps isomorphically down to  $X_1(N)_{\geq r}$  by  $\pi_1$  again, by the theory of the canonical subgroup.

The pullbacks of the parameters  $T_x$  that we chose on the supersingular discs of  $X_1(N)^{\text{an}}$  give us a function on the supersingular part of  $X_1(N, p)_{\geq r}$ , and on a point in this region,  $v'$  agrees with the valuation of this function.

Things are equally simple at the other end of the annulus: if we consider the preimage of  $X_1(N)_{\geq r}$  under  $\pi_2$ , then the component not containing  $\infty$  is an affinoid mapping one-to-one down to  $X_1(N)_{\geq r}$ , and pulling back the functions  $T_x$  onto the supersingular annuli in this affinoid gives us a function defined on these annuli, such that  $1 - v'$  agrees with the valuation of this function. So “at either end” of the supersingular annuli in  $X_1(N, p)$ , the function  $v'$  looks like the valuation of an isomorphism to an explicit annulus in the affine line, and part (iv) of Theorem 3.3 guarantees that things match up in the middle of the annuli.

For  $n \geq 0$  define  $V_n \subset \mathcal{W}_0(p)$  to be the points  $x$  of  $\mathcal{W}_0(p)$  such that  $v'(x) \leq 1 - 1/p^{n-1}(p+1)$ .

**Proposition 4.1.** *All of the  $V_n$  are admissible open affinoid subsets of the curve  $X_1(N; p)^{\text{an}}$ . Moreover, all the  $V_n$  are connected. We have  $V_0 \subset V_1 \subset V_2 \subset \dots$ , and the union of the  $V_i$  is  $\mathcal{W}_0(p)$ . Moreover, the  $V_i$  form an admissible cover of  $\mathcal{W}_0(p)$ .*

*Proof.* One painless way of seeing this is as follows. That  $V_0$  is an affinoid is standard—in fact  $V_0$  is isomorphic to  $X_1(N)_{\geq p^{-1}/(p+1)}$ . For  $n \geq 1$ , we consider a supersingular annulus in  $\mathcal{W}_0(p)$ , and equip it with the function  $\pi_2^*(T_x)$ , where  $T_x$  is our parameter on the corresponding supersingular disc in  $X_1(N)^{\text{an}}$ . Near the “far” end of the annulus, this function is an isomorphism onto an annulus in affine 1-space, and so we can use it to glue a disc onto  $\mathcal{W}_0(p)$ , as in the proof of Proposition 3.3 (ii) of [Coll]. Doing this to all the supersingular annuli in  $\mathcal{W}_0(p)$  we construct a proper rigid space, for which  $V_n$  are affinoid subspaces obtained by removing finitely many open discs. Now everything follows without too much difficulty.  $\square$

We now prove a result about these affinoids which shall be important in the sequel.

**Lemma 4.2.** *Let  $L/K$  be a finite field extension.*

1. *Let  $x \in X_1(N; p)_K(L)$  be a point corresponding to the data  $(E, i, C)$ , such that for some  $n \geq 2$  we have  $v(E, i) \in [1/p^{n-2}(p+1), p/(p+1))$ . Then  $x \in V_n$ .*
2. *Let  $E/L$  be an elliptic curve equipped with  $i$  and  $C$  as usual, corresponding to the point  $x \in V_{n+1}$  for some  $n \geq 0$ . Let  $D$  be any cyclic subgroup of order  $p$  of  $E$  such that  $D \neq C$ , and let  $y$  correspond to the data  $(E/D, i \bmod D, C + D/D)$ . Then  $y \in V_n$ .*

*Proof.* 1) First we observe that  $E$  has a canonical subgroup, so let  $H$  be this subgroup. If  $H = C$  then  $v'(x) < p/(p+1) < 1 - 1/p^{n-1}(p+1)$  and we are home. If not, then  $v(E/C, i) = (1/p)v(E, i) \geq 1/p^{n-1}(p+1)$  by assumption and

part (v) of Theorem 3.3, and so by definition we have  $v'(x) \leq 1 - 1/p^{n-1}(p+1)$  and again we are home.

2) We consider first the case  $n = 0$ . In this case,  $v'(x) \leq p/(p+1)$ , and the theorem is an easy corollary of Theorem 3.3, as follows. There are two possibilities: Either  $v'(x) < p/(p+1)$ , in which case  $E$  is not too supersingular and we use part (v) of the theorem, or  $v'(x) = p/(p+1)$ , so  $E$  has no canonical subgroup and we use part (vi).

Next we deal with the case  $n = 1$ . It suffices to prove the result for  $x \in V_2 \setminus V_1$ . In this case,  $v'(x) \in (1 - 1/(p+1), 1 - 1/p(p+1))$  and so  $v(E/C, \bar{i}) \in [1/p(p+1), 1/(p+1))$ . Our assumption that  $x \notin V_1$  implies that  $E$  is not too supersingular, and that  $E$  has a canonical subgroup  $H$  such that  $H \neq C$ . We now consider two subcases. Firstly if  $v(E/C) > 1/p(p+1)$  then by part (v) of Theorem 3.3 we see that  $v(E, i) \in (1/(p+1), p/(p+1))$ , and so  $E/D$  has a canonical subgroup which is  $E[p]/D = C + D/D$  by (iv) of Theorem 3.3. So in this case,  $(E/D, \bar{i}, C + D/D) \in V_1$ . We are left with the case  $v(E/C, i) = 1/p(p+1)$ . In this case, we deduce from Theorem 3.3 that  $v(E) = 1/(p+1)$ . There are now two cases. If  $D = H$  then  $v(E/D) \geq p/(p+1)$  and we deduce that  $(E/D, \bar{i}, C + D/D) \in V_1$ . Finally, if  $D \neq H$  we have  $v(E/D, \bar{i}) = 1/p(p+1)$  and then  $(E/D, \bar{i}, C + D/D) \in V_0$ .

Finally we deal with the case  $n \geq 2$ . We may assume that  $x \in V_{n+1} \setminus V_n$ , and so  $v'(x) \in (1 - 1/p^{n-1}(p+1), 1 - 1/p^n(p+1))$ . We deduce that  $v(E/C) \in [1/p^n(p+1), 1/p^{n-1}(p+1))$  and so  $E$  has a canonical subgroup  $H$ . Because  $v'(x) \neq v(E)$  we must have  $C \neq H$ , and we see that  $v(E) \in [1/p^{n-1}(p+1), 1/p^{n-2}(p+1))$ . Hence either  $v(E/D) \in [1/p^n(p+1), 1/p^{n-1}(p+1))$  or  $v(E/D) \in [1/p^{n-2}(p+1), 1/p^{n-3}(p+1))$ . In the former case,  $D \neq H$  so  $(E/D)$  has a canonical subgroup, namely the image of  $C$ , and  $(E/D, \bar{i}, C + D/D) \in V_1 \subseteq V_n$ . In the latter case we apply part 1 of the Lemma to deduce that  $(E/D, \bar{i}, C + D/D) \in V_n$ .  $\square$

We now move to a higher level. Let  $m$  be a positive integer and let  $\pi : X_1(Np^m)_K \rightarrow X_1(N; p)_K$  be the map which on non-cuspidal points corresponds to the morphism of moduli problems sending  $(E, i, P)$  to  $(E/\langle pP \rangle, \bar{i}, \langle \bar{P} \rangle)$ . Set  $\mathcal{W}_1(p^m) = (\pi^{\text{an}})^{-1}(\mathcal{W}_0(p))$  and for  $n \geq 0$  set  $Z_n = (\pi^{\text{an}})^{-1}(V_n) \subset X_1(Np^m)^{\text{an}}$ . Note that if  $x \in X_1(Np^m)$  corresponds to  $(E, i, P)$ , then  $x$  is in  $\mathcal{W}_1(p^m)$  iff  $E$  has either got supersingular reduction, or  $E$  has ordinary reduction and  $P$  reduces to the identity in characteristic  $p$ . So  $\mathcal{W}_1(p^m)$  corresponds to the preimage of a characteristic  $p$  “outer component” in the Katz-Mazur canonical balanced model of  $X_1(Np^m)$  over  $\mathbf{Z}_p[\zeta_{p^m}]$ . Note that  $\mathcal{W}_1(p^m)$  contains all the supersingular locus, which is in general rather complicated geometrically for  $m \geq 2$ . Indeed, already in the case of  $X_0(Np^2)$ , Edixhoven’s arguments in [E] show that the special fibre of a stable model for  $X_0(Np^2)$  contains hyperelliptic curves of high genus which reduce to supersingular points in the Katz-Mazur model.

