COMPOSITIO MATHEMATICA

Wiles defect for Hecke algebras that are not complete intersections

Gebhard Böckle, Chandrashekhar B. Khare and Jeffrey Manning

Compositio Math. **157** (2021), 2046–2088.

doi:10.1112/S0010437X21007454

FOUNDATION
COMPOSITIO
MATHEMATICA



LONDON
MATHEMATICAL
SOCIETY
EST. 1865

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Compositio Math. 157 (2021) 2046–2088
doi:10.1112/S0010437X21007454

Wiles defect for Hecke algebras that are not complete intersections

Gebhard Böckle, Chandrashekhar B. Khare and Jeffrey Manning

Abstract

In his work on modularity theorems, Wiles proved a numerical criterion for a map of rings $R \to T$ to be an isomorphism of complete intersections. He used this to show that certain deformation rings and Hecke algebras associated to a mod p Galois representation at non-minimal level are isomorphic and complete intersections, provided the same is true at minimal level. In this paper we study Hecke algebras acting on cohomology of Shimura curves arising from maximal orders in indefinite quaternion algebras over the rationals localized at a semistable irreducible mod p Galois representation $\bar{\rho}$. If $\bar{\rho}$ is scalar at some primes dividing the discriminant of the quaternion algebra, then the Hecke algebra is still isomorphic to the deformation ring, but is not a complete intersection, or even Gorenstein, so the Wiles numerical criterion cannot apply. We consider a weight-2 newform f which contributes to the cohomology of the Shimura curve and gives rise to an augmentation λ_f of the Hecke algebra. We quantify the failure of the Wiles numerical criterion at λ_f by computing the associated *Wiles defect* purely in terms of the local behavior at primes dividing the discriminant of the global Galois representation ρ_f which f gives rise to by the Eichler–Shimura construction. One of the main tools used in the proof is Taylor–Wiles–Kisin patching.

1. Introduction

Let $p > 2$ be prime. Let E be a finite extension of $\mathbf{Q}_p$ with valuation ring $\mathcal{O}$, residue field k and a uniformizer ϖ. All our rings will be $\mathcal{O}$-algebras. We say that an $\mathcal{O}$-algebra R that is a complete Noetherian local ring, and is finite flat as an $\mathcal{O}$-module, with residue field k, is a complete intersection if $R \simeq \mathcal{O}[[T_1, \dots, T_n]]/(f_1, \dots, f_n)$ for some n.

Wiles's numerical isomorphism criterion played an important role in his work [Wil95] on modularity of elliptic curves. Wiles characterized rings that were finite flat over $\mathcal{O}$ which were complete intersections and used the criterion to deduce modularity lifting theorems in non-minimal level from those in minimal level.

Wiles and Lenstra proved the following result [Len95, Wil95].

Theorem 1.1. *Let R and T be complete Noetherian local $\mathcal{O}$-algebras with residue field k, T finite flat as an $\mathcal{O}$-module and $\phi : R \to T$ a surjective map of local $\mathcal{O}$-algebras. Let $\lambda : T \to \mathcal{O}$ be a homomorphism of local $\mathcal{O}$-algebras, and set $\Phi_R = \ker(\lambda\phi)/\ker(\lambda\phi)^2$ and*

Received 14 November 2019, accepted in final form 30 March 2021, published online 16 August 2021.
2020 Mathematics Subject Classification 11F80 (primary), 13C99, 13F99 (secondary).
Keywords: wiles defect, modularity theorem, Cohen–Macaulay.
© 2021 The Author(s). The publishing rights in this article are licensed to Foundation Compositio Mathematica under an exclusive licence.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

$\eta_T = \lambda(\mathrm{Ann}_T(\ker(\lambda)))$. *Then* $|\mathcal{O}/\eta_T| \leq |\Phi_R|$, *where in the case* $|\mathcal{O}/\eta_T|$ *is infinite this is interpreted to mean that* Φ_R *is infinite too. Assume that* η_T *is not zero. Then the following are equivalent.*

- *The equality* $|\Phi_R| = |\mathcal{O}/\eta_T|$ *is satisfied.*
- *The rings* R *and* T *are complete intersections, and* ϕ *is an isomorphism.*

This paper arose out of trying to answer the following question, which arises in a situation studied in [Rib90b]. Consider the elliptic curve $X_1(11)$ defined by $y^2z + yz^2 = x^3 - x^2z$, and the mod $p = 3$ representation $\bar{\rho}$ arising from it. Its Serre invariants are $k(\bar{\rho}) = 2, N(\bar{\rho}) = 11, \epsilon(\bar{\rho}) = 1$ and its image is $\mathrm{GL}_2(\mathbf{F}_3)$.

Let $q \neq 3, 11$ be a prime at which $\mathrm{tr}(\bar{\rho}(\mathrm{Frob}_q)) \equiv \pm(q+1) \pmod{p}$. By Ribet's level-raising results [Rib84], we know that there is a newform $f \in S_2(\Gamma_0(11q))$ which gives rise to $\bar{\rho}$. More precisely, we fix an embedding $\iota : K_f \hookrightarrow \bar{\mathbf{Q}}_p$, with $K_f = \mathbf{Q}(a_n(f))$ the number field generated by the Fourier coefficients $f = \Sigma_n a_n(f) q^n$ of f, and assume that $\mathcal{O}$ is the ring of integers of the completion of $\iota(K_f)$. Then, by f gives rise to $\bar{\rho}$, we mean that the Galois representation $\rho_{f,\iota} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O})$ arising from f, ι by the Eichler–Shimura construction is residually isomorphic to $\bar{\rho}$.

Consider the Shimura curve X over $\mathbf{Q}$ arising from the maximal order of the indefinite quaternion algebra that is ramified at 11 and q. Let $\mathbf{T}$ be the $\mathcal{O}$-algebra generated by Hecke operators T_r for r prime, acting on $H^1(X, \mathcal{O})$. Then f gives rise (via Jacquet–Langlands) to a homomorphism $\lambda_f : \mathbf{T} \to \mathcal{O}$ with kernel $\mathfrak{p}_f$ contained in a maximal ideal $\mathfrak{m}$. Write $Q = \{11, q\}$, let $\mathbf{T}^Q := \mathbf{T}_{\mathfrak{m}}$ and let R^Q be the appropriate Galois deformation ring of $\bar{\rho}$ parameterizing lifts of $\bar{\rho}$ of fixed determinant which are Steinberg at 11 and q (which we define explicitly in § 4.3 as R^{St}). Then there is a Galois representation $\rho_{\mathfrak{m}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}^Q)$ with residual representation $\bar{\rho}$ inducing a surjective morphism $\phi^Q : R^Q \twoheadrightarrow \mathbf{T}^Q$ of $\mathcal{O}$-algebras.

This is now the setting of Theorem 1.1, applied to these rings and the morphism $\lambda_f : \mathbf{T}^Q \to \mathcal{O}$. Thus, one can consider $\Phi_f^Q := \Phi_{R^Q}$ and $\eta_f^Q := \eta_{\mathbf{T}^Q}$, referred to as the *cotangent space* and *congruence ideal* of f, and Theorem 1.1 implies that $|\Phi_f^Q| \geq |\mathcal{O}/\eta_f^Q|$. We also refer to $\mathcal{O}/\eta_f^Q$ as the *congruence module* of f.

In the case when $\bar{\rho}(\mathrm{Frob}_q) \neq \pm\mathrm{Id}$, automorphy lifting techniques, in the form of the Taylor–Wiles–Kisin patching method, imply that R^Q and $\mathbf{T}^Q$ are complete intersections and ϕ^Q is an isomorphism. In this case, Theorem 1.1 gives the equality $|\Phi_f^Q| = |\mathcal{O}/\eta_f^Q|$.

On the other hand, we say that q is a *trivial prime* for $\bar{\rho}$ if $\bar{\rho}(\mathrm{Frob}_q) = \pm\mathrm{Id}$, which implies that $q \equiv 1 \pmod{p}$, as $\det \bar{\rho}$ is the cyclotomic character (as shown in [Rib90b], the first such prime for this specific $\bar{\rho}$ is $q = 193$, at which $\bar{\rho}(\mathrm{Frob}_{193}) = -\mathrm{Id}$). For such a q, patching techniques no longer imply that R^Q and $\mathbf{T}^Q$ are complete intersections. In fact, computations of Shotton in [Sho16], combined with patching, still imply that ϕ^Q is an isomorphism, but also show that $\mathbf{T}^Q$ is *not* a complete intersection (nor is it Gorenstein). Theorem 1.1 then implies that $|\Phi_f^Q| > |\mathcal{O}/\eta_f^Q|$.

Trivial primes have played an important role in automorphy lifting (for instance, Taylor's method of Ihara avoidance [Tay08]) and lifting Galois representations [HR08, FKP19]. Trivial primes are very useful, because of their versatility in allowing lifts of different types or 'killing dual Selmer classes' as in [HR08, FKP19] Trivial primes behave rather differently from the other primes, and our paper explores this difference in the context of the Wiles numerical criterion.

Thus, our primary goal in this paper is to determine the ratio $|\Phi_f^Q|/|\mathcal{O}/\eta_f^Q|$, or equivalently the $\mathcal{O}$-module $\eta_f^Q/\mathrm{Fitt}_{\mathcal{O}}(\Phi_f^Q)$, in this situation. This is called the *Wiles defect* of $\mathbf{T}^Q$ relative to λ_f as defined in [TU18].

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Specializing our main theorem (Theorem 10.1) to this particular case gives the following result.

Theorem 1.2. *Let n_q be the maximal integer such that $\rho_{f,\iota}$ locally at q is $\pm\mathrm{Id}$ mod ϖ^{n_q} (i.e. the maximal n_q such that $\rho_{f,\iota}|_{G_{\mathbf{Q}_q}}$ is congruent to $\pm\mathrm{Id}$ mod ϖ^{n_q}). Then $\eta_f^Q/\mathrm{Fitt}_{\mathcal{O}}(\Phi_f^Q)$ is $\mathcal{O}/\varpi^{2n_q}$.*

We use a different (logarithmic) normalization of the Wiles defect (see Definition 3.10 below) to make it behave well with change of coefficient ring $\mathcal{O}$. With our definition, the Wiles defect $\delta_{\lambda_f,\mathbf{T}^Q}(\mathbf{T}^Q)$ (of $\mathbf{T}^Q$ relative to λ_f) in the theorem above becomes $2n_q/e$, where e is the ramification index of $\mathcal{O}$.

In the full version of our results, see Theorem 10.1, we allow f to be an arbitrary newform of squarefree level and trivial nebentypus, for which the residual representation $\bar\rho$ satisfies the Taylor–Wiles conditions, as well as a condition on the ramification, and we allow the quaternion algebra D to be ramified at multiple trivial primes for $\bar\rho$. For ease of exposition, we will continue to focus on the case outlined above for the remainder of the introduction.

Remark 1.3. As a small point, we remark that while Theorem 1.2 is technically a complete answer to the question of determining the Wiles defect, the numbers n_q may be difficult to compute in the specific case considered here. The reason for this is that while the original representation $\bar\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_3)$ at level 11 is defined by a modular form with rational coefficients, the coefficient field K_f of the newform $f \in S_2(\Gamma_0(11q))$ may be arbitrarily large.

For example, in the case of the first trivial prime $q = 193$, one can show by process of elimination that the newform $f \in S_2(\Gamma_0(11 \cdot 193))$ giving rise to $\bar\rho$ is the form computed in [LMF19, Newform 2123.2.a.h], for which we have $[K_f : \mathbf{Q}] = 46$. This may make a precise determination of n_{193} computationally infeasible.

In § 11 we consider a different specialization of Theorem 10.1 to the case when the newform f has rational coefficients (but we do not assume that the residual representation $\bar\rho$ arises from a modular form with rational coefficients at minimal level). In this case, f corresponds to a semistable elliptic curve E_f, and we show that n_q may be computed in terms of the Tate uniformization of E at q. This allows us to give several numerical computations of Wiles defects via our main theorem.

In the study of congruences between modular forms (cf. [Hid81a]), and computing congruence ideals attached to a newform f, it is found that *cohomological* congruence ideals $\eta_f^{Q,\mathrm{coh}}$, defined similarly to η_f^Q but in terms of the $\mathbf{T}^Q$ module $H^Q := H^1(X,\mathcal{O})_{\mathfrak{m}}$ rather than $\mathbf{T}^Q$,[1] are easier to study (and are related directly to algebraic parts of L-values) than the congruence ideal η_f^Q of the relevant Hecke algebra $\mathbf{T}^Q$.

In [Dia97], Diamond gives a generalization of Theorem 1.1 which shows that $|\Phi_f^Q| = |\mathcal{O}/\eta_f^{Q,\mathrm{coh}}|$ if and only if $\phi^Q : R^Q \to \mathbf{T}^Q$ is an isomorphism of complete intersections and H^Q is free of rank 2 over $\mathbf{T}^Q$.

To relate the congruence ideals η_f^Q and $\eta_f^{Q,\mathrm{coh}}$, the method in most references (see e.g. [Hid81b]) is to prove by delicate arithmetic geometric means that H^Q is actually free of rank 2 over $\mathbf{T}^Q$, which trivially implies that $\eta_f^Q = \eta_f^{Q,\mathrm{coh}}$.

[1] We are suppressing some subtleties here to do with the fact that $H^1(X,\mathcal{O})_{\mathfrak{m}}$ has generic rank 2 as a $\mathbf{T}^Q$-module, rather than 1.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

In the case when q is not a trivial prime, the same patching arguments that imply that R^Q and $\mathbf{T}^Q$ are complete intersections were shown in [Dia97] and [Fuj06] to imply that H^Q is free over $\mathbf{T}^Q$ (a fact which was also shown more directly in [Rib90b]). On the other hand, when q is a trivial prime, the main result of [Rib90b] shows that H^Q is not free over $\mathbf{T}^Q$. Instead it is proved that $\dim_k H^Q/\mathfrak{m} = 4$, whereas $H^Q[1/p]$ is free of rank 2 over $\mathbf{T}^Q \otimes \mathbf{Q}_p$, which is referred to as a multiplicity-two phenomenon.

The work of the third author in [Man21] uses patching techniques to get another proof of Ribet's result in a more general situation. In addition to re-proving Ribet's multiplicity-two result, this approach also yields an additional statement about the endomorphism ring of H^Q as a $\mathbf{T}^Q$-module (see [Man21, Theorem 2]). In Theorem 3.12, we show that this endomorphism statement implies that $\eta_f^Q = \eta_f^{Q,\mathrm{coh}}$, even in the case when H^Q is not free over $\mathbf{T}^Q$.[2]

This shows that a weaker condition than freeness can suffice to show that the two congruence modules are the same, which holds in certain cases when freeness is actually false. Moreover, the techniques of [Man21] give a possible approach for proving this condition in more generality, by carefully analyzing the geometry of certain local Galois deformation rings. We hope that these ideas will be useful more generally to relate congruence ideals of Hecke algebras and congruence ideals of cohomology groups arising from arithmetic manifolds on which they naturally act.

1.1 Strategy of Proof of Theorem 10.1

Here is a sketch of the strategy of the proof of our main theorem, Theorem 10.1. The notation we use here is not the same as in the main text, and we simplify the arguments, skipping over some nuances, to indicate the broad strategy. We consider the full deformation ring $R(Q)$ which arises from considering semistable deformations of determinant the p-adic cyclotomic character ϵ, which are unramified outside $Q = \{3, 11, q\}$ and finite flat at 3. We consider the anemic Hecke algebra $\mathbf{T}(Q)$ (without operators U_q for $q \in Q$) that acts faithfully on the cohomology $H^1(X_0(11q^2), \mathcal{O})_{\mathfrak{m}_q}$ of the modular curve. (Here $\mathfrak{m}_Q$ is a maximal ideal of the full Hecke algebra, which is a pull back of the maximal ideal $\mathfrak{m}$ of $\mathbf{T}^Q$.) There is a surjective map of $\mathcal{O}$-algebras $R(Q) \twoheadrightarrow \mathbf{T}(Q)$ that is an isomorphism of complete intersections $R(Q) \xrightarrow{\sim} \mathbf{T}(Q)$ by [Wil95] and [Dia97].

The rings R^Q and $\mathbf{T}^Q$ are quotients of $R(Q)$ and $\mathbf{T}(Q)$, respectively. We can therefore pull our augmentation $\lambda_f : \mathbf{T}^Q \to \mathcal{O}$ back to $\lambda_f : \mathbf{T}(Q) \to \mathcal{O}$, and use that for this pull back, the Wiles defect is trivial by Theorems 1.1 and 5.2. For this one uses another augmentation $\lambda_g : \mathbf{T}(Q) \to \mathcal{O}$ arising from a newform $g \in S_2(\Gamma_0(N(\bar{\rho}))$ which exists because of the level-lowering results of [Rib90a]. Thus, to compute the Wiles defect of $\mathbf{T}^Q$, it suffices to compute the change in cotangent spaces (at λ_f) between $R(Q)$ and R^Q and the change in congruence modules between $\mathbf{T}(Q)$ and $\mathbf{T}^Q$.