**Proposition 4.3.** *The subsets  $\mathcal{W}_1(p^m)$  and  $Z_n$  of  $X_1(Np^m)^{\text{an}}$  are all admissible open. The space  $\mathcal{W}_1(p^m)$  is a connected wide-open, and the  $Z_n$  are connected affinoids for  $n \geq 0$ . We have  $Z_0 \subset Z_1 \subset Z_2 \dots$ , and the  $Z_n$  form an admissible cover of  $\mathcal{W}_1(p^m)$ .*

*Proof.* This follows essentially immediately from Proposition 4.1. Connectedness comes from the fact that the ordinary locus of  $\mathcal{W}_1(p^m)$  is connected, which follows from Proposition 3.3 (i) of [Col1] applied to the canonical balanced model of  $X_1(Np^m)$  over  $\mathbf{Z}_p[\zeta_{p^m}]$ . Indeed, we see that  $\mathcal{W}_1(p^m)$  is the preimage of a component of this model, and the ordinary locus of  $\mathcal{W}_1(p^m)$  is contained in  $Z_0$ .  $\square$

Recall that there are morphisms  $\pi_1$  and  $\pi_2 : X_1(Np^m; p)_K \rightarrow X_1(Np^m)_K$ , where  $\pi_1$  forgets the subgroup of order  $p$  and  $\pi_2$  quotients out by it. We can now state the main result of this section:

**Proposition 4.4.** *We have  $(\pi_1^{\text{an}})^{-1}(Z_{n+1}) \subseteq (\pi_2^{\text{an}})^{-1}(Z_n)$  for all  $n \geq 0$ .*

*Proof.* The maps  $\pi_1$  and  $\pi_2$  are finite, and hence the claimed inclusions are between affinoids. So it suffices to check on non-cuspidal points, where the result is immediate from part 2 of Lemma 4.2.  $\square$

## 5 Correspondences on rigid spaces and the first analytic continuation results.

In this section we explain a construction the Hecke operator  $U_p$  in the rigid-analytic category, and observe that it gives us a rather easy proof that certain eigenvectors for  $U_p$  satisfy certain analytic continuation properties.

Let  $C$  and  $D$  be smooth proper geometrically irreducible algebraic curves over  $K$ , and let  $f : C \rightarrow D$  be a finite flat  $K$ -morphism. If  $\mathcal{F}$  is an invertible sheaf on  $D$ , then there is a natural  $\mathcal{O}_D$ -linear trace map  $f_* f^* \mathcal{F} \rightarrow \mathcal{F}$ , defined locally using the fact that if  $A$  is Noetherian and  $B$  is a finite flat  $A$ -algebra then there is an  $A$ -linear trace map  $B \rightarrow A$ .

Now say  $g : C \rightarrow D$  is another  $K$ -morphism, and that we are given a map of sheaves of  $\mathcal{O}_C$ -modules  $g^* \mathcal{F} \rightarrow f^* \mathcal{F}$ . This map induces a map  $f_* g^* \mathcal{F} \rightarrow f_* f^* \mathcal{F}$  and by composition with the trace map we obtain a map  $f_* g^* \mathcal{F} \rightarrow \mathcal{F}$ .

Now let us analytify the situation. There are rigid spaces  $C^{\text{an}}$  and  $D^{\text{an}}$  associated to  $C$  and  $D$ . The coherent sheaf  $\mathcal{F}$  on  $D$  gives rise to a coherent analytic sheaf  $\mathcal{F}^{\text{an}}$  on  $D^{\text{an}}$ , and the morphisms  $f$  and  $g$  give rise to morphisms  $f^{\text{an}}$  and  $g^{\text{an}}$ . Moreover, one can define the analytic pushforward, or pullback, of a sheaf of  $\mathcal{O}^{\text{an}}$ -modules, and check that  $(f_* g^* \mathcal{F})^{\text{an}}$  is canonically isomorphic to  $f_*^{\text{an}} g^{\text{an},*} \mathcal{F}^{\text{an}}$ . In particular, we have a natural map  $f_*^{\text{an}} g^{\text{an},*} \mathcal{F}^{\text{an}} \rightarrow \mathcal{F}^{\text{an}}$  and hence, for all admissible open  $U$  in  $D^{\text{an}}$ , a map  $(f_*^{\text{an}} g^{\text{an},*} \mathcal{F}^{\text{an}})(U) \rightarrow \mathcal{F}^{\text{an}}(U)$ .

By definition, we have  $(f_*^{\text{an}} g^{\text{an},*} \mathcal{F}^{\text{an}})(U) = (g^{\text{an},*} \mathcal{F}^{\text{an}})(f^{-1}(U))$ , and hence if  $V$  is any admissible open subset of  $D^{\text{an}}$  such that  $f^{-1}U \subseteq g^{-1}V$ , then we have natural maps

$$\mathcal{F}^{\text{an}}(V) \rightarrow (g^{\text{an},*} \mathcal{F}^{\text{an}})(g^{-1}V) \xrightarrow{\text{res}} (g^{\text{an},*} \mathcal{F}^{\text{an}})(f^{-1}U) \rightarrow \mathcal{F}^{\text{an}}(U).$$

Let us simply refer to the composite of these maps as “the trace map”.

We now apply this theory to the subspaces of the modular curves defined in the previous section to deduce some analytic continuation results. As far as

we are aware, these analytic continuation results are new, although they could have been noticed and proved 25 years ago.

Recall the sheaf  $\omega$  on  $X_1(Np^m)_K$ , which on non-cuspidal points is the push-forward of the differentials on the universal elliptic curve. It is standard that if  $k$  is an integer then the global sections of  $\omega^{\otimes k}$  are modular forms of weight  $k$ . In the notation used at the beginning of this section, set  $C = X_1(Np^m; p)_K$  and  $D = X_1(Np^m)_K$ . Set  $f = \pi_1$  and  $g = \pi_2$ . Finally, let  $k$  be any integer, and define  $\mathcal{F} = \omega^{\otimes k}$ . We deduce from Proposition 4.4 and the remarks at the beginning of this section that we have, for every  $n \geq 0$ , a trace map  $tr : (\omega^{\text{an}})^{\otimes k}(Z_n) \rightarrow (\omega^{\text{an}})^{\otimes k}(Z_{n+1})$ .

We can see this morphism explicitly as follows. Let  $x$  be a point in  $Z_n$ . Then  $x$  corresponds to an elliptic curve  $E'$  over a finite extension  $L$  of  $K$ , equipped with an embedding of  $\mu_N$  and a point of order  $p^m$ . The fibre of  $\omega^{\otimes k, \text{an}}$  at  $x$  is canonically isomorphic to  $H^0(E', \Omega^1)^{\otimes k}$  and hence one can think of  $f(x)$  as being an element of  $H^0(E', \Omega^1)^{\otimes k}$ . Similarly if  $y$  is a point of  $Z_{n+1}$  corresponding to  $(E, i, P)$  then the value of  $tr(f)$  at  $y$  can be thought of as an element of  $H^0(E, \Omega^1)^{\otimes k}$ . We can actually compute  $tr(f)$  explicitly, as follows. For all cyclic subgroups  $C$  of order  $p$  in  $E$  such that  $C$  does not contain  $p^{m-1}P$ , the point  $(E/C, \bar{i}, \bar{P})$  corresponds to a point of  $Z_n$ , by Theorem 4.4. Let  $pr : E \rightarrow E/C$  denote the projection. Then one can check that

$$tr(f)(E, i, P) = \sum_C (pr)^*(f(E/C, \bar{i}, \bar{P}))$$

where  $(pr)^*$  denotes pulling back  $k$ -fold differentials, and the sum is over all  $C$  of order  $p$  not containing  $p^{m-1}P$ .

The Tate curve  $E_q := \mathbf{G}_m/q^{\mathbf{Z}}$  can be thought of as an elliptic curve over the base  $\mathbf{Z}((q))$  or alternatively, if  $q \in \bar{K}$  is an element of norm less than 1, as an elliptic curve over a finite extension of  $K$ . The Tate curve has a canonical differential which we shall call  $dt/t$ , where  $t$  can be thought of as a parameter on  $\mathbf{G}_m$ . From now on,  $q$  will be an element of  $\bar{K}$  with norm less than 1.

If  $i : \mu_N \rightarrow \mathbf{G}_m$  is the canonical inclusion, and  $P \in \mathbf{G}_m$  is our fixed  $p^m$ th root of unity, then  $(E_q, i, P)$  corresponds to a point on  $X_1(Np^m)^{\text{an}}$  close to the cusp  $\infty$ . Evaluating a weight  $k$  modular form  $f$  on  $(E_q, i, P)$  gives us an element of  $H^0(E_q, \Omega_{E_q}^1)^{\otimes k}$  of the form  $c(q)(dt/t)^k$ , and  $c(q)$  is an analytic function of  $q$ . In fact  $c(q) = \sum_{n \geq 0} a_n q^n$  if  $f$  is holomorphic at the cusp  $\infty$ , and  $c(q)$  is said to be the  $q$ -expansion of  $f$ .

**Proposition 5.1.** *The natural trace map  $(\omega^{\text{an}})^{\otimes k}(Z_n) \rightarrow (\omega^{\text{an}})^{\otimes k}(Z_{n+1})$ , calculated on  $q$ -expansions, is the map  $\sum a_n q^n \mapsto p \sum a_{np} q^n$ . In other words, the trace map is just an analytic manifestation of the Hecke operator  $pU_p$ .*

*Proof.* This is now an easy corollary of the equation above and the theory of the Tate curve. If the  $q$ -expansion of  $f$  is  $\sum a_n q^n$  then  $f$  evaluated at  $(E_q, i, \zeta_{p^m})$  is  $\sum a_n q^n (dt/t)^k$  for any  $q$  with  $|q| < 1$ . Hence the  $q$ -expansion of  $tr(f)$  is  $b(q)$



where

$$\begin{aligned}
b(q)(dt/t)^k &= \text{tr}(f)(E_q, i, \zeta_{p^m}) \\
&= \sum_C (pr)^* f(E_q/C, \bar{i}, \zeta_{p^m}) \\
&= \sum_{j=0}^{p-1} (pr)^* f(\mathbf{G}_m / \langle \zeta_p^j q^{1/p} \rangle, \bar{i}, \zeta_{p^m}) \\
&= \sum_{j=0}^{p-1} (pr)^* \sum_{n \geq 0} a_n (\zeta_p^j q^{1/p})^n (dt/t)^k \\
&= \sum_{j=0}^{p-1} \sum_{n \geq 0} a_n \zeta_p^{nj} q^{n/p} (dt/t)^k \\
&= p \sum_{n \geq 0, p|n} a_n q^{n/p} (dt/t)^k \\
&= p \sum_{r \geq 0} a_{rp} q^r (dt/t)^k.
\end{aligned}$$

□

Recall that  $\mathcal{W}_1(p^m) = \bigcup_{n \geq 0} Z_n \subset X_1(Np^m)^{\text{an}}$ .

**Theorem 5.2 (analytic continuation).** 1. *If  $m \geq 1$  is an integer, and  $h$  is an overconvergent modular form of weight  $k$  and level  $Np^m$ , which is an eigenvector for  $U_p$  with non-zero eigenvalue, then  $h$  extends to a section of  $(\omega^{\text{an}})^{\otimes k}$  on  $\mathcal{W}_1(p^m)$ .*

2. *If  $f$  is an overconvergent modular form of level  $N$  and weight  $k$  which is an eigenvector for  $U_p$  with non-zero eigenvalue, then  $f$  extends to a section of  $(\omega^{\text{an}})^{\otimes k}(\mathcal{W}_0(p))$ .*

*Proof of theorem.* 1) By definition,  $h$  is a section of  $\omega^{\otimes k}$  over some region of  $X_1(Np^m)$  which strictly contains the component of the ordinary locus containing infinity. Because  $U_p$  preserves overconvergence, it is well-known that  $h$  extends to a section of  $\omega^{\otimes k}$  on  $Z_0$ . Let the eigenvalue of  $U_p$  on  $h$  be  $\lambda$ . Now by induction define  $h_r \in (\omega^{\text{an}})^{\otimes k}(Z_r)$  for all  $r \geq 0$  by  $h_0 = h$  and  $h_{r+1} = \text{tr}(h_r)/p\lambda$ . By Proposition 5.1, the  $q$ -expansion of  $h_r$  is equal to the  $q$ -expansion of  $h$ . Now because  $Z_{\min\{r,s\}}$  is connected, a section of  $(\omega^{\text{an}})^{\otimes k}$  on  $X_{\min\{r,s\}}$  is determined by its  $q$ -expansion. Hence  $h_r$  and  $h_s$  agree on  $Z_{\min\{r,s\}}$ . Finally,  $\mathcal{W}_1(p^m)$  is admissibly covered by the  $Z_n$ . Hence the  $h_r$  glue to give a section of  $(\omega^{\text{an}})^{\otimes k}$  on  $\mathcal{W}_1(p^m)$ .

2) One could use similar techniques to prove this part, but in fact it can be deduced from part 1 by thinking of  $f$  as having level  $Np$ , extending  $f$  to a section of  $\omega^{\otimes k}$  on  $\mathcal{W}_1(p)$  and then using the fact that  $f$  is invariant under the Diamond operators at  $p$  to descend  $f$  to  $\mathcal{W}_0(p)$ . □

*Remarks.* • We have worked for simplicity with points of order  $N$ , but the result, and the proof, are valid for modular curves associated to any sufficiently fine level  $N$  structure.

- This result seems stronger than any analytic continuation result of this form in the literature, in the sense that the continuation goes further than any other results that we are aware of.
- Of course, an overconvergent form of level  $N$  with non-zero  $U_p$ -eigenvalue cannot in general be extended over the supersingular discs in  $X_1(N)$ , as this would imply that it was classical by rigid GAGA ([Kö]). Also, a classical newform of level  $Np^2$  with trivial character at  $p$  can be regarded as overconvergent of level  $N$ , and will have  $U_p$ -eigenvalue equal to 0. Then Theorem 5.2 will not apply, and indeed one cannot expect  $f$  to extend to  $\mathcal{W}_0(p)$  because in general the value of  $f$  on a supersingular elliptic curve will depend on the subgroup of order  $p^2$ .

## 6 More analytic continuation.

We do not know how to improve on the results of Section 5 in the level  $Np$  cases, that is, we do not know whether one can extend eigenforms beyond the supersingular locus. But if we are working on level  $Np^m$  with  $m > 1$  then, rather surprisingly, we can do even better, and the object of this section is to explain how to do so.

Let  $0 \leq r < m$  be integers (we allow  $m = 1$  for the time being, but only for notational reasons). We define  $\mathcal{U}_r$  to be the wide-open subset of  $X_1(Np^m)^{\text{an}}$  whose non-cuspidal locus parameterises  $(E/L, i, P)$  where either  $E$  has supersingular reduction, or  $E$  has ordinary reduction and  $H_{m-r}(E) = \langle p^r P \rangle$ . In other words, the non-cuspidal locus of  $\mathcal{U}_r$  parameterises elliptic curves with additional structure  $(E/L, i, P)$  such that the order of the reduction of  $P$  has order at most  $p^r$ . It is clear that  $\mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \dots \subseteq \mathcal{U}_{m-1}$ , and moreover all the  $\mathcal{U}_i$  are connected because they are preimages of connected closed subsets of the canonical balanced Katz-Mazur model of  $X_1(Np^m)$  over  $\mathbf{Z}_p[\zeta_{p^m}]$ . Moreover,  $\mathcal{U}_0$  is just the rigid space  $\mathcal{W}_1(p^m)$ . Let  $\mathcal{U}_1(p^m)$  denote the union of the  $\mathcal{U}_i$  for  $0 \leq i \leq m-1$ . Then the  $\mathcal{U}_i$  admissibly cover  $\mathcal{U}_1(p^m)$ .

As in the previous section, let  $\pi_1$  and  $\pi_2 : X_1(Np^m; p) \rightarrow X_1(Np^m)$  denote the two degeneracy maps.

**Lemma 6.1.** *For  $0 \leq r \leq m-2$  we have  $\pi_1^{-1}(\mathcal{U}_{r+1}) \subseteq \pi_2^{-1}(\mathcal{U}_r)$ .*

*Proof.* It suffices to check on non-cuspidal points. We first unravel the lemma. Let  $(E, i, P)$  be an elliptic curve over a finite extension of  $K$ , equipped with an embedding of  $\mu_N$  and a point of exact order  $p^m$ . Assume that  $E$  has an  $H_{m-r-1}$  and that it is generated by  $p^{r+1}P$ . Because  $r \leq m-2$  we deduce that  $E$  has an  $H$  and it is generated by  $p^{m-1}P$ .

Now let  $D \subset E$  be a cyclic subgroup of order  $p$ , such that  $D$  intersects  $H$  trivially. Then by Proposition 3.5 we see that  $E/D$  has an  $H_{m-r}$  and furthermore that it is generated by the image of  $p^r P$ . This proves that the Lemma holds on the ordinary locus. On the supersingular locus the Lemma is clear, because all the supersingular locus is contained in all of the  $\mathcal{U}_i$ .  $\square$

**Corollary 6.2 (more analytic continuation).** *If  $m \geq 2$  and  $f$  is an overconvergent modular form of level  $Np^m$ , such that  $U_p f = \lambda f$  for some  $\lambda \neq 0$ , then  $f$  has an analytic continuation to all of  $\mathcal{U}_1(p^m)$ .*

*Proof.* Exactly the same techniques as before prove this result. We have already proved that  $f$  extends to a section of  $\mathcal{U}_0$ , and then for  $0 \leq i \leq m-1$  the section  $U_p^i f / \lambda^i$  is a section of  $\mathcal{U}_i$  which agrees with  $f$  on  $\mathcal{U}_j$  for  $j < i$ , as  $\mathcal{U}_j$  is connected.  $\square$

The complement of  $\mathcal{U}_1(p^m)$  in  $X_1(Np^m)^{\text{an}}$  is an affinoid, because it is the preimage of an affine open in the Katz-Mazur model of  $X_1(Np^m)_{\mathcal{O}_K}$ . In fact, if  $\bar{X}_0$  denotes the “outer” component of  $X_1(Np^m)_{\bar{\mathbb{F}}_p}$  not containing the cusp  $\infty$ , then the complement of  $\mathcal{U}_1(p^m)$  is precisely the preimage of the ordinary locus of  $\bar{X}_0$ .