The change in cotangent spaces is considered in § 7. The main idea is that the Taylor–Wiles–Kisin patching method (see Theorem 6.3 for details) implies that $R(Q)$ and R^Q are quotients of certain explicit rings R_∞ and R_∞^{St} by a regular sequence. Here R_∞ and R_∞^{St} are both power series rings over completed tensor products of various local Galois deformation rings. In all cases relevant to us, the local deformation rings were computed precisely in [Sho16], and thus we can determine R_∞ and R_∞^{St} completely explicitly.

Pulling back λ_f to augmentations on R_∞^{St} and R_∞, one can then consider the cotangent spaces $\Phi_{R_\infty^{\mathrm{St}}}$ and Φ_{R_∞}, which depend only on the rings R_∞^{St} and R_∞ and the map $\lambda : R_\infty^{\mathrm{St}} \to \mathcal{O}$, and not

[2] Note that this does not contradict Diamond's numerical criterion from [Dia97], since in the case when H^Q is not free $\mathbf{T}^Q$ will not be a complete intersection and so $|\Phi_f^Q| > |\mathcal{O}/\eta_f^Q| = |\mathcal{O}/\eta_f^{Q,\mathrm{coh}}|$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

on the structures of $R(Q)$ and R^Q. While it is not possible to directly determine Φ_{R^Q} and $\Phi_{R(Q)}$ from $\Phi_{R_\infty^{\mathrm{St}}}$ and Φ_{R_∞} (at least without knowing the regular sequence generating $\ker(R_\infty \twoheadrightarrow R(Q))$ and $\ker(R_\infty^{\mathrm{St}} \twoheadrightarrow R^Q)$), we can show via a commutative algebra argument (Theorem 7.14) that $\ker(\Phi_{R(Q)} \twoheadrightarrow \Phi_{R^Q})$ is isomorphic to $\ker(\Phi_{R_\infty} \twoheadrightarrow \Phi_{R_\infty^{\mathrm{St}}})$. The cardinality of the latter kernel can then be computed precisely by a computation with the local deformation rings (Proposition 7.9), which gives the desired comparison between $|\Phi_{R(Q)}|$ and $|\Phi_{R^Q}|$.

The change in congruence modules between $\mathcal{O}/\eta_{R(Q)}$ and $\mathcal{O}/\eta_{R^Q}$ is considered in §§ 8 and 9. We first equate these with the corresponding cohomological congruence modules, via Theorem 5.2 in the modular curve case, and the work of [Man21] combined with Theorem 3.12 (as described above) in the Shimura curve case (see Theorem 8.1 and Corollary 8.3). The change in cohomological congruence modules is then computed by an application of the work of [Rib90a] and [RT97] (see Proposition 9.1).

Relying on the results of [RT97] in this way requires us to assume that $\bar\rho$ is ramified at least at one prime in the discriminant of D. By using the Jacquet–Langlands isomorphism to add or remove certain primes to or from the discriminant of D, we may reduce this condition to simply requiring that $\bar\rho$ is ramified and Steinberg at enough primes, regardless of whether they are in the discriminant of D (see Corollary 9.3).

For this paper we have, for simplicity, restricted to the case when $\bar\rho$ is a semistable representation of $G_{\mathbf{Q}}$, so this is only a fairly minor restriction. However, this would be a much more restrictive condition if we were to give the natural generalizations of our results to possibly non-semistable representations of $\bar\rho : G_F \to \mathrm{GL}_2(k)$, for F a totally real number field. It would thus be worthwhile to find an alternative approach to computing the change in congruence modules, or to directly computing the Wiles defect, which does not depend on the results of [RT97].

1.2 Future work

Theorem 10.1 shows that the Wiles defect depends solely on the local properties of the representation $\rho_{f,\iota}$ at the primes in the discriminant of D, and thus only on the ring R_∞^{St} and the induced augmentation $\lambda : R_\infty^{\mathrm{St}} \to \mathcal{O}$ from § 7. Based on this, one might hope that the Wiles defect could be computed directly from the local deformation rings, similarly to how the change in cotangent spaces was computed in § 7. Such a computation would remove the reliance on the results [RT97] from our approach, and would thus allow us to significantly weaken the hypotheses of our main theorem. It would also potentially allow us to compute the Wiles defect in contexts where there is no natural generalization of [RT97]. We intend to pursue these ideas in a follow-up paper.

The work of Hida [Hid81a] proves a formula of the type

$$\frac{L(1, \mathrm{ad}(f))}{i\pi\Omega} = |\mathcal{O}/\eta_f(\mathbf{T}(Q))|$$

(up to a unit in $\mathcal{O}$) relating L-values and congruence modules (we blur here the distinction between the congruence module for the Hecke algebra $\mathbf{T}(Q)$ and the congruence module for its module $H^1(X_0(NQ), \mathcal{O})_{\mathfrak{m}}$). Here $L(s, \mathrm{ad}(f))$ is the L-function and Ω a suitable period of the adjoint motive associated to f. The work of Wiles [Wil95], when combined with [Hid81a], gives a proof of the Bloch–Kato conjecture (in many cases) for the corresponding L-value by showing the equality of the orders of the congruence module $\mathcal{O}/\eta_f(\mathbf{T}(Q))$ and the cotangent space $\Phi_{R(Q)}$ associated to f (the latter is the Bloch–Kato Selmer group associated to the adjoint motive of f). Our work does not directly involve L-functions, but one could ask if the Wiles defect we study has a relation to L-values (perhaps related to the quotient of 'cohomological' and 'motivic' periods associated to the form f or its Jacquet–Langlands correspondent on the

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

quaternion algebra D_Q). The observations in § 11 that relate the Wiles defect, in the case of augmentations of Hecke algebras arising from semistable elliptic curves, via our main theorem, to tame regulators of local periods of the Tate elliptic curves at the primes in Q (cf. [MT87]), give a geometric description of the Wiles defect in the case we study and are suggestive of such a relationship.

We end the introduction by making a few remarks contrasting the contexts of [TU18] and our paper, which both study the Wiles defect. The paper [TU18] considers Hecke algebras which act on the Betti cohomology of hyperbolic 3-manifolds. Thus, the reason for the failure of complete intersection of the Hecke algebras studied in [TU18] is global, namely that the relevant Betti cohomology (the relevant degrees being 1 and 2) is not concentrated in a single degree (and thus the defect $\ell_0 = 1$). In § 5 of [TU18], the Wiles defect of the Hecke algebras studied there is conjecturally related to (integral) regulators of motivic elements, and to the ratio of cohomological and motivic periods.

In the classical situation studied in this paper, of Hecke algebras acting on the cohomology $H^1(X_Q, \mathcal{O})$ of Shimura curves $X_{Q/\mathbf{Q}}$ arising from an indefinite quaternion algebra D_Q over $\mathbf{Q}$, the reason for the failure of the Hecke algebra to be a complete intersection is local, namely that the *Steinberg* deformation rings at *trivial* primes dividing the discriminant of D_Q are not complete intersections. We compute in many cases the Wiles defect (cf. Theorem 10.1) and give an explicit formula for the Wiles defect (which is a global quantity) as a sum of local terms at such primes. The local terms at a prime q measure the depth of the congruence of $\rho_f|_{G_q}$ to the trivial representation up to sign, where G_q is a decomposition group at q.

1.3 Structure of this paper

We describe the contents of this paper. The key sections of this paper are as follows. In § 3 we develop the commutative algebra we need in this paper related to congruence modules (cf. Theorem 3.12) and define the Wiles defect (cf. Definition 3.10). In §§ 4 and 5 we introduce the deformation rings and Hecke algebras which are the main objects studied in this paper. In §§ 6 and 7 we use patching to show that certain relative local cotangent spaces are isomorphic to related relative global cotangent spaces (cf. Theorems 6.3 and 7.14 and Corollary 7.15). In § 7.2 we give the computation of the cardinality of relative local cotangent spaces at a trivial prime q (cf. Proposition 7.9). In § 8 we show that to prove our main theorem about the Wiles defect $\delta_\lambda(\mathbf{T}^Q)$ we study, Theorem 10.1, it is enough to compute a related cohomological Wiles defect $\delta_{\lambda,T^Q}(H^1(X_Q, \mathcal{O})_\mathfrak{m})$ related to the $\mathbf{T}^Q$ module $H^1(X_Q, \mathcal{O})_\mathfrak{m}$ (cf. Corollary 8.3). In § 9 we use [RT97] to compute the change of the cohomological η-invariant when we pull back our augmentation $\mathbf{T}^Q$ to $\mathbf{T}(Q)$ (cf. Proposition 9.1). In § 10 we pull all our earlier work together to prove Theorem 10.1. We end in § 11 by computing the Wiles defect for semistable elliptic curves in terms of tame regulators of local Tate periods (cf. Proposition 11.1 and Corollary 11.2) and giving some numerical examples.

Notation

We will fix a prime $p > 2$ throughout this paper and fix an embedding $\iota : \bar{\mathbf{Q}} \hookrightarrow \bar{\mathbf{Q}}_p$.

For any field F, we will let G_F denote its absolute Galois group. For any prime q, we will write $G_q := G_{\mathbf{Q}_q}$, and let $I_q \leq G_q$ be the inertia subgroup. For the entirety of this paper, we will fix a choice of embedding $G_q \hookrightarrow G_{\mathbf{Q}}$ for each prime q. All representations of G_F will be assumed to be continuous.

We denote by ϵ the p-adic cyclotomic character $\epsilon : G_{\mathbf{Q}} \to \mathbf{Z}_p^\times$ and also any of the characters to $R^\times$ we get for any $\mathbf{Z}_p$-algebra R by composing it with the map $\mathbf{Z}_p \to R$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

For a newform f, we denote by $\rho_{f,\iota}$ (or simply ρ_f as we have fixed the embedding ι) the Galois representation associated by Eichler–Shimura to f, ι.

We will denote by $\mathcal{O}$ the valuation ring of a finite extension E of $\mathbf{Q}_p$ and by k the residue field of $\mathcal{O}$, and we let $\mathrm{CNL}_{\mathcal{O}}$ be the category of complete, Noetherian local $\mathcal{O}$-algebras with residue field k.

For an $\mathcal{O}$-algebra R, an augmentation λ of R will always mean a surjective $\mathcal{O}$-algebra homomorphism $\lambda : R \to \mathcal{O}'$, where $\mathcal{O}'$ is the ring of integers in a finite extension of E (we will almost always take $\mathcal{O} = \mathcal{O}'$). For an $\mathcal{O}$-module M that is a finite abelian group, we denote by $\ell_{\mathcal{O}}(M)$ the length of M as an $\mathcal{O}$-module. For $\alpha \in \mathcal{O}$, we denote by $\mathrm{ord}_{\mathcal{O}}(\alpha) = \ell_{\mathcal{O}}(\mathcal{O}/(\alpha))$.

For a Galois representation $\bar{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\bar{\mathbf{F}}_p)$ which is finite flat at p, we will let $N(\bar{\rho})$ represent its Artin conductor.

2. Trivial primes for $\bar{\rho}$

Let $\bar{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ be a continuous and irreducible mod p representation.

Definition 2.1. We say that a prime q is *trivial* for $\bar{\rho}$ if q is 1 mod p, $\bar{\rho}$ is unramified at q and $\bar{\rho}(\mathrm{Frob}_q) = \pm\mathrm{Id}$. We say that a prime q is *Steinberg* for $\bar{\rho}$ if $\bar{\rho}|_{G_q}$ up to an unramified twist is of the form

$$\begin{pmatrix} \epsilon & * \\ 0 & 1 \end{pmatrix}$$

with ϵ the mod p cyclotomic character and G_q a decomposition group at q.

We impose that $\bar{\rho}$ satisfies the following conditions.

- $\det(\bar{\rho}) = \epsilon$.
- Each prime $\ell(\neq p)$ at which $\bar{\rho}$ is ramified is Steinberg for $\bar{\rho}$.
- $\bar{\rho}|_{G_p}$ is finite flat.
- $\bar{\rho}$ is modular, i.e. arises from a newform in $S_2(\Gamma_0(N(\bar{\rho})))$ with respect to an embedding $\iota_p : \bar{\mathbf{Q}} \to \bar{\mathbf{Q}}_p$ that we fix. This is implied by the previous hypotheses by Serre's modularity conjecture, which is now a theorem [KW09].

Note that $N(\bar{\rho}) > 1$ and is square-free, and our conditions imply that $\mathrm{Ad}^0(\bar{\rho})$ is absolutely irreducible and the order of $\mathrm{im}(\bar{\rho})$ is divisible by p (otherwise $\bar{\rho}$ has Serre weight 2 and Artin conductor 1 by assumptions on determinant and ramification above). This implies that $\bar{\rho}|_{\mathbf{Q}(\zeta_p)}$ is irreducible and thus satisfies the 'Taylor–Wiles hypothesis' needed for the arguments below.

In our work below we will fix positive square-free and coprime integers N and Q with the property that:

- N divides $N(\bar{\rho})$;
- $N(\bar{\rho})$ divides NQ;
- all primes dividing Q are Steinberg for $\bar{\rho}$.

We also fix a prime $r \gg 0$ such that no lift of $\bar{\rho}|_{G_r}$ is ramified at r. This exists by our hypothesis on $\bar{\rho}$ and [DT94b, Lemma 3]. This allows us to impose $\Gamma_1(r^2)$ level structure whenever needed, and so we may ignore below any issues arising from isotropy.

3. Cotangent spaces, congruence modules and the Wiles defect

Definition 3.1. Given any $R \in \mathrm{CNL}_{\mathcal{O}}$ and any continuous homomorphism $\lambda : R \to \mathcal{O}$, we define the *cotangent space of* R *with respect to* λ to be $\Phi_{\lambda,R} := (\ker \lambda)/(\ker \lambda)^2$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Note that [Sta19, Lemma 00RP, Lemma 02HP] implies that $\Phi_{\lambda,R} = \Omega_{R/\mathcal{O}} \otimes_\lambda \mathcal{O}$, where $\Omega_{R/\mathcal{O}}$ is the R-module of Kähler differentials.

First note the following result.

LEMMA 3.2. *If R is a finite, local, reduced $\mathcal{O}$-algebra, then $\Phi_{\lambda,R}$ is a finite group.*

Proof. As R is reduced, $R[1/\varpi] = R \otimes_{\mathcal{O}} E$ is as well. It follows that $R[1/\varpi]$ is a product of finite extensions of E and so $\Omega_{R/\mathcal{O}} \otimes_{\mathcal{O}} E = \Omega_{R[1/\varpi]/E} = 0$. Thus,

$$\Phi_{\lambda,R} \otimes_{\mathcal{O}} E = (\Omega_{R/\mathcal{O}} \otimes_\lambda \mathcal{O}) \otimes_{\mathcal{O}} E = (\Omega_{R/\mathcal{O}} \otimes_{\mathcal{O}} E) \otimes_\lambda E = 0.$$

But now as R is a finite $\mathcal{O}$-algebra, $\Omega_{R/\mathcal{O}}$ is a finitely generated R-module and hence $\Phi_{\lambda,R}$ is a finitely generated $\mathcal{O}$-module. It follows that $\Phi_{\lambda,R}$ is indeed finite. □

For the remainder of this section, let R denote a finite, local $\mathcal{O}$-algebra. Assume that R is ϖ-torsion free and reduced. This implies that the map $R \to R[1/\varpi] = R \otimes_{\mathcal{O}} E$ is injective, and $R[1/\varpi]$ is a product of finite extensions of E. Also fix an augmentation $\lambda : R \to \mathcal{O}$.

For any finitely generated R-module M, let $M^* = \mathrm{Hom}_{\mathcal{O}}(M, \mathcal{O})$ with its natural R-module structure. We say that M is *reflexive* if the natural map $M \to M^{**}$ is an isomorphism (note that this is equivalent to simply saying that $M \cong M^{**}$, since M^*, and hence M^{**}, is reflexive for any M). Note that if M is finitely generated over R, then it is also finitely generated over $\mathcal{O}$, so it follows that M is a reflexive R-module if and only if it is finitely generated over $\mathcal{O}$ and ϖ-torsion free.

For any reflexive R-module M, note that we have natural isomorphisms

$$(M[\ker\lambda]) \otimes_{\mathcal{O}} E \cong (M \otimes_{\mathcal{O}} E)/\ker\lambda \cong (M/\ker\lambda) \otimes_{\mathcal{O}} E.$$

We will define the λ-rank of M to be $\mathrm{rank}_\lambda M = \dim_E(M[\ker\lambda]) \otimes_{\mathcal{O}} E$, so that if $\mathrm{rank}_\lambda M = d$, then $M[\ker\lambda] \cong \mathcal{O}^d$ as an $\mathcal{O}$-module. It is not hard to see that $\mathrm{rank}_\lambda M = \mathrm{rank}_\lambda M^*$.

Now let $\omega = R^* = \mathrm{Hom}_{\mathcal{O}}(R, \mathcal{O})$ be the dualizing module of R. Then ω is a reflexive R-module and we have $\omega^* = R$.

For any reflexive R-module M, we have a natural, R-equivariant perfect pairing $\langle\,,\rangle_M : M^* \times M \to \mathcal{O}$ (given by $\langle f, x\rangle_M = f(x)$). Note that this induces an isomorphism $M[\ker\lambda] \cong \mathrm{Hom}_{\mathcal{O}}(M^*/\ker\lambda, \mathcal{O})$. Also, identifying M with M^{**}, we have $\langle x, y\rangle_{M^*} = \langle y, x\rangle_M$.

We now make the following definition.

DEFINITION 3.3. For any reflexive R-module M, the *congruence module of M with respect to* λ, denoted $\Psi_\lambda(M)$, is the cokernel of the composition

$$\iota_{M,\lambda} : M[\ker\lambda] \to M \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(M^*, \mathcal{O}) \to \mathrm{Hom}_{\mathcal{O}}(M^*[\ker\lambda], \mathcal{O}).$$

The map $\iota_{M,\lambda}$ is an isomorphism after inverting ϖ, so it is always injective, and $\Psi_\lambda(M)$ is always finite.

LEMMA 3.4. *If M is reflexive and $x_1, \dots, x_d$ (respectively $f_1, \dots, f_d$) is an $\mathcal{O}$-basis for $M[\ker\lambda]$ (respectively $M^*[\ker\lambda]$), then $\Psi_\lambda(M)$ is isomorphic to the cokernel of the map $\mathcal{O}^d \to \mathcal{O}^d$ given by the matrix $(\langle f_i, x_j\rangle_M)_{ij}$. In particular, $|\Psi_\lambda(M)| = |\mathcal{O}/\big(\det(\langle f_i, x_j\rangle_M)_{ij}\big)|$.*

Proof. If we identify $M[\ker\lambda]$ and $M^*[\ker\lambda]$ with $\mathcal{O}^d$ using the bases, $\iota_{M,\lambda}$ is identified with the map $\mathcal{O}^d \to \mathcal{O}^d$ given by the matrix $(\langle f_i, x_j\rangle_M)_{ij}$. The claim follows. □

In light of this lemma, in the case where $\mathrm{rank}_{\mathcal{O}} M[\ker\lambda] = \mathrm{rank}_{\mathcal{O}} M^*[\ker\lambda] = 1$, we define the ideal $\eta_\lambda(M) = (\langle f_1, x_1\rangle_M) \subseteq \mathcal{O}$, where $M[\ker\lambda] = \mathcal{O}x_1$ and $M^*[\ker\lambda] = \mathcal{O}f_1$. Then we have $\Psi_\lambda(M) \cong \mathcal{O}/\eta_\lambda(M)$, and $\eta_\lambda(M)$ is independent of the choice of x_1 and f_1.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Lemma 3.5. *If M is a reflexive R-module with an R-equivariant perfect pairing $\langle\,,\rangle : M \times M \to \mathcal{O}$ (i.e. M is* self-dual*) and $x_1, \dots, x_d$ is a basis for $M[\ker\lambda]$, then $\Psi_\lambda(M)$ is isomorphic to the cokernel of the map $\mathcal{O}^d \to \mathcal{O}^d$ given by the matrix $(\det(\langle x_i, x_j\rangle))_{ij}$. In particular, $|\Psi_\lambda(M)| = |\mathcal{O}/\big(\det(\langle x_i, x_j\rangle)_{ij}\big)|$. (Note that $\Psi_\lambda(M)$ is independent of this choice of pairing $\langle\,,\rangle$, by definition.)*

Proof. The pairing $\langle\,,\rangle$ induces an isomorphism $\varphi : M \xrightarrow{\sim} M^*$ of R-modules for which $\langle x, y\rangle = \langle\varphi(x), y\rangle_M$. Letting $f_i = \varphi(x_i)$, the claim now follows from the previous lemma. □

Lemma 3.6. *If M_0 and M_1 are reflexive R-modules, then $\Psi_\lambda(M_0 \oplus M_1) \cong \Psi_\lambda(M_0) \oplus \Psi_\lambda(M_1)$.*

Proof. Let $M = M_0 \oplus M_1$ and note that we have natural isomorphisms $M[\ker\lambda] = (M_0 \oplus M_1)[\ker\lambda] = M_0[\ker\lambda] \oplus M_1[\ker\lambda]$, $M^* = \mathrm{Hom}_{\mathcal{O}}(M_0 \oplus M_1, \mathcal{O}) \cong M_0^* \oplus M_1^*$, $M^*[\ker\lambda] = M_0^*[\ker\lambda] \oplus M_1^*[\ker\lambda]$ and $\mathrm{Hom}_{\mathcal{O}}(M^*[\ker\lambda], \mathcal{O}) \cong \mathrm{Hom}_{\mathcal{O}}(M_0^*[\ker\lambda], \mathcal{O}) \oplus \mathrm{Hom}_{\mathcal{O}}(M_1^*[\ker\lambda], \mathcal{O})$.

Under these identifications, we have

$$\iota_{M,\lambda} = \iota_{M_0,\lambda} \oplus \iota_{M_1,\lambda} : M_0[\ker\lambda] \oplus M_1[\ker\lambda] \to \mathrm{Hom}_{\mathcal{O}}(M_0^*[\ker\lambda], \mathcal{O}) \oplus \mathrm{Hom}_{\mathcal{O}}(M_1^*[\ker\lambda], \mathcal{O}),$$

from which the claim follows. □

Lemma 3.7. *For any reflexive module M, $\Psi_\lambda(M) \cong \Psi_\lambda(M^*)$.*

Proof. This follows from Lemma 3.4 and the fact that $\langle f_i, x_j\rangle_M = \langle x_j, f_i\rangle_{M^*}$. □

Lemma 3.8. *$\eta_\lambda(R) = \lambda(R[\ker\lambda])$.*

Proof. Let $R[\ker\lambda] = \mathcal{O}\alpha$ and note that $R^*[\ker\lambda] = \omega[\ker\lambda] = \mathrm{Hom}_{\mathcal{O}}(R/\ker\lambda, \mathcal{O}) = \mathcal{O}\lambda \subseteq R^*$. Then, by Lemma 3.4, $\eta_\lambda(R) = (\langle\lambda, \alpha\rangle_R) = (\lambda(\alpha)) = \lambda(R[\ker\lambda])$. □

Lemma 3.8 shows that our definition agrees with the definition used in [Wil95]. We also note the following, which implies that our definition of $\Psi_\lambda(M)$ agrees with the module Ω considered in [Dia97, Theorem 2.4].

Lemma 3.9. *Let $I_\lambda = R[\ker\lambda] \subseteq R$. For any reflexive R-module M, we have*

$$\Psi_\lambda(M) \cong \frac{M}{M[\ker\lambda] + M[I_\lambda]}.$$

Proof. Write $\mathfrak{p}_\lambda = \ker\lambda$. We first claim that the natural map $M \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(M^*, \mathcal{O}) \to \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$ induces an isomorphism $M/M[I_\lambda] \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$.

The map $M \to \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$ clearly factors through $M/\mathfrak{p}_\lambda M$. Also, as $M^*[\mathfrak{p}_\lambda]$ is saturated in M^*, this map is surjective, so we have a surjection $M/\mathfrak{p}_\lambda M \twoheadrightarrow \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$.

It is easy to see that this map is an isomorphism after inverting ϖ (this follows from the fact that $M[1/\varpi] = \mathfrak{p}_\lambda M[1/\varpi] \oplus I_\lambda M[1/\varpi]$ and $M^*[1/\varpi] = \mathfrak{p}_\lambda M^*[1/\varpi] \oplus I_\lambda M^*[1/\varpi]$) and so the map identifies $\mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$ with the maximal ϖ-torsion free quotient of $M/\mathfrak{p}_\lambda M$. But now as $M[1/\varpi]/\mathfrak{p}_\lambda M[1/\varpi] = (M[1/\varpi])[I_\lambda]$, $M[I_\lambda]$ is the saturation of $\mathfrak{p}_\lambda M$ in M as an $\mathcal{O}$-module and so $M/M[I_\lambda]$ is the maximal ϖ-torsion free quotient of $M/\mathfrak{p}_\lambda M$. Thus, we indeed have the desired natural isomorphism $M/M[I_\lambda] \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$.

But now identifying $M/M[I_\lambda] = \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$ under this isomorphism, the map $\iota_{M,\lambda} : M[\mathfrak{p}_\lambda] \to \mathrm{Hom}_{\mathcal{O}}(M^*[\mathfrak{p}_\lambda], \mathcal{O})$ from Definition 3.3 is identified with the map $M[\mathfrak{p}_\lambda] \hookrightarrow M \twoheadrightarrow M/M[I_\lambda]$ and so

$$\Psi_\lambda(M) = \mathrm{coker}(\iota_{M,\lambda}) = \frac{M/M[I_\lambda]}{(M[\mathfrak{p}_\lambda] + M[I_\lambda])/M[I_\lambda]} \cong \frac{M}{M[\mathfrak{p}_\lambda] + M[I_\lambda]}.$$ □

Our main goal for this paper is to compare the orders of $\Phi_{\lambda,R}$, $\Psi_\lambda(R)$ and $\Psi_\lambda(M)$ (for a specific choice of M). We thus make the following definition.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

DEFINITION 3.10. Let R denote a finite, local $\mathcal{O}$-algebra that is ϖ-torsion free and reduced. Let M be a reflexive R-module with $\operatorname{rank}_\lambda M = d > 0$.

The *Wiles defect* of M is the quantity

$$\delta_{\lambda,R}(M) = \frac{d\log|\Phi_{\lambda,R}| - \log|\Psi_\lambda(M)|}{d\log|\mathcal{O}/p|},$$

which we will denote by $\delta_\lambda(M)$ when R is clear from context.

Note that $\delta_{\lambda,R}(M)$ is a rational number. Also note that by this definition, if E'/E is a finite extension with ring of integers $\mathcal{O}'$, then $\delta_{\lambda\otimes\mathcal{O}',R\otimes\mathcal{O}'}(M\otimes\mathcal{O}') = \delta_{\lambda,R}(M)$. Essentially, $\delta_{\lambda,R}(M)$ is independent of our choice of coefficient ring. This allows us to assume that $\mathcal{O}$ is as large as needed in the arguments below.

Also note that for any $k \geq 1$, $\delta_{\lambda,R}(M^{\oplus k}) = \delta_{\lambda,R}(M)$ by Lemma 3.6.

As shown in [Dia97], we have the following result.

THEOREM 3.11. *Let R be as above and let M be a reflexive R-module. Let $d = \operatorname{rank}_\lambda M$ and assume that M has rank at most d at each generic point of R. Then $\delta_{\lambda,R}(M) \geq 0$ and we have $\delta_{\lambda,R}(M) = 0$ if and only if R is a complete intersection and M is free of rank d over R.*

In the situation we will consider in this paper, R will not be a complete intersection (or even Gorenstein) and M will not be free over R and so both of the Wiles defects $\delta_{\lambda,R}(R)$ and $\delta_{\lambda,R}(M)$ will be positive. Our goal will be to determine these defects.

We will use the following theorem below to compare $\delta_{\lambda,R}(R)$ and $\delta_{\lambda,R}(M)$.

THEOREM 3.12. *Let M be a reflexive R-module of λ-rank 1. Then $|\Psi_\lambda(R)| \geq |\Psi_\lambda(M)|$ and so $\delta_{\lambda,R}(R) \leq \delta_{\lambda,R}(M)$. Equality holds if the natural map $R \to \operatorname{End}_R(M)$ is an isomorphism.*

Proof. First note that $M^*[\ker\lambda] \subseteq M^* = \operatorname{Hom}_\mathcal{O}(M,\mathcal{O})$ consists precisely of the $\mathcal{O}$-module homomorphisms $f : M \to \mathcal{O}$ satisfying $f(rx) = \lambda(r)f(x)$ for all $r \in R$, $x \in M$.

Let $R[\ker\lambda] = \mathcal{O}\alpha$ and $M^*[\ker\lambda] = \mathcal{O}\beta$ for some $\alpha \in R$, $\beta \in M^*$. Regarding β as a map $\beta : M \to \mathcal{O}$, note that β must be surjective (as otherwise we would have $\varpi^{-1}\beta \in M^*[\ker\lambda]$, contradicting the choice of β).

By Lemmas 3.4 and 3.8, $\eta_\lambda(R) = (\lambda(\alpha))$ and $\eta_\lambda(M) = \beta(M[\ker\lambda])$. Now, as $\alpha \in R[\ker\lambda]$, we clearly have $\alpha M \subseteq M[\ker\lambda]$ and thus

$$\eta_\lambda(M) = \beta(M[\ker\lambda]) \supseteq \beta(\alpha M) = \lambda(\alpha)\beta(M) = \mathcal{O}\lambda(\alpha) = \eta_\lambda(R),$$

which automatically implies that $|\Psi_\lambda(R)| \geq |\Psi_\lambda(M)|$.

Moreover, if we have $|\Psi_\lambda(R)| > |\Psi_\lambda(M)|$, then we must have $\beta(\alpha M) \subsetneq \beta(M[\ker\lambda])$ and so $\beta(\alpha M) \subseteq \varpi\beta(M[\ker\lambda]) = \beta(\varpi M[\ker\lambda])$. As $\alpha M \subseteq M[\ker\lambda]$ and $\beta|_{M[\ker\lambda]} : M[\ker\lambda] \to \mathcal{O}$ is injective (as $M[\ker\lambda]$ is a free rank-1 $\mathcal{O}$-module), this implies that $\alpha M \subseteq \varpi M[\ker\lambda] \subseteq \varpi M$.

As M is ϖ-torsion free, it follows that we may define an endomorphism $f : M \to M$ by $f(x) = \varpi^{-1}\alpha x$. Clearly, $f \in \operatorname{End}_R(M)$ and, if $i : R \to \operatorname{End}_R(M)$ is the natural map, we have $i(\alpha) = \varpi f$. In particular, $i(\alpha) \in \varpi\operatorname{End}_R(M)$. But now as $R[\ker\lambda]$ is saturated in R, $\alpha \notin \varpi R$ and so $i : R \to \operatorname{End}_R(M)$ cannot be an isomorphism.

Thus, if we assume that i is an isomorphism, it follows that $|\Psi_\lambda(R)| = |\Psi_\lambda(M)|$. □

Remark 3.13. It is worth pointing out that the converse to the second claim in Theorem 3.12 is not true. That is, it is possible that $\operatorname{rank}_\lambda M = 1$ and $|\Psi_\lambda(R)| = |\Psi_\lambda(M)|$ for some λ (and, in fact, for *every* λ), but $R \to \operatorname{End}_R(M)$ fails to be an isomorphism.

To see this, let $\tilde{R}$ be any reduced, local ϖ-torsion free $\mathcal{O}$-algebra, and let $\tilde{\lambda}_1 : \tilde{R} \to \mathcal{O}^{(1)}, \ldots, \tilde{\lambda}_r : \tilde{R} \to \mathcal{O}^{(r)}$ be all of the augmentations of $\tilde{R}$ (where each $\mathcal{O}^{(i)}$ is a finite extension

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

of $\mathcal{O}$). For the sake of simplicity, assume that $\mathcal{O}^{(1)} = \cdots = \mathcal{O}^{(r)} = \mathcal{O}$ (and so $\tilde{R}[1/\varpi] \cong E^r$). For each i, write $\tilde{R}[\ker \tilde{\lambda}_i] = \mathcal{O}\alpha_i$.

Now let $R \subseteq \tilde{R}$ be the $\mathcal{O}$-subalgebra of $\tilde{R}$ generated by $\alpha_1, \ldots, \alpha_r$. It is not hard to see that R is also a reduced, local ϖ-torsion free $\mathcal{O}$-algebra, that $R[1/\varpi] = \tilde{R}[1/\varpi] \cong E^r$ and that the augmentations of R are all of the form $\lambda_i := \tilde{\lambda}_i|_R : R \to \mathcal{O}$. But one can now easily check that $R[\ker \lambda_i] = \mathcal{O}\alpha_i = \tilde{R}[\ker \tilde{\lambda}_i] = \tilde{R}[\ker \lambda_i]$ for all i. So, treating $\tilde{R}$ as an R-module, we now see that $\operatorname{rank}_{\lambda_i} \tilde{R} = 1$ and $\Psi_{\lambda_i}(R) = \mathcal{O}/(\lambda_i(\alpha_i)) = \Psi_{\tilde{\lambda}_i}(\tilde{R})$ for all i. On the other hand, we clearly have $\tilde{R} \subseteq \operatorname{End}_R(\tilde{R})$, so, provided that $R \neq \tilde{R}$ (that is, that $\tilde{R}$ is not generated as an $\mathcal{O}$-algebra by $\alpha_1, \ldots, \alpha_r$), the map $R \to \operatorname{End}_R(\tilde{R})$ cannot be an isomorphism.