## 7 Application of results of Wiles and others.

In this section and the next we explain how, given a representation as in the Main Theorem of the introduction, one can construct some  $p$ -adic modular objects (weight one overconvergent  $p$ -adic modular forms). Useful references are §1 of [BT] and §§2–3 of [BDST].

Assume we are given  $\rho : \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\mathcal{O})$  satisfying the hypotheses of the Main Theorem in the introduction. In particular we have  $\rho|_{D_p} \cong \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$  for characters  $\alpha$  and  $\beta : \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow \mathcal{O}^\times$  which are distinct modulo  $\lambda$  and such that  $\alpha(I_p)$  and  $\beta(I_p)$  are finite.

Let  $\beta_0 : \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow \mathcal{O}^\times$  denote the character of finite order and conductor a power of  $p$  such that  $\beta_0|_{I_p} = \beta|_{I_p}$ . Let  $\bar{\rho}_{\beta_0}$  denote the mod  $\lambda$  reduction of  $\rho \otimes \beta_0^{-1}$ . Let  $\Sigma$  be a finite set of primes containing  $p$  and all the primes where  $\rho$  is ramified. Consider deformations  $\tilde{\rho}_\beta$  of  $\bar{\rho}_{\beta_0}$  to complete noetherian local  $W(\mathcal{O}/\lambda)$ -algebras with residue field  $\mathcal{O}/\lambda$ , which are unramified outside  $\Sigma$  and satisfy

$$\tilde{\rho}_\beta|_{D_p} \cong \begin{pmatrix} * & * \\ 0 & \psi \end{pmatrix}$$

with  $\psi$  an unramified lifting of  $(\beta/\beta_0)|_{D_p}$ .

It is a theorem of Mazur that there is a universal such deformation, and a theorem of Diamond (for  $p > 2$ ) and Dickinson (for  $p = 2$ ), building on work of Wiles and Taylor-Wiles, that under the assumptions of our theorem, the universal deformation ring is isomorphic to a  $\Lambda$ -adic Hecke algebra of the

type considered by Hida. In particular, this shows us that to the representation  $\rho \otimes \beta_0^{-1}$ , which is a deformation of  $\bar{\rho}_\beta$  of the type considered above, there corresponds a homomorphism from this Hecke algebra to  $\mathcal{O}^\times$ .

We say a little about this Hecke algebra. Define  $\Lambda = \mathbf{Z}_p[[T]]$ . Then there is an integer  $N$ , a finite extension  $\mathcal{L}$  of the field of fractions of  $\Lambda$ , a height 1 prime  $\wp$  of the integral closure  $\mathcal{O}_\mathcal{L}$  of  $\Lambda$  in  $\mathcal{L}$ , a Hecke algebra  $h^0(N)$  of the type considered by Hida, and a  $\Lambda$ -adic modular form  $G : h^0(N) \rightarrow \mathcal{L}$  such that the reduction mod  $\wp$  of the Galois representation to  $\mathrm{GL}_2(\mathcal{O}_\mathcal{L})$  attached to  $G$  is isomorphic to  $\rho \otimes \beta_0^{-1}$ . A reference for more of the details of this construction is §1 of [BT].

Switching the roles of  $\alpha$  and  $\beta$ , and applying the deformation theory techniques to the mod  $\lambda$  reduction of  $\rho \otimes \alpha_0^{-1}$  gives us a second Hida eigenform  $F$ , called a “ $\Lambda$ -adic companion form” for  $G$  in [BT].

Let  $R$  denote the ring  $\mathcal{O}_\mathcal{L}/\wp$  and let  $f = \sum_{n>0} a_n q^n$  denote the mod  $\wp$  reduction of  $F$ , that is, the power series in  $R[[q]]$  satisfying  $a_n = F(T(n)) \bmod \wp$ . Similarly, let  $g = \sum_{n>0} b_n q^n$  denote the mod  $\wp$  reduction of  $G$ . Next, let  $p^m$  be the conductor of the character  $\alpha_0/\beta_0$ , or  $p$  if this character is unramified.

We have defined  $f$  and  $g$  as formal  $q$ -expansions. Let us assume that we have proved that the  $q$ -expansion of  $f$  is the  $q$ -expansion of a classical modular form of weight 1 and level  $Np^m$ . Then by construction we have that  $f$  is an eigenform, and  $b_l$  equals the trace of  $\rho \otimes \alpha_0^{-1}(\mathrm{Frob}_l)$  for almost all primes  $l$ , and hence the representation associated to  $f$  by Deligne and Serre will be isomorphic to  $\rho \otimes \alpha_0^{-1}$ . Now twisting  $f$  will show that  $\rho$  is also modular, and the Main Theorem of the introduction will be proved. In the final sections of this paper, we explain various methods that allow us to show in certain cases that the  $q$ -expansion of  $f$  is that of a classical modular form, and these sections complete the proof of the main theorem.

## 8 Geometric interpretation of $f$ and $g$ .

In this section we explain how one can interpret the formal  $q$ -expansions  $f$  and  $g$  as  $p$ -adic analytic objects, and hence how one can use the methods of earlier sections of this paper to manipulate them.

The Hida families which arose in the deformation theory arguments above can be interpreted in terms of the theory of  $p$ -adic modular forms. We shall use the very geometric interpretation of this theory, which Coleman has developed in the last few years. In Coleman’s language, the  $q$ -expansions of the  $p$ -adic forms  $f$  and  $g$  introduced above are  $q$ -expansions of overconvergent  $p$ -adic modular forms of level  $Np^m$ . We shall explain more about what this means, and what it implies, below.

Firstly, let us simplify the situation slightly by dealing with bad primes. So let  $l$  be a prime dividing  $N$ , and assume that  $f = \sum a_n q^n$  with  $a_l \neq 0$ . Then the form  $f(q) - a_l f(q^l)$  is an overconvergent  $p$ -adic eigenform  $f'$  of level  $Nl$  with  $q$ -expansion  $\sum_{l \nmid n} a_n q^n$ ; moreover,  $f'$  is an eigenvector of  $U_l$  with eigenvalue zero. By modifying  $f$  and  $g$  in this way and increasing  $N$ , if necessary, we can assume

that  $f = \sum a_n q^n$  and  $g = \sum b_n q^n$  with  $a_n = b_n = 0$  if  $n$  is not coprime to  $N$ .

Now we do not yet know that  $f$  and  $g$  are classical weight 1 modular forms. But because  $f$  and  $g$  lie in Hida families, we do know that their  $q$ -expansions are limits of  $q$ -expansions of classical modular eigenforms of level  $Np^m$ . Hence by a theorem of Coleman, we see that  $f$  and  $g$  are overconvergent  $p$ -adic modular forms, that is, they are rigid-analytic sections of  $\omega$  on a strict neighbourhood of the component of the ordinary locus of  $X_1(Np^m)^{\text{an}}$  containing  $\infty$ . Perhaps the best reference for this result of Coleman is currently §2 of [BDST], which deals only with the unramified case. However, the arguments of this section go through with little difficulty. In particular, an appropriate generalisation of Lemma 2.9 of [BDST] proves that  $f$  and  $g$  are overconvergent, and because  $U_p$  increases overconvergence, we see that  $f$  and  $g$  can be thought of as rigid sections of  $\omega$  on  $Z_0$ . Moreover, we have proved in Theorem 5.2 that any such section may be analytically extended to all of  $\mathcal{U} = \mathcal{U}_0(p)$  or  $\mathcal{U}_1(p^m)$  as appropriate, and we shall now think of  $f$  and  $g$  as being sections of  $\omega$  defined on  $\mathcal{U}$ .

Let us assume from now on that  $K$  contains all the  $p^m$ th roots of unity, and let us fix a  $p^m$ th root of unity  $\zeta_{p^m}$  in  $K$ . We have already mentioned the automorphism  $w$  of  $X_1(Np^m)$  defined over  $K$ : for an elliptic curve  $E/S/\text{Spec}(K)$  equipped with an embedding  $i : \mu_N \rightarrow E$  and a point  $P \in E(S)$  of exact order  $p^m$ , we define  $w(E, i, P) = (F, \bar{i}, Q)$  where  $\phi : E \rightarrow F := E/\langle P \rangle$  and  $Q \in \ker(\phi)(S) \subset F(S)$  is the point such that  $\langle P, Q \rangle_\phi = \zeta_{p^m}$ . This morphism  $w$  extends to the cusps and induces an automorphism of  $X_1(Np^m)_K$  and hence of the rigid space  $X_1(Np^m)^{\text{an}}$ . Note that with notation as above, we have  $\langle P, Q \rangle_\phi \langle Q, P \rangle_{\bar{\phi}} = 1$  and hence  $w^2(E, i, P) = w(E, \bar{p}i, -P)$ .