It is easy to find examples of rings $\tilde{R}$ for which $R \neq \tilde{R}$, so in general the hypothesis that $R \to \operatorname{End}_R(M)$ is an isomorphism is strictly stronger than the hypothesis that $|\Psi_\lambda(R)| = |\Psi_\lambda(M)|$ for some, or even all, λ.

4. Galois deformation theory

4.1 Generalities

We use § 5 of [Tho16] as reference for the material recalled in this subsection.

Let p be an odd prime and let E be a coefficient field with valuation ring $\mathcal{O}$ and residue field k. We fix a continuous, absolutely irreducible representation $\bar{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ and assume that $\det \bar{\rho} = \epsilon$, the cyclotomic character. We will assume that k contains the eigenvalues of all elements in the image of $\bar{\rho}$. We also fix a finite set S of primes containing p, and all primes at which $\bar{\rho}$ is ramified (we also allow S to contain finitely many other primes in addition to these).

Let $q \in S$. We write $\mathcal{D}_q^{\square} : \mathrm{CNL}_{\mathcal{O}} \to \mathrm{Sets}$ for the functor that associates to $R \in \mathrm{CNL}_{\mathcal{O}}$ the set of all continuous homomorphisms $r : G_{\mathbf{Q}_q} \to \mathrm{GL}_2(R)$ such that $r \bmod \mathfrak{m}_R = \bar{\rho}|_{G_{\mathbf{Q}_q}}$ and $\det r$ is the cyclotomic character ϵ. It is easy to see that $\mathcal{D}_q^{\square}$ is represented by an object $R_q^{\square} \in \mathrm{CNL}_{\mathcal{O}}$.

Definition 4.1. Let $q \in S$. A local deformation problem for $\bar{\rho}|_{G_{\mathbf{Q}_q}}$ is a subfunctor $\mathcal{D}_q \subset \mathcal{D}_q^{\square}$ satisfying the following conditions.

- $\mathcal{D}_q$ is represented by a quotient R_q of $R_q^{\square}$.
- For all $R \in \mathrm{CNL}_{\mathcal{O}}$, $a \in \ker(\mathrm{GL}_2(R) \to \mathrm{GL}_2(k))$ and $r \in \mathcal{D}_q(R)$, we have $ara^{-1} \in \mathcal{D}_q(R)$.

We will write $\rho_q^{\square} : G_{\mathbf{Q}_q} \to \mathrm{GL}_2(R_q^{\square})$ for the universal lifting. If a quotient R_q of $R_q^{\square}$ corresponding to a local deformation problem $\mathcal{D}_q$ has been fixed, we will write $\rho_q : G_{\mathbf{Q}_q} \to \mathrm{GL}_2(R_q)$ for the universal lifting of type $\mathcal{D}_q$.

We recall [Tho16, Lemma 5.12].

Lemma 4.2. *Let $R_q \in \mathrm{CNL}_{\mathcal{O}}$ be a quotient of $R_q^{\square}$ satisfying the following conditions.*

1. *The ring R_q is reduced and not isomorphic to k.*
2. *Let $r : G_{\mathbf{Q}_q} \to \mathrm{GL}_2(R_q)$ denote the specialization of the universal lifting and let $a \in \ker(\mathrm{GL}_2(R_q) \to \mathrm{GL}_2(k))$. Then the homomorphism $R_q^{\square} \to R_q$ associated to the representation ara^{-1} by universality factors through the canonical projection $R_q^{\square} \to R_q$.*

Then the subfunctor of $\mathcal{D}_q^{\square}$ defined by R_q is a local deformation problem.

We now specialize to the situation of this paper. Consider our fixed $\bar{\rho}$, and the integers N and Q, with $N|N(\bar{\rho})|NQ$, that we have fixed. (Recall that all the primes in Q are Steinberg for $\bar{\rho}$.) Let S be the set of primes which divide $N(\bar{\rho})pQ$. We briefly recall the local and global deformation rings that we consider in this paper and some of their known properties. For convenience, we

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

will also assume that the coefficient ring E contains all (q^2-1)th roots of unity for each $q|Q$ (recall that as shown in the previous section, the Wiles defect is unaffected by expanding the coefficient ring, so this assumption is harmless).

4.2 Local deformation rings

For any prime q, let $R_q^{\square} \in \mathrm{CNL}_{\mathcal{O}}$ be the full framed local deformation ring of $\bar{\rho}|_{G_q}$ for lifts of fixed determinant ϵ. We consider below quotients of $R_q^{\square}$ which are defined as in [Kis09] as reduced, flat over $\mathcal{O}$ quotients of $R_q^{\square}$, and are characterized by the $\bar{\mathbf{Q}}_p$-valued points of their generic fiber; [Kis09] computed the dimension of generic fibers of the quotients we consider, and proved that they are regular. The quotients we consider satisfy Lemma 4.2 and thus give rise to deformation conditions in the sense of Definition 4.1. We need finer integral properties of the quotients we consider that are stated in Proposition 4.3.

For $\ell|N$:

- let $R_\ell^{\min}$ be the minimally ramified quotient of R_ℓ (which parametrizes minimally ramified deformations) and let R_p^{fl} be the flat quotient of R_p (which parametrizes flat deformations).

For $q|Q$, the Steinberg quotient R_q^{St} of $R_q^{\square}$ is defined as follows.

- If q is ramified in $\bar{\rho}$, then the Steinberg quotient R_q^{St} is defined as the minimally ramified quotient of $R_q^{\square}$.
- If q is unramified in $\bar{\rho}$ and $q \not\equiv -1 \pmod{p}$, then the Steinberg quotient R_q^{St} is the unique reduced quotient of $R_q^{\square}$ characterized by the fact that the L-valued points of its generic fiber, for any finite extension L/E, correspond to representations which up to an unramified twist are of the form $\left(\begin{smallmatrix} \epsilon & * \\ 0 & 1 \end{smallmatrix}\right)$.
- If q is unramified in $\bar{\rho}$ and $q \equiv -1 \pmod{p}$, then $\bar{\rho}(\mathrm{Frob}_q)$ takes two distinct eigenvalues $\alpha_q, \beta_q \in k$ such that $\alpha_q/\beta_q = q$ (note that $\alpha_q, \beta_q = \pm 1$). Thus, $\bar{\rho}|_{G_q}$ in some basis e_1, e_2 is of the form
$$\begin{pmatrix} \varepsilon\bar{\chi} & 0 \\ 0 & \bar{\chi} \end{pmatrix},$$
where we assume that $\bar{\chi}(\mathrm{Frob}_q) = \beta_q$. We define $R_q^{\mathrm{St}(\beta_q)}$ as the unique reduced quotient of $R_q^{\square}$ characterized by the fact that the L-valued points of its generic fiber, for any finite extension L/E, correspond to representations of the form
$$\begin{pmatrix} \varepsilon\chi & * \\ 0 & \chi \end{pmatrix}$$
such that χ reduces to $\bar{\chi}$.

  Below we will fix a β_q and, for uniformity of notation, generally we write R_q^{St} for $R_q^{\mathrm{St}(\beta_q)}$.

For $q|Q$ unramified in $\bar{\rho}$, the unipotent quotient R_q^{uni} of $R_q^{\square}$ is defined as:

- the unique reduced quotient of $R_q^{\square}$ such that $\operatorname{Spec} R_q^{\mathrm{uni}} = \operatorname{Spec} R_q^{\mathrm{St}} \cup \operatorname{Spec} R_q^{\min}$ inside $\operatorname{Spec} R_q^{\square}$, where here $R_q^{\min}$ is the quotient of $R_q^{\square}$ that parametrized unramified deformations.

  If $q \equiv -1 \pmod{p}$, we write $R_q^{\mathrm{uni}(\beta_q)}$ for R_q^{uni} if we wish to indicate the β_q used to define R_q^{St}.

Recall from [BGHT11, § 3] that $R \in \mathrm{CNL}_{\mathcal{O}}$ is called *geometrically integral* if, for all finite field extensions $E' \supset E$ with integral closure $\mathcal{O}'$ of $\mathcal{O}$, the ring $R \otimes_{\mathcal{O}} \mathcal{O}'$ is an integral domain.

The following result summarizes basic ring-theoretic properties of the $R_q^?$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Proposition 4.3. *The following hold.*

1. *We have* $R_q^{\mathrm{fl}} \cong \mathcal{O}[[X_1, X_2, X_3, X_4]]$ *for* $q = p$ *and* $R_q^{\min} \cong \mathcal{O}[[X_1, X_2, X_3]]$ *for all* $q \neq p$.
2. *For* $q \neq p$, *the ring* $R_q^{\square}$ *is a complete intersection, reduced and flat over* $\mathcal{O}$ *of relative dimension* 3.
3. *For* $q|Q$, *the ring* R_q^{St} *is Cohen–Macaulay, flat of relative dimension* 3 *over* $\mathcal{O}$ *and geometrically integral and, if* q *is not a trivial prime for* $\bar{\rho}$, *we in fact have* $R_q^{\mathrm{St}} \cong \mathcal{O}[[X_1, X_2, X_3]]$.
4. *For each* $q|Q$ *and each minimal prime* $\mathfrak{p}$ *of* $R_q^{\square}$, $R_q^{\square}/\mathfrak{p}$ *is flat over* $\mathcal{O}$ *and geometrically integral.*
5. *For* $q|Q$ *such that* $\bar{\rho}$ *is unramified, the reduced ring* R_q^{uni} *is Cohen–Macaulay and flat over* $\mathcal{O}$ *of relative dimension* 3.

Moreover, the rings $R_q^{?}$ *in 1–5 are the completion of a finite-type* $\mathcal{O}$*-algebra at a maximal ideal.*

Proof. Assertion 1 is from [CHT08, §§ 2.4.1 and 2.4.4] and Assertion 2 from [Sho18, Theorem 2.5], except that one has to modify the statements and proofs there by including the further condition that the determinant be ϵ, which in turn results in a drop of dimensions by 1. Assertion 3 is from [Sho16, § 5], where here one has to impose the additional condition that the image of Frobenius has determinant 1 and one has to modify the proofs accordingly. That R_q^{St} is geometrically integral is not stated explicitly in [Sho16]. However, the base changes $R_q^{\mathrm{St}} \otimes_{\mathcal{O}} \mathcal{O}'$ have again an interpretation as a universal Steinberg ring and are thus all integral domains. Assertion 4 follows similarly from the results in [Sho16, § 5].

Assertion 5 will be addressed in the proof of Proposition 7.9, where we give a description of R_q^{uni} as a quotient of a power series ring over $\mathcal{O}$ by some explicit relations. In that proof we also recall more detailed information about R_q^{St} and $R_q^{\square}$; this would easily provide a direct proof of Assertion 2 except for the reducedness.

In all cases it is obvious from the explicit nature of the rings $R_q^{?}$ given in the references that the rings are the completion of a ring of finite type over $\mathcal{O}$ at a maximal ideal. □

Define $R_q = R_q^{\square}$ for $q \in Q$ and further:

$$R_{\mathrm{loc}} = \left[\widehat{\bigotimes_{q|Q}} R_q\right] \widehat{\otimes} \left[\widehat{\bigotimes_{\ell|N}} R_\ell^{\min}\right] \widehat{\otimes} R_p^{\mathrm{fl}} \quad \text{and} \quad R_{\mathrm{loc}}^{\mathrm{St}} = \left[\widehat{\bigotimes_{q|Q}} R_q^{\mathrm{St}}\right] \widehat{\otimes} \left[\widehat{\bigotimes_{\ell|N}} R_\ell^{\min}\right] \widehat{\otimes} R_p^{\mathrm{fl}}.$$

Lemma 4.4. *Suppose that* $\mathcal{R}_i$, $i = 1, 2$, *lie in* $\mathrm{CNL}_{\mathcal{O}}$ *and are flat over* $\mathcal{O}$. *Let* $\mathcal{R} = \mathcal{R}_1 \widehat{\otimes}_{\mathcal{O}} \mathcal{R}_2$ *be their completed tensor product. Then the following hold.*

1. *The ring* $\mathcal{R}$ *is flat over* $\mathcal{O}$.
2. *If both* $\mathcal{R}_i$ *are Cohen–Macaulay, then so is* $\mathcal{R}$.
3. *If both* $\mathcal{R}_i$ *are Gorenstein, then so is* $\mathcal{R}$.
4. *If both* $\mathcal{R}_i$ *are local complete intersection rings, then so is* $\mathcal{R}$.
5. *If both* $\mathcal{R}_i$ *are reduced and if* $\mathcal{R}_2$ *is the completion of a finite-type* $\mathcal{O}$*-algebra* R_2 *at a maximal ideal* $\mathfrak{m}_2$, *then* $\mathcal{R}$ *is reduced.*
6. *If both* $\mathcal{R}_i$ *are geometrically integral, then so is* $\mathcal{R}$.
7. *If the minimal primes of* $\mathcal{R}_1$ *are* $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ *and the minimal primes of* $\mathcal{R}_2$ *are* $\mathfrak{q}_1, \dots, \mathfrak{q}_s$ *and each* $\mathcal{R}_1/\mathfrak{p}_i$ *and* $\mathcal{R}_2/\mathfrak{q}_j$ *is flat over* $\mathcal{O}$ *and geometrically integral, then the distinct minimal primes of* $\mathcal{R}$ *are* $\mathfrak{p}_i\mathcal{R} + \mathfrak{q}_j\mathcal{R}$ *for* $i = 1, \dots, r$ *and* $j = 1, \dots, s$.
8. *If both* $\mathcal{R}_i$ *are the completion of a finite-type* $\mathcal{O}$*-algebra* R_i *at a maximal ideal* $\mathfrak{m}_i$, *then so is* $\mathcal{R}$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Proof. Part 1 is proved in the first paragraph of [Kis09, Proof of Lemma (3.4.12)].

For part 2, recall from [Sta19, Lemma 045J] that for a flat local homomorphism $R \to S$ of local Noetherian rings, the ring S is Cohen–Macaulay if and only if R and $S/\mathfrak{m}_R S$ are Cohen–Macaulay. We deduce on the one hand that the rings $\mathcal{R}_i/\pi\mathcal{R}_i$ are Cohen–Macaulay for $i = 1, 2$ and on the other that to show that $\mathcal{R}$ is Cohen–Macaulay, we need to establish that $\mathcal{R}/\pi\mathcal{R} = (\mathcal{R}_1/\pi\mathcal{R}_1)\widehat{\otimes}_k(\mathcal{R}_2/\pi\mathcal{R}_2)$ is Cohen–Macaulay.

Now, by the Cohen structure theorem in the form given in [Sta19, Lemma 032D], there exist power series rings $\mathcal{S}_i \subset \mathcal{R}_i/\pi\mathcal{R}_i$ over k such that $\mathcal{R}_i/\pi\mathcal{R}_i$ is finite over $\mathcal{S}_i$, $i = 1, 2$.[3] Since $\mathcal{R}_i/\pi\mathcal{R}_i$ is Cohen–Macaulay, it follows from [Sta19, Lemma 00R4] that $\mathcal{R}_i/\pi\mathcal{R}_i$ is free of finite rank over $\mathcal{S}_i$. Hence, $\mathcal{R}/\pi\mathcal{R}$ is a ring extension of the power series ring $\mathcal{S} = \mathcal{S}_1\widehat{\otimes}_k\mathcal{S}_2$ that is free of finite rank over $\mathcal{S}$. The ring $\mathcal{S}$ is regular and in particular Cohen–Macaulay, and all fibers of $\mathcal{S} \to \mathcal{R}/\pi\mathcal{R}$ are Artin rings and thus also Cohen–Macaulay. Invoking again [Sta19, Lemma 045J], we see that $\mathcal{R}/\pi\mathcal{R}$ is Cohen–Macaulay. Let us mention that our proof of part 2 is inspired by [CEG+16, Appendix A].

The first half of the proof of part 2 goes through with Gorenstein or local complete intersection instead of Cohen–Macaulay, as well by appealing to [Sta19, Proposition 0BJL] and [Sta19, Proposition 09Q2]. The remaining argument for part 3 is [WITO69, Theorem 6]. There is also a simple direct proof. One can combine systems of parameters for $\mathcal{R}_i/\pi\mathcal{R}_i$, $i = 1, 2$, to obtain a system of parameters for $(\mathcal{R}_1/\pi\mathcal{R}_1)\widehat{\otimes}_k(\mathcal{R}_2/\pi\mathcal{R}_2)$ and then deduce that the socle of the resulting quotient has k-dimension 1 using that this holds for the quotients of the individual factors. Even more elementary is the completion of the argument for part 4. We write $\mathcal{R}_i/\pi\mathcal{R}_i$ as a quotient of a power series ring $\mathcal{S}_i$ over k by an ideal $\mathcal{I}_i = (f_{i,1}, \dots, f_{i,h_i})$ such that $h_i = \dim \mathcal{S}_i - \dim \mathcal{R}_i/\pi\mathcal{R}_i$. Then $\mathcal{R}/\pi\mathcal{R} = \mathcal{S}/\mathcal{I}$ for $\mathcal{S}$ the power series ring $\mathcal{S}_1\widehat{\otimes}_k\mathcal{S}_2$ over k and $\mathcal{I} = (\{f_{1,j} \otimes 1 : j \in \{1, \dots, h_1\}\} \cup \{1 \otimes f_{2,j} : j \in \{1, \dots, h_2\})$, so that $\dim \mathcal{S} - \dim \mathcal{R}/\pi\mathcal{R} = h_1 + h_2$.