Now  $f$  is defined in a neighbourhood of  $\infty$ , and  $w^*g$  is a rigid section of  $w^*\omega$  defined in a neighbourhood of  $w(\infty)$ . Because  $p$  is invertible on the base, the natural map  $w^*\omega \rightarrow \omega$  induced by pulling back differentials on the universal curve is an isomorphism. Let  $g|w$  denote the rigid section of  $\omega$  defined by the image of  $w^*g$  under this natural map. Then  $g|w$  is defined in a neighbourhood of  $w\infty$ .

More precisely, we have extended  $f$  and  $g$  until they are sections of  $\omega$  on the wide-open space  $\mathcal{U} \subset X$ , where  $\mathcal{U} = \mathcal{W}_0(p)$  and  $X = X_1(N; p)$  if  $\alpha/\beta$  is unramified, and  $\mathcal{U} = \mathcal{U}_1(p^m)$  and  $X = X_1(Np^m)$  if  $\alpha/\beta$  is ramified of conductor  $p^m$ . It is not hard to see that in either case,  $\mathcal{U}$  is the complement of the component of the ordinary locus of  $X$  containing the cusp  $w(\infty)$ . Moreover, the automorphism  $w$  sends  $w(\infty)$  to a cusp on the same connected component of the ordinary locus as  $\infty$ . It is not too difficult to deduce from this that  $\{\mathcal{U}, w\mathcal{U}\}$  is an admissible covering of  $X$ .

We now summarise what we know about the  $q$ -expansions of  $f$  and  $g$ . We have that

$$f = \sum a_n q^n, \text{ and } g = \sum b_n q^n$$

with

- $a_1 = b_1 = 1$ , and  $a_n = b_n = 0$  if  $n$  and  $N$  are not coprime,
- $a_l = \text{trace}(\rho(\text{Frob}_l))/\alpha_0(l)$  and  $b_l = \text{trace}(\rho(\text{Frob}_l))/\beta_0(l)$  if  $l$  is a prime not dividing  $Np$ ,

- $a_n \alpha_0(n) = b_n \beta_0(n)$  if  $n$  is prime to  $Np$ ,
- The character of  $f$  at  $p$  is  $\beta_0/\alpha_0$ , and the character of  $g$  at  $p$  is  $\alpha_0/\beta_0$ . The characters of  $f$  and  $g$  away from  $p$  are equal.
- $a_p = (\alpha/\alpha_0)(\text{Frob}_p)$  and  $b_p = (\beta/\beta_0)(\text{Frob}_p)$ , and  $a_p b_p = \chi^p(p)$ , where  $\chi^p$  is the character of  $f$  (and  $g$ ) away from  $p$ .

These facts can be read off from what is known about the Galois representations attached to Hida families of modular forms. See section 1 of [BT], for example.

Next we come to the “glueing” procedure. The claim here is that there is a constant  $c \in K$  such that  $f$  and  $cg|w$  agree on  $\mathcal{U} \cap w\mathcal{U}$ . Assuming this, we can finish the proof of the theorem. For  $f$  and  $cg|w$  can be glued on this intersection to give a rigid section of  $\omega^{\text{an}}$  on all of  $X$ , and by rigid GAGA ([Kö]) we deduce that this section is in fact an algebraic section, that is, an element of  $H^0(X_{K(\zeta_p^m)}, \omega)$ .

To finish the proof of the theorem, we just have to show that  $c$  exists such that  $f$  and  $cg|w$  agree on  $\mathcal{U} \cap w\mathcal{U}$ . In the case where  $\alpha/\beta$  is unramified there is already a proof of this in [BT] but we shall give another proof in the next section. In the case where  $\alpha/\beta$  is tamely ramified but not unramified, we shall glue using a new method, explained in Section 10. These two methods do not in fact need the full strength of the analytic continuation techniques that have been developed so far in this paper. Finally, we shall prove the Main Theorem in the case that  $\alpha/\beta$  is wildly ramified, using another method which relies strongly on the analytic continuation theorems. In fact, one of the motivations of this paper was to present these new glueing methods in the hope that they will be useful more generally.

## 9 The unramified case.

In this section we shall explain another proof of the theorem in [BT], although in fact it is perhaps just another way of looking at the same proof. Note that we shall explain all our glueing results for overconvergent forms of a general integer weight  $k$ , although in all the applications in this paper we will have  $k = 1$ .

**Theorem 9.1.** *Let  $f = \sum a_n q^n$  and  $g = \sum b_n q^n$  be cuspidal overconvergent  $U_p$ -eigenforms of weight  $k$  and level  $N$ , where  $N \geq 5$  is an integer prime to  $p$ . Assume that*

- $a_1 = b_1 = 1$ .
- $a_p$  and  $b_p$  are distinct and non-zero.
- $a_n = b_n$  for all  $n$  prime to  $p$ .

*Then  $f$  and  $g$  are classical forms of level  $Np$ .*

*Proof.* Because  $a_p$  and  $b_p$  are non-zero, we know by Theorem 5.2 that  $f$  and  $g$  both have analytic continuation to all of  $\mathcal{W}_0(p)$ . It suffices to prove that there is a constant  $c$  such that  $f$  and  $cg|w$  agree on the supersingular locus of  $\mathcal{W}_0(p)$ .

Let  $W$  denote a supersingular annulus in  $\mathcal{W}_0(p)$ , lying over a supersingular disc  $B$  in  $X_1(N)$ . Choose a rational number  $q$  with  $p/(p+1) < q < 1$  and consider the subannulus  $A \subset B$  defined by  $|T| > p^{-q}$ , where  $T$  is our fixed parameter on  $B$ . The preimage of  $A$  under the map  $\pi_2 : W \rightarrow B$  is two annuli—in fact it is the annuli defined by  $0 < v' < p(1-q)$  and  $q < v' < 1$ , where  $v'$  is the function on  $W$  defined earlier. Let  $C$  denote the annulus  $0 < v' < p(1-q)$ , considered as a subannulus of  $W$ . Then we have a commutative diagram

$$\begin{array}{ccc} C & \hookrightarrow & W \\ \downarrow \pi_2 & & \downarrow \pi_2 \\ A & \hookrightarrow & B \end{array} ,$$

where the horizontal maps are inclusions, and the vertical ones are induced by  $\pi_2$ . Note that the left hand vertical map is finite and flat of degree  $p$ , and the right hand one is finite and flat of degree  $p+1$ .

If we define  $f_0 = (a_p f - b_p g)/(a_p - b_p)$  then  $f_0$  is an overconvergent section of  $\omega^{\otimes k}$  and it suffices to prove that it extends to  $B$ , because then  $f_0$  will be classical of level  $N$ , and  $f = f_0(q) - b_p f_0(q^p)$  and  $g = f_0(q) - a_p f_0(q^p)$  will be classical of level  $Np$ . This is in fact what we shall prove.

Choose a trivialisation of  $\omega^{\otimes k}$  on  $B$ , and note that this induces a trivialisation of  $\omega^{\otimes k}$  on  $A$  and of  $\pi_2^* \omega^{\otimes k}$  on  $W$  and on  $C$ . The natural map from  $\pi_2^* \omega^{\otimes k}$  to  $\omega^{\otimes k}$  is an isomorphism, because  $p$  is invertible in this situation, and we identify  $\pi_2^* \omega^{\otimes k}$  with  $\omega^{\otimes k}$ . Now consider  $f_1 = p \langle p \rangle_N (f - g)/(a_p - b_p)$ , thought of as a section of  $\omega^{\otimes k}$  on  $C$ . An easy calculation on Tate curves shows that  $f_1 = \pi_2^* f_0$ , and the fact that  $f$  and  $g$  extend to  $\mathcal{W}_0(p)$  means that  $f_1$  extends to a section of  $\omega^{\otimes k}$  on  $W$ . Our fixed trivialisations mean that we may consider  $f_0$  and  $f_1$  as functions on  $A$  and  $W$ , whose pullbacks to  $C$  are equal. Taking fields of fractions of everything gives us a diagram

$$\begin{array}{ccc} K(C) & \hookleftarrow & K(W) \\ \uparrow p & & \uparrow p+1 \\ K(A) & \hookleftarrow & K(B) \end{array}$$

of fields, where the vertical maps are finite field extensions of degrees  $p$  and  $p+1$ , and all maps are inclusions. It suffices to prove that  $K(A) \cap K(W) = K(B)$  as subfields of  $K(C)$ , as this will establish that  $f_0$  extends to  $B$ . Set  $L = K(A) \cap K(W)$  and  $M = K(A).K(W)$ . Firstly note that certainly  $K(W)$  is not contained within  $K(A)$ , as there are injective functions from  $W$  to the affine line, and any function on  $A$  is not injective when pulled back to  $C$ . Hence  $M \neq K(A)$  and so must be  $K(C)$ , as  $[K(C) : K(A)] = p$  is prime. We deduce that  $[K(W) : L] \geq p$ . But  $[K(W) : K(B)] = p+1$  and  $L$  contains  $K(B)$ , hence  $L = K(B)$ .  $\square$

We remark that in fact the full strength of Theorem 5.2 is not in fact needed in this result, as we can replace  $B$  by a disc of radius slightly larger

than  $p^{-p/(p+1)}$ , replace  $A$  by an annulus at the boundary of  $B$ , and replace  $C$  and  $W$  by appropriate preimages of  $A$  and  $B$ , where  $f$  and  $g$  were already known to extend, and the argument still works.