To establish part 5, observe first that the $\mathcal{R}_i$ as well as R_2 and $\mathcal{R}_1 \otimes R_2$ are Nagata rings by [Sta19, Proposition 0335]. By [Sta19, Lemma 07NZ], the completion of a Nagata ring is reduced if and only if the ring itself is reduced. From this and our hypothesis, we see that $\mathcal{R}_1$ and R_2 are reduced and it suffices to show that $\mathcal{R}_1 \otimes_{\mathcal{O}} R_2$ is reduced. Now $\mathcal{R}_1$ and R_2 each embed into a finite product of fields $\mathcal{K}_j$ and $K_{j'}$. These fields must contain E and it follows that $\mathcal{R}_1 \otimes_{\mathcal{O}} R_2$ embeds into the finite product of the rings $\mathcal{K}_j \otimes_E K_{j'}$. The latter are reduced because E is of characteristic 0 and thus $K_{j'}$ is separable over E; see [Sta19, Lemma 030U]. Therefore, $\mathcal{R}_1 \otimes_{\mathcal{O}} R_2$ has no non-zero nilpotent elements and is hence reduced.

Parts 6 and 7 are from [BGHT11, Lemma 3.3] and part 8 is obvious. □

Proposition 4.3 and Lemma 4.4 yield the following result.

PROPOSITION 4.5. *The ring R_{loc} is a reduced complete intersection, the ring $R^{\mathrm{St}}_{\mathrm{loc}}$ is a Cohen–Macaulay domain and both are flat over $\mathcal{O}$.*

Moreover, each irreducible component of $\operatorname{Spec} R_{\mathrm{loc}}$ *is of the form*

$$\operatorname{Spec}\left[\widehat{\bigotimes_{q|Q}} R^{\square}_q/\mathfrak{p}^{(q)}\right]\widehat{\otimes}\left[\widehat{\bigotimes_{\ell|N}} R^{\min}_{\ell}\right]\widehat{\otimes} R^{\mathrm{fl}}_p,$$

where each $\operatorname{Spec} R_q/\mathfrak{p}^{(q)}$ *is an irreducible component of* $\operatorname{Spec} R^{\square}_q$ *(that is, each $\mathfrak{p}^{(q)}$ is a minimal prime of R_q).*

[3] That statement of [Sta19, Lemma 032D] assumes the ring to be a domain. However, its proof (Case I, with $\Lambda = k$) does not make use of this hypothesis.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

4.3 Global deformation rings

Now we set up the notation for the corresponding global deformation rings that we use later.

Let $R^{\square}$ (respectively $R^{\square,\mathrm{St}}$) be the framed global deformation rings that parametrize strict equivalence classes of tuples $(\rho, \{t_q\}_{q\in S})$, where:

- $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(R)$ is a lifting of $\bar{\rho}$, with $R \in \mathrm{CNL}_{\mathcal{O}}$, of determinant ϵ;
- ρ is unramified outside S, of determinant ϵ and, for each $\ell \in S$, $\rho|_{G_\ell}$ satisfies the local deformation condition arising from:
  - $R_\ell^{\min}$ (for $\ell|N$), R_p^{fl} (for $\ell = p$), $R_\ell^{\square}$ (for $\ell \in Q$);
  - respectively, $R_\ell^{\min}$ (for $\ell|N$), R_p^{fl} (for $\ell = p$), R_ℓ^{St} (for $\ell \in Q$);
- t_q is an element of $\ker(\mathrm{GL}_2(R) \to \mathrm{GL}_2(k))$.

Here two framed liftings $(\rho_1, \{t_q\}_{q\in S})$ and $(\rho_2, \{u_q\}_{q\in S})$ are said to be strictly equivalent if there is an element $a \in \ker(\mathrm{GL}_2(R) \to \mathrm{GL}_2(k))$ such that $\rho_2 = a\rho_1 a^{-1}$ and $u_q = at_q$ for each $q \in S$.

As $\bar{\rho}$ is irreducible, we also have the corresponding unframed rings R and R^{St}, which are canonically $\mathcal{O}$-subalgebras of $R^{\square}$ and $R^{\square,\mathrm{St}}$, and furthermore (non-canonically) $R^{\square} = R[[X_1, \ldots, X_{4j-1}]]$ and $R^{\square,\mathrm{St}} = R^{\mathrm{St}}[[X_1, \ldots, X_{4j-1}]]$ with j the cardinality of S. (We will also by harmless abuse of notation sometimes view R and R^{St} as (non-canonical) quotients of the corresponding framed deformation rings by the framing variables X_i.)

The natural transformation $(\rho, \{t_q\}_{q\in S}) \mapsto (t_q^{-1}\rho|_{G_{\mathbf{Q}_q}} t_q)_{q\in S}$ induces canonical maps $R_{\mathrm{loc}} \to R^{\square}$ and $R_{\mathrm{loc}}^{\mathrm{St}} \to R^{\square,\mathrm{St}}$, which are homomorphisms in $\mathrm{CNL}_{\mathcal{O}}$, and we have $R^{\square,\mathrm{St}} = R^{\square} \otimes_{R_{\mathrm{loc}}} R_{\mathrm{loc}}^{\mathrm{St}}$.

5. Hecke algebras and newforms

5.1 Modular curves and Shimura curves

From now on, let D be a quaternion algebra over $\mathbf{Q}$ (either definite or indefinite) and let Q be the set of finite primes at which D is ramified. (By abuse of notation, we will also frequently use Q to denote the product of all the primes in the set Q. The context will make clear which meaning is intended.) As all of our results are trivial in the case when $Q = \varnothing$, we will assume that $Q \neq \varnothing$, for convenience. Fix a square-free integer N not divisible by any prime in Q. Let $\Gamma_0(N\Pi_{q\in Q} q^2)$ and $\Gamma_0^Q(N)$ be congruence subgroups for GL_2 and $D^\times$, respectively, where $\Gamma_0^Q(N)$ is maximal compact at primes in Q and upper triangular mod ℓ for all $\ell|N$. We will denote $\Gamma_0(N\Pi_{q\in Q} q^2)$ by $\Gamma_0(NQ^2)$ using our abuse of notation.

We consider a prime p below not dividing $2NQ$ and a finite extension $E/\mathbf{Q}_p$, and let $\mathcal{O}$ be the ring of integers in E, ϖ a uniformizer and $k = \mathcal{O}/\varpi$ the residue field. We will assume below that E is sufficiently large so that $\mathcal{O}$ contains the Fourier coefficients of all newforms in $S_2(\Gamma_0(NQ))$.

Let $K_0(NQ^2) \subseteq \mathrm{GL}_2(\mathbf{A}_{\mathbf{Q},f})$ and $K_0^Q(N) \subseteq D^\times(\mathbf{A}_{\mathbf{Q},f})$ be the corresponding compact open subgroups. If D is definite, let

$$S^Q(\Gamma_0^Q(N)) = \{f : D^\times(\mathbf{Q})\backslash D^\times(\mathbf{A}_{\mathbf{Q},f})/K_0^Q(N) \to \mathcal{O}\}.$$

If D is indefinite, let $X_0^Q(N)$ be the (compact) Riemann surface $D^\times(\mathbf{Q})\backslash\big(D^\times(\mathbf{A}_{\mathbf{Q},f}) \times \mathcal{H}\big)/K_0^Q(N)$ (where $\mathcal{H}$ is the complex upper half plane). Give $X_0^Q(N)$ its canonical structure as an algebraic curve over $\mathbf{Q}$ and let $S^Q(\Gamma_0^Q(N)) = H^1(X_0^Q(N), \mathcal{O})$. We also consider $S(NQ^2) = H^1(X_0(NQ^2), \mathcal{O})$. Let $\mathbf{T}(NQ^2)$ and $\mathbf{T}^Q(N)$ be the $\mathcal{O}$-algebras at levels $\Gamma_0(NQ^2)$ and $\Gamma_0^Q(N)$, respectively, generated by the Hecke operators T_r for primes r coprime to pNQ acting on

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

$H^1(X_0(NQ^2),\mathcal{O})$ and $S^Q(\Gamma_0^Q(N))$. Note that by the Jacquet–Langlands correspondence, $\mathbf{T}^Q(N)$ is a quotient of $\mathbf{T}(NQ^2)$.

Let $f \in S_2(\Gamma_0(NQ))$ be a newform of level NQ such that all its Fourier coefficients lie in E, and consider the corresponding $\mathcal{O}$-algebra homomorphism $\lambda_f : \mathbf{T}(NQ^2) \to \mathcal{O}$. We will fix this newform and our main results will be in relation to f. By the Jacquet–Langlands correspondence, this also gives a related homomorphism $\mathbf{T}^Q(N) \to \mathcal{O}$ that we denote by the same symbol λ_f. We denote the corresponding maximal ideals which contain the prime ideal $\ker(\lambda_f)$ by $\mathfrak{m}$. Let $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O})$ be the Galois representation associated by Eichler and Shimura to f and assume that the corresponding residual Galois representation $\bar{\rho}_f = \bar{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ is absolutely irreducible. By enlarging $\mathcal{O}$ if necessary, we may assume that k contains all eigenvalues of $\bar{\rho}(\sigma)$ for all $\sigma \in G_{\mathbf{Q}}$. We assume that $N|N(\bar{\rho})|NQ$. The Galois representation $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(E)$, with irreducible residual representation $\bar{\rho}$, is locally at q of the form

$$\begin{pmatrix} \epsilon & * \\ 0 & 1 \end{pmatrix},$$

up to twist by an unramified character χ of order dividing 2. The $\beta_q \in \{\pm 1\}$ of the previous section will be chosen so that $\rho_f|_{G_q}$ gives rise to a point of $\mathrm{Spec}\, R_q^{\mathrm{St}(\beta_q)}$ in what follows (and thus depends on whether χ is trivial at q or not).

There is an oldform f^Q in $S_2(\Gamma_0(NQ^2))$ with corresponding newform f which is characterized by the property that it is an eigenform for the Hecke operators T_ℓ for ℓ prime with $(\ell, NQ^2) = 1$ and U_ℓ for $\ell|NQ^2$ and such that $a_q(f^Q) = 0$, i.e. $f^Q|U_q = 0$ for $q \in Q$. Let $\lambda_{f^Q} : \mathbf{T}^{\mathrm{full}}(NQ^2) \to \mathcal{O}$ be the induced homomorphism of the full Hecke algebra $\mathbf{T}^{\mathrm{full}}(NQ^2)$ acting on $H^1(X_0(NQ^2),\mathcal{O})$ which is generated as an $\mathcal{O}$-algebra by the action of the Hecke operators T_ℓ for $(\ell, NQ^2) = 1$ and U_ℓ for $\ell|NQ^2$ on $S(NQ^2) = H^1(X_0(NQ^2),\mathcal{O})$. Let $\mathfrak{m}_Q$ be the maximal ideal of $\mathbf{T}^{\mathrm{full}}(NQ^2)$ that contains $\ker(\lambda_{f^Q})$.

By Ribet's level-lowering result [Rib90a], there is a newform $g \in S_2(\Gamma_0(N(\bar{\rho})))$ which gives rise to $\bar{\rho}$ and an augmentation $\lambda_g : \mathbf{T}(NQ^2) \to \mathcal{O}$ whose kernel is contained in the maximal ideal $\mathfrak{m}$. There is an oldform g^Q in $S_2(\Gamma_0(NQ^2))$ with corresponding newform g which is characterized by the property that it is an eigenform for the Hecke operators T_ℓ for $(\ell, NQ^2) = 1$ and U_ℓ for $\ell|NQ^2$ and such that $a_q(g^Q) = 0$, i.e. $g^Q|U_q = 0$ for $q \in Q$. Let $\lambda_{g^Q} : \mathbf{T}^{\mathrm{full}}(NQ^2) \to \mathcal{O}$ be the induced homomorphism of the full Hecke algebra $\mathbf{T}^{\mathrm{full}}(NQ^2)$ acting on $H^1(X_0(NQ^2),\mathcal{O})$ which is generated as an $\mathcal{O}$-algebra by the action of the Hecke operators T_ℓ for $(\ell, NQ^2) = 1$ and U_ℓ for $\ell|NQ^2$ on $S(NQ^2) = H^1(X_0(NQ^2),\mathcal{O})$. The kernel of λ_{g^Q} is contained in the maximal ideal $\mathfrak{m}_Q$ that f^Q gives rise to.

The homomorphism $\lambda_f : \mathbf{T}^Q(N) \to \mathcal{O}$ extends to the full Hecke algebra acting on $S^Q(\Gamma_0^Q(N))$, and we denote by $\mathfrak{m}'_Q$ the maximal ideal of $\mathbf{T}^Q(N)$ which contains its kernel.

We define $\mathbf{T}$ and $\mathbf{T}^{\mathrm{St}}$ to be the images of $\mathbf{T}(NQ^2)$ and $\mathbf{T}^Q(N)$ in the endomorphisms of the finitely generated $\mathcal{O}$-modules $H^1(X_0(NQ^2),\mathcal{O})_{\mathfrak{m}_Q}$ and $S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}$, respectively. We are primarily interested in $\mathbf{T}^{\mathrm{St}}$ in this paper, but we study it via looking at the surjective map of $\mathcal{O}$-algebras $\mathbf{T} \twoheadrightarrow \mathbf{T}^{\mathrm{St}}$ produced by the Jacquet–Langlands correspondence that yields a Hecke equivariant inclusion $S_2(\Gamma_0^Q(N)) \hookrightarrow S_2(\Gamma_0(NQ^2))$.

We have the corresponding universal modular deformation $\rho^{\mathrm{mod}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T})$ by results of Carayol, which is a specialization of the universal representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(R)$.

Define

$$M(NQ^2) = \mathrm{Hom}_{\mathbf{T}[G_{\mathbf{Q}}]}\left(\rho^{\mathrm{mod}}, S(NQ^2)^*_{\mathfrak{m}_Q}\right).$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

If D is definite, define $M^{\mathrm{St}}(N) = S^Q(\Gamma_0^Q(N))^*_{\mathfrak{m}'_Q}$ and, if D is indefinite, define

$$M^{\mathrm{St}}(N) = \mathrm{Hom}_{\mathbf{T}[G_{\mathbf{Q}}]}\left(\rho^{\mathrm{mod}}, S^Q(\Gamma_0^Q(N))^*_{\mathfrak{m}'_Q}\right).$$

As in [Car94], we have the following result.

Lemma 5.1. *The evaluation map $M(NQ^2) \otimes_{\mathbf{T}} \rho^{\mathrm{mod}} \to S(NQ^2)^*_{\mathfrak{m}_Q}$ is an isomorphism, as is $M^{\mathrm{St}}(N) \otimes_{\mathbf{T}} \rho^{\mathrm{mod}} \to S^Q(\Gamma_0^Q(N))^*_{\mathfrak{m}'_Q}$ when D is indefinite. In particular, as $\mathbf{T}$-modules we have $S(NQ^2)^*_{\mathfrak{m}_Q} = M(NQ^2)^{\oplus 2}$ and, when D is indefinite, $S^Q(\Gamma_0^Q(N))^*_{\mathfrak{m}'_Q} = M^{\mathrm{St}}(N)^{\oplus 2}$.*

We remark that, using the notation of § 8, below this implies that for any $\lambda : \mathbf{T} \to \mathcal{O}$, $\delta_{\lambda,\mathbf{T}}(S(NQ^2)_{\mathfrak{m}_Q}) = \delta_{\lambda,\mathbf{T}}(M(NQ^2))$ and, for any $\lambda : \mathbf{T}^{\mathrm{St}} \to \mathcal{O}$, $\delta_{\lambda,\mathbf{T}^{\mathrm{St}}}(S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}) = \delta_{\lambda,\mathbf{T}^{\mathrm{St}}}(M^{\mathrm{St}}(N))$.

We have the following key theorem due to Wiles [Wil95] and Diamond [Dia97].

Theorem 5.2. *The surjective map $R \twoheadrightarrow \mathbf{T}$ of complete Noetherian local $\mathcal{O}$-algebras is an isomorphism of complete intersections. Furthermore, $S(NQ^2)_{\mathfrak{m}_Q} = H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}$ is free of rank 2 over $\mathbf{T}$ and $M(NQ^2)$ is free of rank 1 over $\mathbf{T}$.*

Proof. This follows from the arguments of [Dia97, Theorem 3.4]. We sketch Diamond's argument. One first proves an $R_\varnothing = \mathbf{T}_\varnothing$ theorem as in [Dia97, Theorem 3.1] for minimal deformation and Hecke rings for $\bar\rho$ that come with an augmentation induced by a newform $g \in S_2(\Gamma_0(N(\bar\rho))$ as above. One also proves that these rings are complete isomorphisms and the appropriate localization of the cohomology of a modular curve $X_\varnothing$, $H^1(X_\varnothing, \mathcal{O})_{\mathfrak{m}}$, is free as a $\mathbf{T}_\varnothing$ module. Then one observes that $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q} \otimes E$ is free of rank 2 over $\mathbf{T} \otimes E$, which follows from the choice of $\mathfrak{m}_Q$ above so as to contain the kernel of the augmentation of the full Hecke algebra $\mathbf{T}^{\mathrm{full}}(NQ^2)$ induced by an oldform g^Q with associated newform g as above (see also [DDT97, Proposition 4.7]). Using level-raising arguments due to [Rib84], one then deduces that if x, y is a basis of the free $\mathcal{O}$ module $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}[\lambda_{g^Q}] = H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}[\lambda_g]$ of rank 2, then $\#\ker(\lambda_R)/\ker(\lambda_R)^2 \leq \#\mathcal{O}/\langle x, y\rangle_Q$, where the pairing $\langle\ ,\ \rangle_Q$ is a certain alternating $\mathbf{T}$-bilinear, $\mathcal{O}$-perfect pairing on $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}$. By appealing to [Dia97, Theorem 2.4], one deduces that $R \twoheadrightarrow \mathbf{T}$ is an isomorphism of complete intersections and $S(NQ^2)_{\mathfrak{m}_Q} = H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}$ is free of rank 2 over $\mathbf{T}$. In particular, $\mathbf{T}$ is Gorenstein. As $S(NQ^2)^*_{\mathfrak{m}_Q} = M(NQ^2)^{\oplus 2}$, we deduce from this that $M(NQ^2)$ is projective and, as $\mathbf{T}$ is a local ring, that $M(NQ^2)$ is a free $\mathbf{T}$ module of rank 1. □

Note that classical generic multiplicity-one results for modular forms imply that $M(NQ^2)$ and $M^{\mathrm{St}}(N)$ have rank 1 at each generic point of $\mathbf{T}^{\mathrm{St}}$. In contrast to what we have proved about $M(NQ^2)$, when Q contains a prime q that is trivial for $\bar\rho$, it is proved in [Man21] that $M^{\mathrm{St}}(N)$ is not free of rank 1 over $\mathbf{T}^{\mathrm{St}}$, nor is $\mathbf{T}^{\mathrm{St}}$ a complete intersection.

Remark 5.3. In [Dia97], a slightly less restricted deformation functor is considered than the one represented by R above (for instance, at primes dividing $N(\bar\rho)$ he puts no restriction on ramification while we have imposed minimality at primes dividing N). Whilst we could also work with his conditions, we have chosen to work with our slightly more restrictive deformation conditions at primes dividing N, and the arguments of [Dia97] we have quoted apply *mutatis mutandis*.

Remark 5.4. In [Dia97], the analogs of the rings R and $\mathbf{T}$ above are denoted R_Σ and $\mathbf{T}_\Sigma$ (where Σ is the set Q here), the maximal ideal $\mathfrak{m}_Q$ is denoted by $\mathfrak{m}_\Sigma$ and the analog of the modular

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

curve $X_0(NQ^2)$ we consider here is denoted by X_Σ. We remark that as in [Dia97], although $\mathbf{T}$ is an 'anemic Hecke algebra', defined without including the operators U_q for $q \in Q$, the maximal ideal $\mathfrak{m}_Q$ is a suitable maximal ideal of the full Hecke algebra with $U_q \in \mathfrak{m}_Q$. This ensures that $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q} \otimes E$ is free of rank 2 over $\mathbf{T} \otimes E$, which is a pre-requisite to apply [Dia97, Theorem 2.4] and get the refined integral statement that $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}$ is free of rank 2 over $\mathbf{T}$. This way of avoiding use of Hecke operators U_q for $q|Q$, when proving that $R \twoheadrightarrow \mathbf{T}$ is an isomorphism with non-minimal conditions at Q, goes back to the original arguments of Wiles in [Wil95, Chapter 3] that prove modularity results in the non-minimal case. Wiles used the numerical criterion Theorem 1.1, modularity results in the minimal case and level raising. Diamond further enhances the argument in [Dia97] to get that $H^1(X_0(NQ^2), \mathcal{O})_{\mathfrak{m}_Q}$ is free of rank 2 over $\mathbf{T}$, without having to know *a priori* Gorenstein properties of Hecke algebras.

It is an interesting feature of the proof of the main result of this paper, Theorem 10.1 below, that to prove results about the Wiles defect of the weight-2 newform f of level NQ that we are interested in, we have to use another newform of weight 2 and minimal level $N(\bar{\rho})$ via its use in the proof of Theorem 5.2.

5.2 Abelian varieties of GL_2 type with multiplicative reduction at q

We note a useful result about Tate modules of abelian varieties A over $\mathbf{Q}$ of GL_2-type, with field of endomorphisms over $\mathbf{Q}$ a number field F, and which have multiplicative reduction at a prime q. Let p be a prime and consider $F \otimes \mathbf{Q}_p = \prod_v F_v$, which induces a factorization $\mathrm{Ta}_p(A) \otimes \mathbf{Q}_p = \prod_v \mathrm{Ta}_v(A)$ and $\mathrm{Ta}_v(A) = F_v^2$, and we denote by $\rho_{A,v} : G_{\mathbf{Q}} \to \mathrm{GL}_2(F_v)$ the representation of $G_{\mathbf{Q}}$ arising from its action on $\mathrm{Ta}_v(A)$. If $\mathcal{O}_v$ as usual is the valuation ring of F_v, $\rho_{A,v}$ has an integral model with values in $\mathrm{GL}_2(\mathcal{O}_v)$. We assume that the determinant of $\rho_{A,v}$ is the p-adic cyclotomic character ε, which is assured in our applications as $A = A_f$ arise from a newform f in $S_2(\Gamma_0(M))$ for some positive integer M.

Lemma 5.5.

1. *If $\rho_{A,v}$ is residually irreducible, then $\rho_{A,v}$ is integrally well defined, i.e. $G_{\mathbf{Q}}$ preserves a unique lattice of E_v^2 up to scaling.*
2. *Furthermore, the restriction $\rho_{A,v}|_{G_q}$ of an integral model of $\rho_{A,v}$ is $\mathrm{GL}_2(\mathcal{O}_v)$-conjugate to an unramified twist of*

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}.$$

The reducible representation $\rho_{A,v}|_{G_q}$ when regarded as valued in $\mathrm{GL}_2(E)$ is ramified and non-split.

Proof. The first part is a well-known result of Carayol [Car94]. The second part is a consequence of the Mumford–Raynaud–Tate q-adic uniformization of the abelian variety A over $\mathbf{Q}_q$ which has multiplicative reduction at q and weight-monodromy results. Namely, consider the character group $\mathcal{X}(A, q)$ (with values in $\mathbf{Z}_p$) of the torus of the reduction of the Néron model of the abelian variety A at the prime q. We denote $\mathcal{X}(A, q)_{\mathcal{O}_v} = \mathcal{X}(A_q) \otimes \mathcal{O}_v$, where the tensor product is over $\mathrm{End}(A) \otimes \mathbf{Z}_p$. Note that as A has purely toric reduction at q, we have the exact sequence of $\mathcal{O}_v[G_q]$-modules

$$0 \to \mathrm{Hom}_{\mathcal{O}_v}((\mathcal{X}(A, q)_{\mathcal{O}_v}, \mathcal{O}_v(1)) \to \mathrm{Ta}_p(A)_{\mathcal{O}_v} \to \mathcal{X}(A^d, q)_{\mathcal{O}_v} \to 0.$$

Note that $\mathrm{Hom}_{\mathcal{O}_v}((\mathcal{X}(A, q)_{\mathcal{O}_v}, \mathcal{O}_v(1)), \mathcal{X}(A^d, q)_{\mathcal{O}_v}$ are both free $\mathcal{O}_v$-modules of rank 1 (as they are both torsion free and after inverting p are both isomorphic to F_v), where A^d denotes the

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

dual abelian variety of A. The fact that the representation is ramified follows from the q-adic uniformization. □

We will use this when $A = A_f$ with A_f the abelian variety associated by Shimura to a newform $f \in S_2(\Gamma_0(M))$ that is Steinberg at a place q. The shape of the Galois representation $\rho_f|_{G_q}$ helps to apply the computations of § 7.2 to Theorem 7.14.

6. Patching

Let R and R^{St} be the full (respectively Steinberg) unframed global deformation rings for $\bar{\rho}$, as defined in § 4. Note that the representations $\rho^{\mathrm{mod}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T})$ and $\rho^{Q,\mathrm{mod}} = \rho^{\mathrm{mod}} \otimes_{\mathbf{T}} \mathbf{T}^{\mathrm{St}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}^{\mathrm{St}})$ induce surjective maps $R \to \mathbf{T}$ and $R^{\mathrm{St}} \to \mathbf{T}^{\mathrm{St}}$, which are compatible with quotient maps $R \to R^{\mathrm{St}}$ and $\mathbf{T} \to \mathbf{T}^{\mathrm{St}}$.

As noted in § 4.3, we may (non-canonically) view R and R^{St} as quotients of $R^{\square}$ and $R^{\square,\mathrm{St}}$, respectively, and moreover these quotient maps fit into the following commutative diagram.

$$\begin{array}{ccccccc} R_{\mathrm{loc}} & \longrightarrow & R^{\square} & \twoheadrightarrow & R & \twoheadrightarrow & \mathbf{T} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ R_{\mathrm{loc}}^{\mathrm{St}} & \longrightarrow & R^{\square,\mathrm{St}} & \twoheadrightarrow & R^{\mathrm{St}} & \twoheadrightarrow & \mathbf{T}^{\mathrm{St}} \end{array}$$

Using these maps, we may view $M(NQ^2)$ and $M^{\mathrm{St}}(N)$ as R_{loc}- and $R_{\mathrm{loc}}^{\mathrm{St}}$-modules, respectively.

The following results are fairly standard; we only sketch the proofs.

Lemma 6.1. *Treating $M(NQ^2)$ as an R_{loc}-module, $\mathrm{Spec}(R_{\mathrm{loc}} \otimes_{\mathcal{O}} E)$ is formally smooth at each point in the support of $M(NQ^2) \otimes_{\mathcal{O}} E$. In particular, each point in the support of $M(NQ^2) \otimes_{\mathcal{O}} E$ lies in exactly one irreducible component of $\mathrm{Spec}\, R_{\mathrm{loc}}$.*

Proof. To see this, it is enough to observe that for a classical newform g, the associated Galois representation $\rho_{g,\iota}$ for $\iota : \bar{\mathbf{Q}} \to \bar{\mathbf{Q}}_p$ is generic (see [All16, Lemma 1.1.5]) at places $q \in Q$. This is a standard consequence of the weight-monodromy conjectures that are known in the case of Galois representations attached to classical newforms (and which yield that $H^2(G_q, \mathrm{ad}(\rho_{g,\iota})) = 0$). In the special case of forms g of weight 2 that are Steinberg at a prime q, the genericity of $\rho_{g,\iota}|_{G_q}$ follows from the fact that the associated GL_2-type abelian variety A_g has multiplicative reduction at q (see Lemma 5.5(2)). This shows, for instance see [All16, Proposition 1.2.2], that $\rho_{g,\iota}$ corresponds to a smooth point of $\mathrm{Spec}\, R_q^{\square}$. □

Lemma 6.2. *Each irreducible component of $\mathrm{Spec}\, R_{\mathrm{loc}}$ contains at least one point in the support of $M(NQ^2) \otimes_{\mathcal{O}} E$.*

Proof. The irreducible components of $\mathrm{Spec}\, R_{\mathrm{loc}}$ are given by Proposition 4.5, combined with the description of the irreducible components of $\mathrm{Spec}\, R_q^{\square}$ given in [Sho16, § 5]. Thus, the components of $\mathrm{Spec}\, R_q^{\square}$ for $q \in Q$ are labelled by inertial Weil–Deligne parameters compatible with the residual representation, with the exception that when q is congruent to -1 mod p and $\bar{\rho}$ is unramified at q, there are two Steinberg components $R_q^{\mathrm{St}(+1)}$ and $R_q^{\mathrm{St}(-1)}$, the related inertial Weil–Deligne parameter being $(\tau, N) = (0, N)$ with N a non-zero nilpotent matrix in $M_2(E)$, corresponding to the two eigenvalues ± 1 of $\bar{\rho}(\mathrm{Frob}_q)$.

We claim that by using the analysis in [DT94b] (see [DT94b, Theorem 1]) of the local behavior of lifts of $\bar{\rho}$ (at all primes away from p) arising from newforms, one sees that the support of $M(NQ^2) \otimes_{\mathcal{O}} E$ contains a point in each of the components of $\mathrm{Spec}\, R_{\mathrm{loc}}$. To justify

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

this claim, in addition to the work of [DT94b], we have to take care of the additional components that arise in the case of q congruent to -1 mod p. For this, we need a slight refinement of [DT94b, Theorem 3]. Namely, we can find a newform g of weight 2 and level NQ that gives rise to $\bar{\rho}$, that is in particular Steinberg at all places in Q, as in [DT94b], with the additional flexibility that for each $q \in Q$ that is -1 mod p and unramified in $\bar{\rho}$, we can further specify the eigenvalue of U_q acting g to be either 1 or -1. To prove this refinement, using the level-raising methods of [Rib84] and [DT94b], we use the fact that if h is a q-old eigenform in $S_2(\Gamma_0(NQ/q), \mathcal{O})$ that gives rise to $\bar{\rho}$, then it has two stabilizations of level NQ that are eigenforms for U_q with eigenvalues that are respectively 1 and -1 modulo the maximal ideal of $\mathcal{O}$, and the reductions of both these forms are in the kernel of $U_q^2 - 1$. □

The Taylor–Wiles–Kisin method applied in our situation now gives the following result.

THEOREM 6.3. *There exist integers $g, d \geq 0$, rings*

$$R_\infty = R_{\mathrm{loc}}[[x_1, \ldots, x_g]],$$
$$R_\infty^{\mathrm{St}} = R_{\mathrm{loc}}^{\mathrm{St}}[[x_1, \ldots, x_g]],$$
$$S_\infty = \mathcal{O}[[y_1, \ldots, y_d]]$$

and modules M_∞ and M_∞^{St} over R_∞ and R_∞^{St}, respectively, satisfying the following.

1. $\dim S_\infty = \dim R_\infty = \dim R_\infty^{\mathrm{St}}$.
2. *There exists an $\mathcal{O}$-module morphism $i : S_\infty \to R_\infty$ (inducing a morphism $i^{\mathrm{St}} : S_\infty \to R_\infty^{\mathrm{St}}$ via the quotient map $R_\infty \to R_\infty^{\mathrm{St}}$) making M_∞ and M_∞^{St} into finite free S_∞-modules. From now on, we will view R_∞ and R_∞^{St} as S_∞-algebras via this map i.*
3. *M_∞ and M_∞^{St} are maximal Cohen–Macaulay modules over R_∞ and R_∞^{St}, respectively.*
4. *R_∞ and R_∞^{St} act faithfully on M_∞ and M_∞^{St}, respectively.*
5. *We have isomorphisms $R_\infty \otimes_{S_\infty} \mathcal{O} \cong R$, $R_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O} \cong R^{\mathrm{St}}$, $M_\infty \otimes_{S_\infty} \mathcal{O} \cong M(NQ^2)$ and $M_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O} \cong M^{\mathrm{St}}(N)$.*
6. *The maps $i : S_\infty \to R_\infty$ and $i^{\mathrm{St}} : S_\infty \to R_\infty^{\mathrm{St}}$ are injective, and these maps make R_∞ and R_∞^{St} into finite free S_∞-modules.*
7. *The surjective maps $R \to \mathbf{T}$ and $R^{\mathrm{St}} \to \mathbf{T}^{\mathrm{St}}$ are isomorphisms. In particular, R and R^{St} are finite over $\mathcal{O}$, reduced and ϖ-torsion free.*

Proof. As explained in § 2, the assumptions on $\bar{\rho}$ (namely that it is absolutely irreducible and that $N(\bar{\rho})$ is squarefree) imply that $\bar{\rho}|_{G_{\mathbf{Q}(\zeta_p)}} : G_{\mathbf{Q}(\zeta_p)} \to \mathrm{GL}_2(k)$, that is, that $\bar{\rho}$ satisfies the 'Taylor–Wiles condition'.