## 10 The tame case.

In this section we explain another method for glueing the forms  $f$  and  $g|w$  which works when the character  $\alpha/\beta$  is tamely ramified but not unramified. Perhaps surprisingly, this case seems to be the trickiest case, the wild case being rather easy because of our strong analytic continuation results.

**Theorem 10.1.** *Let  $f$  and  $g$  be overconvergent cusp forms of level  $Np$  and weight  $k$ , which are eigenforms for the Diamond operators and such that  $f$  has character  $\chi = \chi_N \chi_p : (\mathbf{Z}/Np\mathbf{Z})^\times \rightarrow K$  and  $g$  has character  $\chi_N \chi_p^{-1}$ . Assume that  $f = \sum_n a_n q^n$  and  $g = \sum b_n q^n$  are eigenforms for  $U_p$  with non-zero eigenvalues  $a_p$  and  $b_p$ . Finally, assume that*

- $a_1 = b_1 = 1$  and  $a_n = \chi_p(n)b_n$  for all  $n$  prime to  $p$ ,
- $\chi_p$  is non-trivial, and
- $a_p b_p = \chi_N(p)p^{k-1}$ .

Then  $f$  and  $g$  are classical eigenforms of level  $Np$ .

*Proof.* By Theorem 5.2,  $f$  and  $g$  extend to sections of  $\omega^{\otimes k}$  on  $\mathcal{W}_1(p)$ . We shall glue  $f$  and  $cg|w$  for some constant  $c$ , and this will suffice to prove the theorem. Recall that the curve  $X_1(Np; p)_{\mathbf{Z}[1/Np][\zeta_p]}$  can be thought of as the completion of the curve parameterising elliptic curves equipped with an embedding of  $\mu_N$  and two points of order  $p$ ,  $P$  and  $Q$ , which generate  $E[p]$  and which pair to  $\zeta_p$ . For this section then, let us assume that our base contains  $\zeta_p$ , a fixed  $p$ th root of unity.

We have several maps  $X_1(Np; p)^{\text{an}} \rightarrow X_1(Np)^{\text{an}}$  defined by the following maps of moduli problems:

$$\pi : (E, i, P, Q) \mapsto (E, i, P)$$

and for  $0 \leq j \leq p-1$ ,

$$\sigma_j : (E, i, P, Q) \mapsto (E/\langle jP + Q \rangle, \bar{i}, \bar{P}).$$

Define  $T = \pi^{-1}(\mathcal{W}_1(p)) \subset X_1(Np; p)^{\text{an}}$ . Then  $T$  is a wide-open space, and the non-cuspidal points of  $T$  parameterise elliptic curves  $(E, i, P, Q)$  such that either  $E$  has supersingular reduction, or  $E$  has ordinary reduction and  $H(E) = \langle P \rangle$ . It is easily checked that  $\pi$  and  $\sigma_j$  all induce maps  $T \rightarrow \mathcal{W}_1(p)$ . Note that  $T$  is connected because it is the preimage of a geometrically connected component of the Katz-Mazur model of  $X_1(Np; p)$  over  $\mathbf{Z}_p[\zeta_p]$ , which has reduced special fibre. Note also that there is a natural sheaf  $\omega = \omega_{X_1(Np; p)}$  on  $X_1(Np; p)^{\text{an}}$



and hence on  $T$ . Moreover  $\pi^*\omega = \omega$  (with apologies for the two meanings of  $\omega$ ) and there is a natural map  $\sigma_j^*\omega \rightarrow \omega$  induced on the non-cuspidal locus by pulling back differentials on the universal curves. Hence  $\pi$  and  $\sigma_j$  induce maps of modular forms  $H^0(\mathcal{W}_1(p), \omega^{\otimes k}) \rightarrow H^0(T, \omega^{\otimes k})$ , denoted  $h \mapsto h|\pi$  and  $h \mapsto h|\sigma_j$  respectively. We now compute what these maps do on  $q$ -expansions.

**Sublemma.** *If  $h \in H^0(\mathcal{W}_1(p), \omega^{\otimes k})$  has  $q$ -expansion equal to  $\sum_{n \geq 0} d_n q^n (dt/t)^k$  at the cusp  $(\mathbf{G}_m/q^{\mathbf{Z}}, i, \zeta_p)$  and  $c \in T$  denotes the cusp  $(\mathbf{G}_m/q^{\mathbf{Z}}, i, \zeta_p, q^{1/p})$ , then  $(h|\pi)(c) = \sum d_n q^n (dt/t)^k$  and  $(h|\sigma_j)(c) = \sum d_n \zeta_p^{jn} q^{n/p} (dt/t)^k$ .*

*Proof of sublemma.* The statement for  $\pi$  is obvious. For  $\sigma_j$  we note that

$$\begin{aligned} (h|\sigma_j)(c) &= pr^*h(\mathbf{G}_m/\langle \zeta_p^j q^{1/p} \rangle, \bar{i}, \zeta_p) \\ &= \sum d_n (\zeta_p^j q^{1/p})^n (dt/t)^k \\ &= \sum d_n \zeta_p^{jn} q^{n/p} (dt/t)^k. \end{aligned}$$

□

Hence

$$\begin{aligned} \left( \sum_{j=1}^{p-1} \chi_p(j) f|\sigma_j \right)(c) &= \sum_{n \geq 1} a_n q^{n/p} \sum_{j=1}^{p-1} \zeta_p^{jn} \chi_p(j) (dt/t)^k \\ &= W(\chi_p) \sum_{n \geq 1, p \nmid n} a_n / \chi_p(n) q^{n/p} (dt/t)^k \end{aligned}$$

where  $W(\chi_p) = \sum_{j=1}^{p-1} \zeta_p^j \chi_p(j) = \chi_p(n) \left( \sum_{j=1}^{p-1} \zeta_p^{jn} \chi_p(j) \right)$ . So

$$\begin{aligned} \left( \sum_{j=1}^{p-1} \chi_p(j) f|\sigma_j \right)(c) &= W(\chi_p) \sum_{n \geq 1, p \nmid n} b_n q^{n/p} (dt/t)^k \\ &= W(\chi_p) \left( \sum_{n \geq 1} b_n q^{n/p} - \sum_{m \geq 1} b_{mp} q^m \right) (dt/t)^k \\ &= W(\chi_p) (g|\sigma_0 - b_p g|\pi)(c) \end{aligned}$$

and by connectivity of  $T$  we deduce that

$$\left( \sum_{j=1}^{p-1} \chi_p(j) f|\sigma_j \right) = W(\chi_p) (g|\sigma_0 - b_p g|\pi) \quad (\dagger)$$

as sections of  $\omega^{\otimes k}$  on  $T$ . In particular, this relationship holds on the supersingular locus of  $T$  and it is this relation  $(\dagger)$  that we shall exploit to show that the forms  $f$  and  $g|w$  glue on the supersingular region of  $X_1(Np)$ .

If  $(E, i, P, Q)$  corresponds to a point in  $T$  then

$$\begin{aligned} pa_p f(E, i, Q) - pr^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) &= pU_p f(E, i, Q) - pr^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) \\ &= \sum_{C \subset E, C \neq \langle Q \rangle, C \neq \langle P \rangle} pr^* f(E/C, \bar{i}, \bar{Q}) \end{aligned}$$

where the sum is over subgroups  $C$  of  $E$  of order  $p$ . Hence we see

$$\begin{aligned} pa_p f(E, i, Q) - pr^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) &= \sum_{j=1}^{p-1} pr^* f(E/\langle jP + Q \rangle, \bar{i}, \bar{Q}) \\ &= \chi_p(-1) \sum_{j=1}^{p-1} \chi_p(j) pr^* f(E/\langle jP + Q \rangle, \bar{i}, \bar{P}) \\ &= \chi_p(-1) \sum_{j=1}^{p-1} \chi_p(j) f|_{\sigma_j}(E, i, P, Q) \end{aligned}$$

and hence using  $(\dagger)$  we deduce that

$$\begin{aligned} pa_p f(E, i, Q) - pr^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) &= \chi_p(-1) W(\chi_p)(g|\sigma_0(E, i, P, Q) - b_p g|\pi(E, i, P, Q)) \\ &= \chi_p(-1) W(\chi_p)(pr^* g(E/\langle Q \rangle, \bar{i}, \bar{P}) - b_p g(E, i, P)), \end{aligned}$$

valid in particular for any  $(E, i, P, Q)$  such that  $E$  has supersingular reduction (and such that the Weil pairing  $\langle P, Q \rangle_p$  equals  $\zeta_p$ , by definition of  $T$ ).

From now on, we only consider this equation (which we proved by looking near a cusp) as an identity only on elliptic curves with supersingular reduction. We now use the definition of  $g|w$  and the relation  $a_p b_p = \chi^p(p) p^{k-1}$  to deduce that

$$\begin{aligned} pa_p f(E, i, Q) - pr^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) &= \chi_p(-1) W(\chi_p)((g|w)(E, i, -Q) - (pa_p)^{-1} \chi^p(p) pr^*(g|w)(E/\langle P \rangle, p^{-1}\bar{i}, -\bar{Q})) \\ &= W(\chi_p)((g|w)(E, i, Q) - (pa_p)^{-1} pr^*(g|w)(E/\langle P \rangle, \bar{i}, \bar{Q})) \end{aligned}$$

and hence

$$\begin{aligned} pa_p f(E, i, Q) - W(\chi_p)(g|w)(E, i, Q) &= (pr)^* f(E/\langle P \rangle, \bar{i}, \bar{Q}) - W(\chi_p)(pa_p)^{-1} pr^*(g|w)(E/\langle P \rangle, \bar{i}, \bar{Q}). \quad (*) \end{aligned}$$

We are now finally in a position to glue  $f$  and  $g|w$ . Let  $SS$  denote the region in  $X_1(Np)^{\text{an}}$  corresponding to elliptic curves with supersingular reduction, and define  $h \in H^0(SS, \omega^{\otimes k})$  by

$$h = pa_p f - W(\chi_p)(g|w).$$

If we can prove that  $h$  is identically zero then we will be able to glue  $f$  and  $(pa_p)^{-1} W(\chi_p) g|w$  and we are home. But note that equation  $(*)$  above says

that if  $E$  is an elliptic curve with supersingular reduction and  $P, Q$  are points of order  $p$  such that  $\langle P, Q \rangle_p = \zeta_p$ , then

$$h(E, i, Q) = (pa_p)^{-1} pr^* h(E/\langle P \rangle, \bar{i}, \bar{Q}).$$

This implies that for  $E, P$  and  $Q$  as above, and  $n$  any integer,  $h(E, i, Q) = h(E, i, nP + Q)$ , as  $\langle P, nP + Q \rangle = \zeta_p$ . Now replacing  $P$  by  $P + mQ$  for any integer  $m$  we deduce that

$$h(E, i, Q) = h(E, i, nP + (1 + nm)Q).$$

Setting  $n = 1$  and  $m = -1$  we deduce that  $h(E, i, Q) = h(E, i, P)$ . Finally, choosing  $n \in \mathbf{Z}$  such that  $\chi_p(n) \neq 1$  and choosing  $m \in \mathbf{Z}$  such that  $nm \equiv -1 \pmod{p}$ , we deduce that  $h(E, i, nP) = h(E, i, Q) = h(E, i, P)$ . But  $h(E, i, nP) = \chi_p(n)h(E, i, P)$  because the same is true for  $f$  and  $g|w$ . Hence we must have  $h(E, i, P) = 0$  and  $h$  vanishes identically on the supersingular locus, as was to be proved. □

## 11 The wild case.

In this section we explain another method for glueing the forms that works in the wild case. The point is that in the wild case, we have analytically continued both  $f$  and  $g|w$  to areas whose intersection contains cusps.

**Theorem 11.1.** *Let  $f$  and  $g$  be overconvergent cusp forms of weight  $k$  and level  $Np^m$ , with  $m \geq 2$ , and assume that  $f$  and  $g$  are eigenforms for  $U_p$  with non-zero eigenvalues. Assume moreover that  $f$  and  $g$  are eigenforms for the Diamond operators at  $N$  and at  $p$ , with  $\chi_p : (\mathbf{Z}/p^m\mathbf{Z})^\times \rightarrow K$  the character of  $f$  at  $p$  and  $\chi_p^{-1}$  the character of  $g$  at  $p$ . Assume that  $\chi_p$  has conductor  $p^m$ . Write  $f = \sum a_n q^n$  and  $g = \sum b_n q^n$ . Assume that  $f$  and  $g$  satisfy  $a_1 = b_1 = 1$ , and  $a_n = \chi_p(n)b_n$  for all  $n$  prime to  $p$ . Then  $f$  and  $g$  are classical.*

*Remark 11.2.* As a consequence of the proof, we shall see that, in the situation above,  $a_p$  is determined by  $a_n$  for  $n$  prime to  $p$ .

*Proof.* Note that by Theorem 5.2 we know that  $f$  and  $g$  can be extended to sections of  $\omega^{\otimes k}$  on  $\mathcal{W}_1(p^m)$ .

Let us consider the cusp  $(\mathbf{G}_m/q^{\mathbf{Z}}, i, q^{1/p}\zeta_{p^m})$ , corresponding to a point  $c \in X_1(Np^m)^{\text{an}}$  defined over  $K$ , which we shall assume contains a fixed primitive  $p^m$ th root of unity,  $\zeta = \zeta_{p^m}$ . Then  $c$  is not in  $\mathcal{U}_0$  but is in  $\mathcal{U}_1$ , because  $H_{m-1}(\mathbf{G}_m/q^{\mathbf{Z}})$  is  $\langle \zeta^p \rangle$ . In particular, both  $f$  and  $g|w$  can be analytically continued to an admissible open subset of  $X_1(Np^m)^{\text{an}}$  containing  $c$ . Note that this in fact makes the argument rather easier than the tame cases. Let us compute the  $q$ -expansions of  $f$  and  $g|w$  at  $c$ .

**Sublemma.**  $f(c) = (\sum c_n q^{n/p})(dt/t)^k$  where

$$c_n = a_n / (pa_p) \sum_{j=0}^{p-1} \chi_p(1 - jp^{m-1}) \zeta^{p^{m-1}jn}.$$

*Proof of sublemma.* Write  $\mu = \zeta^{p^{m-1}}$ , a primitive  $p$ th root of unity. Then

$$\begin{aligned} pa_p f(c) &= tr(f)(c) = tr(f)(\mathbf{G}_m/q^{\mathbf{Z}}, i, q^{1/p}\zeta) \\ &= \sum_{C \neq \langle \mu \rangle} pr^* f(\mathbf{G}_m/q^{\mathbf{Z}}/C, \bar{i}, \overline{q^{1/p}\zeta}) \\ &= \sum_{j=0}^{p-1} pr^* f(\mathbf{G}_m/\langle \mu^j q^{1/p} \rangle, i, q^{1/p}\zeta) \\ &= \sum_{j=0}^{p-1} pr^* f(\mathbf{G}_m/\langle \mu^j q^{1/p} \rangle, i, \zeta \mu^{-j}) \\ &= \sum_{j=0}^{p-1} pr^* f(\mathbf{G}_m/\langle \mu^j q^{1/p} \rangle, i, \zeta^{1-jp^{m-1}}) \\ &= \sum_{j=0}^{p-1} \chi_p(1 - jp^{m-1}) pr^* f(\mathbf{G}_m/\langle \mu^j q^{1/p} \rangle, i, \zeta) \\ &= \sum_{j=0}^{p-1} \chi_p(1 - jp^{m-1}) \sum_{n \geq 0} a_n (\mu^j q^{1/p})^n (dt/t)^k \\ &= \sum_{n \geq 0} a_n q^{n/p} \sum_{j=0}^{p-1} \chi_p(1 - jp^{m-1}) \mu^{jn} (dt/t)^k \end{aligned}$$

as required. □

Before we go on, we make some remarks about this  $q$ -expansion. We know that  $\chi_p$  has conductor  $p^m$  and hence  $\chi_p(1 + p^{m-1})$  must be a primitive  $p$ th root of unity. Write  $\chi_p(1 + p^{m-1}) = \mu^r$  with  $0 < r < p$ . Now note that  $\chi_p(1 - jp^{m-1}) = \mu^{-rj}$  and hence

$$\sum_{j=0}^{p-1} \chi_p(1 - jp^{m-1}) \mu^{jn} = \sum_{j=0}^{p-1} \mu^{jn-jr}$$

which vanishes unless  $n \equiv r \pmod{p}$ , in which case it is  $p$ . Hence

**Corollary 11.3 (of sublemma).**

$$f(c) = (a_p)^{-1} \sum_{n \geq 1, n \equiv r \pmod{p}} a_n q^{n/p} (dt/t)^k.$$

□

This corollary implies that, in the situation that we are in,  $f$  is determined by  $a_n$  for  $n$  prime to  $p$ . This is because if  $f_1$  and  $f_2$  were as in the Theorem, with the same  $a_n$  ( $p \nmid n$ ) and distinct non-zero  $a_p$ , then a non-trivial linear combination of  $f_1$  and  $f_2$  would vanish in a neighbourhood of  $c$  and would hence be identically zero. This partially answers a question of Gouvêa.