First, as $\bar{\rho}|_{\mathbf{Q}(\zeta_p)}$ is irreducible, we may appeal to Lemma 2 of [DT94a] to find an arbitrarily large prime r so that no lift of $\bar{\rho}$ is ramified at r. This allows us to impose $\Gamma_1(r^2)$ level structure without affecting any of the objects considered in this theorem, and so we may ignore any issues arising from isotropy (see [Man21, § 4.2] or [EGS15, § 6.2] for details).

We can now apply the Taylor–Wiles–Kisin patching method applied to the modules $M(NQ^2)$ and $M^{\mathrm{St}}(N)$. We will primarily follow the treatment in [Man21, § 4].[4]

First, by the method outlined in [Man21, § 4.3], there exist:

- integers $g, d \geq 0$ satisfying $d + 1 = \dim R_{\mathrm{loc}} + g = \dim R_{\mathrm{loc}}^{\mathrm{St}} + g$ (see [Man21, Lemma 2.5] and [Kis09, Proposition (3.2.5)]);

[4] One may note that [Man21] only handles the 'minimal level' case, i.e. only $M^{\mathrm{St}}(N)$ and not $M(NQ^2)$ in our notation; however, all results we will cite can easily be seen to hold in our situation with no modifications to the proofs.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

- for each $n \geq 1$, an S_∞-algebra $R_n^\square$ with a surjective map $R_\infty \twoheadrightarrow R_n$ and an isomorphism $R_n^\square \otimes_{S_\infty} \mathcal{O} \cong R$, where S_∞ and R_∞ are as in the theorem statement, and the kernel of the structure map $S_\infty \to R_n^\square$ has a specific form (as given in [Man21, Proposition 4.7]);
- for each $n \geq 1$, a finitely generated $R_n^\square$-module $M_n^\square$, which is free over $\mathrm{im}(S_\infty \to R_n^\square)$ and satisfies $M_n^\square \otimes_{S_\infty} \mathcal{O} \cong M(NQ^2)$.

Briefly, $R_n^\square$ is a (framed) global deformation ring constructed exactly as in § 4.3 of this paper, except that we also allow deformations with ramification at a carefully chosen set of primes Q_n (constructed in [Kis09, Proposition (3.2.5)]), disjoint from S. The module $M_n^\square$ is simply constructed from the homology group of an appropriate modular curve just as in [Man21, § 4.3] (although to keep this construction consistent with the construction of $M(NQ^2)$ given in this paper, we must also modify this space by localizing at the maximal ideal $\mathfrak{m}_Q$ of the full Hecke algebra, as in § 5.1 of this paper: this does not cause any issue for the arguments of [Man21]).

The 'ultrapatching' construction described in [Man21, § 4.1] (as well as in the proof of Lemma 4.8) then produces an S_∞-algebra $\widetilde{R}_\infty$ (which would be called $\mathscr{P}(\mathscr{R}^\square)$ in the notation of that paper) as well as an $\widetilde{R}_\infty$-module M_∞, for which:

- M_∞ is finite free over S_∞;
- $\widetilde{R}_\infty \otimes_{S_\infty} \mathcal{O} \cong R$ and $M_\infty \otimes_{S_\infty} \mathcal{O} \cong M(NQ^2)$;
- there is a surjection $R_\infty \twoheadrightarrow \widetilde{R}_\infty$ such that the composition

$$R_{\mathrm{loc}} \hookrightarrow R_\infty \twoheadrightarrow \widetilde{R}_\infty \twoheadrightarrow R$$

  is the map $R_{\mathrm{loc}} \to R$ from the start of this section.

Analogously, for each $n \geq 1$, we may also construct a quotient $R_n^{\square,\mathrm{St}}$ of $R_n^\square$ and an $R_n^{\square,\mathrm{St}}$-module $M_n^{\square,\mathrm{St}}$ satisfying analogous properties to $R_n^\square$ and $M_n^\square$ (this time identically to the construction given in [Man21]). Again applying the ultrapatching construction produces an S_∞-algebra $\widetilde{R}_\infty^{\mathrm{St}}$ and a $\widetilde{R}_\infty^{\square,\mathrm{St}}$-module M_∞^{St}, which is finite free over S_∞, and satisfies $\widetilde{R}_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O} \cong R^{\mathrm{St}}$ and $M_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O} \cong M^{\mathrm{St}}(N)$.

Moreover, as each $R_n^{\square,\mathrm{St}}$ is a quotient of $R_n^\square$, the natural functoriality of the ultrapatching construction (see [Man21, § 4.1]) gives a quotient map $\widetilde{R}_\infty \twoheadrightarrow \widetilde{R}_\infty^{\mathrm{St}}$ and one can easily check that the maps we have described fit into the commutative diagram

$$\begin{array}{ccccccc}
R_{\mathrm{loc}} & \hookrightarrow & R_\infty & \twoheadrightarrow & \widetilde{R}_\infty & \twoheadrightarrow & R \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
R_{\mathrm{loc}}^{\mathrm{St}} & \hookrightarrow & R_\infty^{\mathrm{St}} & \twoheadrightarrow & \widetilde{R}_\infty^{\mathrm{St}} & \twoheadrightarrow & R^{\mathrm{St}}
\end{array}$$

where the compositions of the top and bottom rows are the maps $R_{\mathrm{loc}} \to R$ and $R_{\mathrm{loc}}^{\mathrm{St}} \to R^{\mathrm{St}}$ from before.

Now, as $\widetilde{R}_\infty$ is an S_∞-algebra, R_∞ is a complete local ring and $S_\infty = \mathcal{O}[[y_1, \ldots, y_d]]$ is a power series ring, we may lift the structure map $S_\infty \to \widetilde{R}_\infty$ to a map $i : S_\infty \to R_\infty$ making $\pi_\infty : R_\infty \to \widetilde{R}_\infty$ into an S_∞-module surjection. This proves (1) and (2).

(3) follows by noting that $(\varpi, i(y_1), \ldots, i(y_d))$ and $(\varpi, i^{\mathrm{St}}(y_1), \ldots, i^{\mathrm{St}}(y_d))$ are regular sequences for M_∞ and M_∞^{St}, respectively, as M_∞ and M_∞^{St} are both free over S_∞.

But now by standard properties of maximal Cohen–Macaulay modules, the supports of M_∞ and M_∞^{St} are unions of irreducible components of $\mathrm{Spec}\, R_\infty$ and $\mathrm{Spec}\, R_\infty^{\mathrm{St}}$, respectively. As R_∞^{St} is a domain, this implies that it acts faithfully on M_∞^{St}.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Now, as $R_\infty = R_{\mathrm{loc}}[[x_1, \dots, x_g]]$, the irreducible components of $\operatorname{Spec} R_\infty$ are in bijection with those of $\operatorname{Spec} R_{\mathrm{loc}}$. By Lemma 6.2, it follows that each irreducible component of $\operatorname{Spec} R_\infty$ contains a point in the support of $M_\infty/(i(y_1), \dots, i(y_d)) \otimes_{\mathcal{O}} E = M(NQ^2) \otimes_{\mathcal{O}} E$ (which is not contained in any other component, by Lemma 6.1). Since the support of M_∞ clearly contains the support of $M_\infty/(i(y_1), \dots, i(y_d)) \otimes_{\mathcal{O}} E$, it follows that M_∞ is supported on all of $\operatorname{Spec} R_\infty$. Since R_{loc}, and hence R_∞, is reduced (by Proposition 4.5), it follows that R_∞ acts faithfully on M_∞.

Now, by definition, the action of R_∞ on M_∞ factors through $\pi_\infty : R_\infty \to \widetilde{R}_\infty$. Since R_∞ acts faithfully on M_∞, it follows that π_∞ must be an isomorphism of S_∞-algebras. By an identical argument, $\pi_\infty^{\mathrm{St}} : R_\infty^{\mathrm{St}} \to \widetilde{R}_\infty^{\mathrm{St}}$ is also an isomorphism of S_∞-algebras. (5) now follows immediately from the properties of $\widetilde{R}_\infty$ and $\widetilde{R}_\infty^{\mathrm{St}}$ mentioned above.

Now, as M_∞ is supported on every geometric point of $\operatorname{Spec} R_\infty$, $M(NQ^2) = M_\infty \otimes_{S_\infty} \mathcal{O}$ is supported on every geometric point of $\operatorname{Spec}(R_\infty \otimes_{S_\infty} \mathcal{O}) = \operatorname{Spec} R$ and so, as the action of R on $M(NQ^2)$ factors through the surjection $R \to \mathbf{T}$, we must have $R^{\mathrm{red}} = \mathbf{T}$ (as $\mathbf{T}$ is reduced). In particular, as $\mathbf{T}$ is finite over $\mathcal{O}$, R is finite over $\mathcal{O}$ as well. Similarly, R^{St} is finite over $\mathcal{O}$.

In particular, $R/\varpi = R_\infty/(\varpi, i(y_1), \dots, i(y_d))$ and $R^{\mathrm{St}}/\varpi = R_\infty^{\mathrm{St}}/(\varpi, i^{\mathrm{St}}(y_1), \dots, i^{\mathrm{St}}(y_d))$ are both finite and hence are zero dimensional. It follows that $(\varpi, i(y_1), \dots, i(y_d))$ and $(\varpi, i^{\mathrm{St}}(y_1), \dots, i^{\mathrm{St}}(y_d))$ are systems of parameters for R_∞ and R_∞^{St}, respectively. As both of these rings are Cohen–Macaulay by Proposition 4.5, it follows that these are regular sequences. Equivalently, treating R_∞ and R_∞^{St} as S_∞-modules, these both have $(\varpi, y_1, \dots, y_d)$ as a regular sequence and so are maximal Cohen–Macaulay over S_∞. But all finitely generated maximal Cohen–Macaulay modules over S_∞ are free (by the Auslander–Buchsbaum formula), so (6) follows.

In particular, $R = R_\infty \otimes_{S_\infty} \mathcal{O}$ and $R^{\mathrm{St}} = R_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O}$ are finite flat over $\mathcal{O}$.

Now, similarly to [Kis09, Proposition (3.3.1)], take any $x \in \operatorname{Spec} R[1/\varpi] \subseteq \operatorname{Spec}(R_\infty \otimes_{\mathcal{O}} E)$. As noted above, x is in the support of $M(NQ^2)$ and so Lemma 6.1 gives that $\operatorname{Spec}(R_\infty \otimes_{\mathcal{O}} E)$ is formally smooth at x, which implies that $(M_\infty \otimes_{\mathcal{O}} E)_x$ is projective, and hence free, over $(R_\infty \otimes_{\mathcal{O}} E)_x$. It follows that $(M(NQ^2) \otimes_{\mathcal{O}} E)_x = (M_\infty \otimes_{\mathcal{O}} E)_x/(i(y_1), \dots, i(y_d))$ is free (of rank 1) over $R[1/\varpi]_x$. Thus, $M(NQ^2) \otimes_{\mathcal{O}} E$ is locally free of rank 1 over $R[1/\varpi]$. As $R[1/\varpi]$ has dimension zero, and is thus a direct product of local E-algebras, it follows that $M(NQ^2) \otimes_{\mathcal{O}} E$ is free of rank 1 over $R[1/\varpi]$.

Since the action of $R[1/\varpi]$ on $M(NQ^2) \otimes_{\mathcal{O}} E$ factors through $R[1/\varpi] \twoheadrightarrow \mathbf{T}[1/\varpi]$, it follows that the map $R[1/\varpi] \twoheadrightarrow \mathbf{T}[1/\varpi]$ is an isomorphism and so the kernel of the surjection $R \twoheadrightarrow \mathbf{T}$ is a torsion $\mathcal{O}$-module. As noted above, R is free over $\mathcal{O}$ and thus has no nontrivial torsion submodules. It follows that the map $R \to \mathbf{T}$ is an isomorphism.[5] Similarly, $R^{\mathrm{St}} \to \mathbf{T}^{\mathrm{St}}$ is an isomorphism. This proves (7). □

7. Patching and the growth of cotangent spaces

7.1 Continuous Kähler differentials

The module $\Omega_{B/A}$ of Kähler differentials of B over A (for a ring map, $A \to B$) can be a rather poorly behaved object when B is merely *topologically* finitely generated over A rather than literally finitely generated. As the local deformation rings we will be considering in our arguments are merely topologically finitely generated, it will often be necessary for us to consider the module of *continuous* Kähler differentials $\widehat{\Omega}_{B/A}$ in place of $\Omega_{B/A}$. In this section we review the definition and basic properties of continuous Kähler differentials.

[5] See also the arguments in [Sno18, § 5] or [Man21, § 4.3] for the observation that R_∞ being Cohen–Macaulay implies an integral $R = \mathbf{T}$ theorem.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

For the reminder of this section, for any $A \in \mathrm{CNL}_{\mathcal{O}}$ we will always use $\mathfrak{m}_A$ to denote the unique maximal ideal of A. For any $A, B \in \mathrm{CNL}_{\mathcal{O}}$, we will say that a homomorphism $f : A \to B$ is local if $\mathfrak{m}_A = f^{-1}(\mathfrak{m}_B)$ (or equivalently $f(\mathfrak{m}_A) \subseteq \mathfrak{m}_B$).

For any $A, B \in \mathrm{CNL}_{\mathcal{O}}$ and any local homomorphism $A \to B$, we will define

$$\widehat{\Omega}_{B/A} = \varprojlim_n \Omega_{(B/\mathfrak{m}_B^n)/(A/\mathfrak{m}_A^n)} = \varprojlim_n \Omega_{(B/\mathfrak{m}_B^n)/A}$$

(where the equality $\Omega_{(B/\mathfrak{m}_B^n)/(A/\mathfrak{m}_A^n)} = \Omega_{(B/\mathfrak{m}_B^n)/A}$ follows from [Sta19, Lemma 00RR]). Note that if B is topologically generated as an A-algebra by $b_1, \ldots, b_N \in B$, then each $\Omega_{(B/\mathfrak{m}_B^n)/A}$, and thus $\widehat{\Omega}_{B/A}$ itself, is generated as a B-module by $db_1, \ldots, db_N$.

First we observe the following result.

Lemma 7.1. *Let $A \in \mathrm{CNL}_{\mathcal{O}}$ and let $\mathcal{R}$ be a finitely generated A-algebra. Let $f : A \to \mathcal{R}$ be the structure map, let $\mathfrak{m}_{\mathcal{R}} \subseteq \mathcal{R}$ be a maximal ideal for which $f^{-1}(\mathfrak{m}_{\mathcal{R}}) = \mathfrak{m}_A$ and define $R = \mathcal{R}_{\mathfrak{m}_{\mathcal{R}}} \in \mathrm{CNL}_{\mathcal{O}}$. Then*

$$\widehat{\Omega}_{R/A} = \Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} R.$$

Proof. By [Sta19, Lemma 02HQ], for any k and any $n > k$ we have a canonical isomorphism

$$\Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k \xrightarrow{\sim} \Omega_{(\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n)/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k$$

and so taking inverse limits we have

$$\Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k \xrightarrow{\sim} \varprojlim_n \left(\Omega_{(\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n)/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k\right).$$

Now, for each n, $\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n$ is certainly a finitely generated A-algebra and so $\Omega_{(\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n)/A}$ is a finitely generated $\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n$-module. In particular, it has finite length. As $\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k$ is a finitely presented $\mathcal{R}$-module, the above work, and the Mittag–Leffler conditions (see [Sta19, Section 0594]), imply that

$$\begin{aligned}\left(\Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} R\right) \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k \cong \Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k &\cong \varprojlim_n \left(\Omega_{(\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n)/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k\right) \\ &\cong \varprojlim_n \left(\Omega_{(\mathcal{R}/\mathfrak{m}_{\mathcal{R}}^n)/A}\right) \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k \cong \widehat{\Omega}_{R/A} \otimes_{\mathcal{R}} \mathcal{R}/\mathfrak{m}_{\mathcal{R}}^k\end{aligned}$$

for all k. Since $\Omega_{\mathcal{R}/A} \otimes_{\mathcal{R}} R$ and $\widehat{\Omega}_{R/A}$ are both finitely generated R-modules, the claim follows. □

Lemma 7.2. *If $A, B \in \mathrm{CNL}_{\mathcal{O}}$ are such that B is an A-algebra which is finitely generated as an A-module, then $\widehat{\Omega}_{B/A} \cong \Omega_{B/A}$.*

Proof. B is already finitely generated as an A-algebra, so we may simply take $R = \mathcal{R} = B$ in Lemma 7.1. □

Lemma 7.3. *Take any $R \in \mathrm{CNL}_{\mathcal{O}}$ and any augmentation $\lambda : R \to \mathcal{O}$. Defining $\Phi_{\lambda,R} := (\ker \lambda)/(\ker \lambda)^2$ as in § 3, we have $\Phi_{\lambda,R} \cong \widehat{\Omega}_{R/\mathcal{O}} \otimes_{\lambda} \mathcal{O}$.*

In particular, in the setup of Lemma 7.1 with $A = \mathcal{O}$, if $\lambda_{\mathcal{R}} : \mathcal{R} \to \mathcal{O}$ is an augmentation extending to an augmentation $\lambda : R \to \mathcal{O}$, then we have $\Phi_{\lambda,R} \cong \Omega_{\mathcal{R}/A} \otimes_{\lambda_{\mathcal{R}}} \mathcal{O} \cong (\ker \lambda_{\mathcal{R}})/(\ker \lambda_{\mathcal{R}})^2$.