We now calculate  $g|w(c)$ . Firstly, notice that by definition of  $w$ , we have  $(g|w)(E, i, P) = pr^*g(E/\langle P \rangle, \bar{i}, Q)$ , where  $pr : E \rightarrow E/\langle P \rangle$  is the projection and  $Q$  is chosen such that  $\langle P, Q \rangle = \zeta$ . Next note that raising to the power  $p^{m-1}$  on  $\mathbf{G}_m$  induces an isomorphism  $\mathbf{G}_m/\langle q, q^{1/p}\zeta \rangle \rightarrow \mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle$ , and that the dual of the map

$$pr : \mathbf{G}_m/q^{\mathbf{Z}} \xrightarrow{p^{m-1}} \mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle$$

is a map  $\mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle \rightarrow \mathbf{G}_m/q^{\mathbf{Z}}$  induced by the  $p$ th power map on  $\mathbf{G}_m$ . Moreover,  $q^{1/p}$  generates the kernel of this map. Hence

$$\begin{aligned} (g|w)(c) &= (g|w)(\mathbf{G}_m/q^{\mathbf{Z}}, i, q^{1/p}\zeta) \\ &= pr^*(g(\mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle, p^{m-1}i, q^{1/p})). \end{aligned}$$

Note that the cusp  $(\mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle, p^{m-1}i, q^{1/p})$  is in  $\mathcal{U}_{m-1}$  and so we will be able to evaluate  $g$  on it by iterating the  $U_p$  operator  $m-1$  times. More precisely, we know that

$$(pa_p)^{m-1}g(E, i, P) = (pU_p)^{m-1}g(E, i, P) = \sum_C pr^*g(E/C, \bar{i}, \bar{P})$$

where the sum is over the cyclic subgroups of  $E$  of order  $p^{m-1}$  which have trivial intersection with  $\langle P \rangle$ .

If  $E$  is the curve  $\mathbf{G}_m/\langle q^{p^{m-2}}\mu \rangle$  and  $P = q^{1/p}$  then  $\langle p^{m-1}P \rangle = \langle q^{p^{m-2}} \rangle = \langle \mu^{-1} \rangle = \mu_p$  and we deduce that the  $p^{m-1}$  cyclic subgroups  $C$  over which the sum is taken are the groups generated by  $q^{1/p}\zeta^{1+sp}$  for  $0 \leq s \leq p^{m-1} - 1$ . Hence

$$(g|w)(c) = (pa_p)^{1-m} \sum_{s=0}^{p^{m-1}-1} pr^*g(\mathbf{G}_m/\langle q^{1/p}\zeta^{1+sp} \rangle, p^{m-1}i, q^{1/p})$$

where  $pr : \mathbf{G}_m/q^{\mathbf{Z}} \rightarrow \mathbf{G}_m/\langle q^{1/p}\zeta^{1+sp} \rangle$  is induced by raising to the power  $p^{m-1}$  on  $\mathbf{G}_m$ . In particular, there are non-zero constants  $\kappa_1$  and  $\kappa_2$  (which are easy

to compute, but which are suppressed for simplicity) such that

$$\begin{aligned}
(g|w)(c) &= \kappa_1 \sum_{s=0}^{p^{m-1}-1} pr^* g(\mathbf{G}_m / \langle q^{1/p} \zeta^{1+sp} \rangle, i, q^{-1/p}) \\
&= \kappa_1 \sum_{s=0}^{p^{m-1}-1} pr^* g(\mathbf{G}_m / \langle q^{1/p} \zeta^{1+sp} \rangle, i, \zeta^{1+sp}) \\
&= \kappa_1 \sum_{s=0}^{p^{m-1}-1} \chi_p(1+sp)^{-1} pr^* g(\mathbf{G}_m / \langle q^{1/p} \zeta^{1+sp} \rangle, i, \zeta) \\
&= \kappa_2 \sum_{s=0}^{p^{m-1}-1} \chi_p(1+sp)^{-1} \sum_{n \geq 0} b_n q^{n/p} \zeta^{n+ns} (dt/t)^k \\
&= \kappa_2 \sum_{n \geq 0} b_n q^{n/p} \sum_{s=0}^{p^{m-1}-1} \chi_p(1+sp)^{-1} \zeta^{n+ns} (dt/t)^k.
\end{aligned}$$

Now define  $\gamma_n := \sum_{s=0}^{p^{m-1}-1} \chi_p(1+sp)^{-1} \zeta^{n+ns}$ . Replacing the sum over  $s$  by a sum over  $s' = s - p^{m-2}$ , it is easy to deduce that  $\gamma_n = \mu^{n-r} \gamma_n$ , and hence that  $\gamma_n$  vanishes if  $n$  is not congruent to  $r \pmod p$ . Moreover, a similar easy exercise shows that for  $n \equiv r \pmod p$  and  $j \equiv 1 \pmod p$ , we have  $\gamma_{nj} = \chi_p(j) \gamma_n$ , and hence that  $\gamma_n / \chi_p(n)$  is independent of  $n$  if  $n \equiv r \pmod p$ . So for some non-zero constant  $\kappa_3$  we have

$$\begin{aligned}
(g|w)(c) &= \kappa_3 \sum_{n \geq 1, n \equiv r \pmod p} \chi_p(n) b_n q^{n/p} (dt/t)^k \\
&= \kappa_3 \sum_{n \geq 1, n \equiv r \pmod p} a_n q^{n/p} (dt/t)^k.
\end{aligned}$$

From this and the earlier corollary, we deduce that for some non-zero constant  $\kappa_4$  we have that  $f$  and  $\kappa_4 g|w$  agree on an open disc around the cusp  $c$  and hence that  $f$  and  $g|w$  can be glued together to form a rigid (and hence algebraic) section of  $\omega^{\otimes k}$  on the whole of  $X_1(Np^m)$ , which was what we wanted.  $\square$

We remark that the constants that were involved in the computation above can easily be made explicit (they involve Gauss sums, for example) but some are rather nasty to write down, and all are irrelevant, so we have suppressed them.

## References

- [Be] P. Berthelot, *Cohomologie rigide et cohomologie rigide à supports propres*, preprint available at <http://www.maths.univ-rennes1.fr/~berthelo/>.

- [Bu] J. P. Buhler, *Icosahedral Galois representations*, Springer-Verlag, Berlin, 1978, Lecture Notes in Mathematics, Vol. 654.
- [BGR] S. Bosch, U. Güntzer, R. Remmert, *Non-Archimedean Analysis*, Grundlehren der mathematischen Wissenschaften 261, Springer, 1984.
- [BDST] K. Buzzard, M. Dickinson, N. Shepherd-Barron, R. Taylor, *On Icosahedral Galois Representations*, accepted for publication in Duke Math. Journal.
- [BT] K. Buzzard and R. Taylor, *Companion forms and weight 1 forms*, Annals of Mathematics 149 (1999), 905–919.
- [Col1] R. Coleman, *Reciprocity laws on curves*, Compos. Math. 72 (1989), 205–235.
- [Col2] R. Coleman, *The Monodromy Pairing*, pre-print.
- [Con] B. Conrad, *Modular curves and rigid analytic spaces*, pre-print.
- [D] F. Diamond, *On deformation rings and Hecke rings*, Ann. Math. 144 (1996), 137–166.
- [DR] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, LNM 349, Springer-Verlag (1971), 123–165.
- [E] S. Edixhoven, *Minimal resolution and stable reduction of  $X_0(N)$* , Ann. Inst. Fourier 40(1) (1990), 31–67.
- [G] F. Gouvêa, *Arithmetic of  $p$ -adic modular forms*, Springer LNM 1304 (1988)
- [K] N. Katz,  *$p$ -adic properties of modular schemes and modular forms*, LNM 350 (1973), 69–170.
- [Kö] U. Köpf, *Über eigentliche Familien algebraischer Varietäten über affinoiden Räumen*, /schriftenreihe des Mathematischen Instituts der Universität Münster, 2 Serie, Heft 7 (1974).
- [KM] N. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Math. Stud. 108, Princeton University (1985).