Proof. Let $I = \ker \lambda$, so that $R/I \cong \mathcal{O}$. As in the proof of Lemma 7.1, we get that

$$\widehat{\Omega}_{R/A} \otimes_{\lambda} \mathcal{O} = \widehat{\Omega}_{R/A} \otimes_R R/I = \varprojlim_n \left(\Omega_{(R/\mathfrak{m}_R^n)/\mathcal{O}}\right) \otimes_R R/I = \varprojlim_n \left(\Omega_{(R/\mathfrak{m}_R^n)/\mathcal{O}} \otimes_R R/I\right).$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Now, for each n, $\lambda : R \to \mathcal{O}$ gives a surjection $\bar{\lambda}_n : R/\mathfrak{m}_R^n \to \mathcal{O}/\varpi^n$. Let $\bar{I}_n = \ker \bar{\lambda}_n = (I + \mathfrak{m}_R^n)/\mathfrak{m}_R^n \cong I/(I \cap \mathfrak{m}_R^n)$. Now note that the structure map $\mathcal{O} \to \mathcal{O}/\varpi^n \to R/\mathfrak{m}_R^n$ provides a section to $\bar{\lambda}_n$, so [Sta19, Lemma 02HP] (as well as the fact that $\Omega_{(\mathcal{O}/\varpi^n)/\mathcal{O}} = 0$) gives a natural isomorphism

$$\Omega_{(R/\mathfrak{m}_R^n)/\mathcal{O}} \otimes_R R/I \cong \Omega_{(R/\mathfrak{m}_R^n)/\mathcal{O}} \otimes_{R/\mathfrak{m}_R^n} (R/\mathfrak{m}_R^n)/\bar{I}_n \cong \bar{I}_n/\bar{I}_n^2.$$

Taking inverse limits (and again applying the Mittag–Leffler conditions to $0 \to \bar{I}_n^2 \to \bar{I}_n \to \bar{I}_n/\bar{I}_n^2 \to 0$) gives the isomorphism $\widehat{\Omega}_{R/A} \otimes_\lambda \mathcal{O} \cong I/I^2 = \Phi_{\lambda,R}$.

The second claim follows by noting that $\Omega_{\mathcal{R}/A} \otimes_{\lambda_\mathcal{R}} \mathcal{O} \cong (\ker \lambda_\mathcal{R})/(\ker \lambda_\mathcal{R})^2$ (by a similar, but simpler, argument to the one above) and $(\ker \lambda)/(\ker \lambda)^2 \cong (\ker \lambda_\mathcal{R})/(\ker \lambda_\mathcal{R})^2$. □

LEMMA 7.4. *Given $A, B, C \in \mathrm{CNL}_\mathcal{O}$ and local homomorphisms $A \to B$ and $A \to C$ making B and C into A-algebras, we have an isomorphism*

$$\widehat{\Omega}_{(B\widehat{\otimes}_A C)/A} \cong \left(\widehat{\Omega}_{B/A}\widehat{\otimes}_A C\right) \oplus \left(B\widehat{\otimes}_A \widehat{\Omega}_{C/A}\right).$$

Proof. Let $R = B\widehat{\otimes}_A C$. For any $n \geq 1$, define the ideal $I_n \subseteq R$ by

$$I_n = \mathfrak{m}_B^n R + \mathfrak{m}_C^n R = \ker\left(R \twoheadrightarrow (B/\mathfrak{m}_B^n) \otimes_A (C/\mathfrak{m}_C^n)\right).$$

As $\{I_n\}$ is cofinal with $\{\mathfrak{m}_R^n\}$, we get that

$$\widehat{\Omega}_{R/A} \cong \varprojlim_n \Omega_{(R/I_n)/A} \cong \varprojlim_n \Omega_{[(B/\mathfrak{m}_B^n)\otimes_A(C/\mathfrak{m}_C^n)]/A}.$$

But now, by [Eis95, Proposition 16.5],

$$\Omega_{[(B/\mathfrak{m}_B^n)\otimes_A(C/\mathfrak{m}_C^n)]/A} \cong \left(\widehat{\Omega}_{(B/\mathfrak{m}_B^n)/A} \otimes_A (C/\mathfrak{m}_C^n)\right) \oplus \left((B/\mathfrak{m}_B^n) \otimes_A \widehat{\Omega}_{(C/\mathfrak{m}_C^n)/A}\right),$$

so taking inverse limits gives the result. □

Lastly, a standard property of Kähler differentials (cf. [Sta19, Lemma 00RS]) states that for any ring maps $A \to B \to C$ we have an exact sequence

$$\Omega_{B/A} \otimes_B C \to \Omega_{C/A} \to \Omega_{C/B} \to 0.$$

We will need the following continuous version of this below.

LEMMA 7.5. *Let $A \to B \to C$ be local ring homomorphisms for $A, B, C \in \mathrm{CNL}_\mathcal{O}$. Then there is an exact sequence*

$$\widehat{\Omega}_{B/A} \otimes_B C \to \widehat{\Omega}_{C/A} \to \widehat{\Omega}_{C/B} \to 0.$$

Proof. Since the maps $A \to B \to C$ are local ring homomorphisms, they clearly induce homomorphisms $A \to B/\mathfrak{m}_B^n \to C/\mathfrak{m}_C^k$ for all $n > k$. Thus, we have an exact sequence

$$\Omega_{(B/\mathfrak{m}_B^n)/A} \otimes_B C/\mathfrak{m}_C^k \to \Omega_{(C/\mathfrak{m}_C^k)/A} \to \Omega_{(C/\mathfrak{m}_C^k)/B} \to 0$$

(where we used the fact that $\Omega_{(C/\mathfrak{m}_C^k)/(B/\mathfrak{m}_B^n)} = \Omega_{(C/\mathfrak{m}_C^k)/B}$).

As $C/\mathfrak{m}_C^k$ is a finite group, it is certainly finitely presented as a B-module. Thus, as each $\Omega_{(B/\mathfrak{m}_B^n)/A}$ has finite length, the Mittag–Leffler conditions then imply that

$$\varprojlim_n \left(\Omega_{(B/\mathfrak{m}_B^n)/A} \otimes_B C/\mathfrak{m}_C^k\right) \cong \varprojlim_n \left(\Omega_{(B/\mathfrak{m}_B^n)/A}\right) \otimes_B C/\mathfrak{m}_C^k \cong \widehat{\Omega}_{B/A} \otimes_B C/\mathfrak{m}_C^k$$

and so we have an exact sequence

$$\widehat{\Omega}_{B/A} \otimes_B C/\mathfrak{m}_C^k \to \Omega_{(C/\mathfrak{m}_C^k)/A} \to \Omega_{(C/\mathfrak{m}_C^k)/B} \to 0$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

for each k. As $\widehat{\Omega}_{B/A}$ is a finitely generated B-module, each module in the above sequence again has finite length, so, taking the inverse limit with respect to k and using the Mittag–Leffler conditions again, give the desired exact sequence. □

7.2 Relative local cotangent spaces at an augmentation

We consider primes $q \in Q$; recall that these are Steinberg for $\bar{\rho}$. We begin with a general definition that we need.

Definition 7.6. Given $R, S \in \mathrm{CNL}_{\mathcal{O}}$, a surjective map $\alpha : R \to S$ in $\mathrm{CNL}_{\mathcal{O}}$ and an augmentation $\lambda : S \to \mathcal{O}$ whose pull back to R we denote by λ_R, we define the *relative cotangent space* $\Phi_{\lambda,R/S}$ to be the kernel of the surjective map $\Phi_{\lambda_R,R} \to \Phi_{\lambda,S}$, i.e. the kernel of the map $\widehat{\Omega}_{R/\mathcal{O}} \otimes_{\lambda_R} \mathcal{O} \to \widehat{\Omega}_{S/\mathcal{O}} \otimes_{\lambda_S} \mathcal{O}$.

Definition 7.7. We fix an augmentation $\lambda : R_q^{\mathrm{St}} \to \mathcal{O}$ such that the corresponding representation of $G_q \to \mathrm{GL}_2(\mathcal{O})$ is of the form

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$$

up to twist by an unramified character of order dividing 2, and which is ramified (i.e. generic). We define the local invariants m_q (respectively, n_q) to be the largest integer n such that $\rho_\lambda(I_q)$ (respectively, $\rho_\lambda(G_q)$) mod ϖ^n has trivial projective image.

Lemma 7.8. *We have the inequalities $n_q \leq m_q$ and $n_q \leq \mathrm{ord}_\varpi(q-1)$.*

Proof. The first inequality follows from $I_q \leq G_q$. The second inequality follows from the fact that the characteristic polynomial of $\rho_\lambda(\mathrm{Frob}_q) = x^2 \pm (q+1)x + q$ (where Frob_q is any lift of the Frobenius automorphism of the residue field of q). □

We have the following key proposition.

Proposition 7.9. *Consider a prime $q \in Q$. We fix an augmentation $\lambda : R_q^{\mathrm{St}} \to \mathcal{O}$ such that the corresponding representation of $G_q \to \mathrm{GL}_2(\mathcal{O})$ is of the form*

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$$

up to twist by an unramified character of order dividing 2, and which is ramified (i.e. generic). Then the following hold for the relative cotangent space at λ, where cases 2 and 3 only occur if $\bar{\rho}$ is unramified.

1. *For the morphism $R_q^{\square} \to R_q^{\mathrm{St}}$, the module $\Phi_{\lambda,R_q^{\square}/R_q^{\mathrm{St}}}$ is finite over $\mathcal{O}$ of length $m_q + \mathrm{ord}_{\mathcal{O}}(q^2-1) - 2n_q$.*
2. *For the morphism $R_q^{\square} \to R_q^{\mathrm{uni}}$, the module $\Phi_{\lambda,R_q^{\square}/R_q^{\mathrm{uni}}}$ is finite cyclic over $\mathcal{O}$ of length $\mathrm{ord}_{\mathcal{O}}(q^2-1) - n_q$.*
3. *For the morphism $R_q^{\mathrm{uni}} \to R_q^{\mathrm{St}}$, the module $\Phi_{\lambda,R_q^{\mathrm{uni}}/R_q^{\mathrm{St}}}$ is finite cyclic over $\mathcal{O}$ of length $m_q - n_q$.*

Proof. By possibly twisting with the unramified character of order 2, we may assume that our representation is of the form

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}.$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

We first recall well-known explicit descriptions of the rings $R_q^{\square} \to R_q^{\mathrm{uni}} \to R_q^{\mathrm{St}}$ introduced in § 4.2. We have two limiting conditions: we require our deformations to have determinant the p-adic cyclotomic character. Moreover, the deformations of $\bar{\rho}$ factor via the tame quotient G_q^t of G_q, because the order of inertia under $\bar{\rho}$ divides $p \neq q$. The group G_q^t can be identified with the profinite completion of the group $\langle \sigma, \tau \mid \sigma\tau\sigma^{-1} = \tau^q \rangle$ in such a way that σ maps to Frobenius in the unramified quotient of G_q.

Imposing the conditions just stated on the images of σ and τ in a matrix group, and choosing the augmentation λ as the origin of our coordinates, leads to the following description of the universal deformation $\rho_q^{\square}\colon G_q^t \to \mathrm{GL}_2(R_q^{\square})$: let $\mathcal{R} := \mathcal{O}[[a, b, c, e, \alpha, \beta, \gamma, \delta]]$, let $A, B \in \mathrm{GL}_2(\mathcal{R})$ be the matrices

$$A := \begin{pmatrix} q+a & b+s \\ c & 1-a-e \end{pmatrix}, \quad B := \begin{pmatrix} 1+\alpha & \beta+t \\ \gamma & 1+\delta \end{pmatrix}$$

with $s, t \in \mathcal{O}$ and t non-zero and let $\mathcal{I}_q^{\square} = (r_1^{\square}, \dots, r_6^{\square}) \subset \mathcal{R}$ be the ideal generated by $r_1^{\square} = \det A - q = a - (q+a)(a+e) - (b+s)c$, $r_2^{\square} = \det B - 1 = \alpha + \delta + \alpha\delta - (\beta+t)\gamma$ and

$$\begin{pmatrix} r_3^{\square} & r_4^{\square} \\ r_5^{\square} & r_6^{\square} \end{pmatrix} = AB - B^q A.$$

Then $R_q^{\square} = \mathcal{R}/\mathcal{I}_q^{\square}$ and $\rho_q^{\square}$ is defined by $\sigma \mapsto A \pmod{\mathcal{I}_q^{\square}}$ and $\tau \mapsto B \pmod{\mathcal{I}_q^{\square}}$. It is not hard to see that $\mathcal{I}_q^{\square}$ is generated by $r_1^{\square}$ and $r_2^{\square}$ and three additional elements from $\{r_3^{\square}, r_4^{\square}, r_5^{\square}, r_6^{\square}\}$. The ring $R_q^{\square}$ is a complete intersection ring, as we shall recall from work of Shotton in Proposition 4.3.

The ring R_q^{St} is described as a quotient of $\mathcal{R}$ by the following ideal $\mathcal{I}_q^{\mathrm{St}}$ by using either [Sho16, proof of Proposition 5.8] or Lemma 7.11 below; we give further details in Remarks 7.12 and 7.13. Explicitly, $\mathcal{I}_q^{\mathrm{St}}$ is generated by the six 2×2 minors $r_1^{\mathrm{St}}, \dots, r_6^{\mathrm{St}}$ of

$$\begin{pmatrix} \alpha & \beta+t & (q-1+a) & b+s \\ \gamma & \delta & c & -a-e \end{pmatrix} \tag{7.1}$$

together with $r_7^{\mathrm{St}} = \alpha + \delta$ and $r_8^{\mathrm{St}} = e$. For explicit use below, let r_1^{St} and r_2^{St} denote 2×2 minors of the above matrix formed by the first two columns, and the last two columns, respectively. The ring-theoretic properties of R_q^{St} are given in Proposition 4.3.

To describe R_q^{uni}, let $\mathcal{I}^{\mathrm{unr}}$ denote the ideal of $\mathcal{R}$ containing $\mathcal{I}_q^{\square}$ so that $R_q^{\mathrm{unr}} = \mathcal{R}/\mathcal{I}_q^{\mathrm{unr}}$ describes unramified deformations of $\bar{\rho}$, i.e. $\mathcal{I}_q^{\mathrm{unr}} = (\alpha, \beta+t, \gamma, \delta, r_1^{\square})$. Then $R_q^{\mathrm{uni}} = \mathcal{R}/\mathcal{I}_q^{\mathrm{uni}}$, where $\mathcal{I}_q^{\mathrm{uni}} = \mathcal{I}_q^{\mathrm{St}} \cap \mathcal{I}_q^{\mathrm{unr}}$. Explicitly, one has $\mathcal{I}_q^{\mathrm{uni}} = (r_1^{\mathrm{uni}}, \dots, r_{10}^{\mathrm{uni}})$, where $r_1^{\mathrm{uni}} = r_1^{\square}$ and $r_2^{\mathrm{uni}} = r_2^{\square}$ are the determinant conditions from $\mathcal{I}_q^{\square}$, $r_i^{\mathrm{uni}} = r_i^{\mathrm{St}}$ for $i = 3, \dots, 7$ and $r_8^{\mathrm{uni}} = \alpha e$, $r_9^{\mathrm{uni}} = (\beta+t)e$, $r_{10}^{\mathrm{uni}} = \gamma e$. The exact sequence

$$0 \to \mathcal{R}/I_q^{\mathrm{uni}} \to R_q^{\mathrm{unr}} \times R_q^{\mathrm{St}} \to \mathcal{R}/(\mathcal{I}_q^{\mathrm{St}} + \mathcal{I}_q^{\mathrm{unr}}) \to 0$$

allows one to verify that $\mathcal{R}/I_q^{\mathrm{uni}}$ is Cohen–Macaulay (this also requires [Eis95, Exercise 18.13]), reduced, flat over $\mathcal{O}$ and of relative dimension 3 and that $\operatorname{Spec} \mathcal{R}/I_q^{\mathrm{uni}}$ is the reduced schematic closure in $\operatorname{Spec} R_q^{\square}$ of $\operatorname{Spec} R_q^{\mathrm{St}} \cup R_q^{\mathrm{unr}}$. This proves the claim $R_q^{\mathrm{uni}} = \mathcal{R}/\mathcal{I}_q^{\mathrm{uni}}$.

Note that our generating sets of the ideals $\mathcal{I}_q^?$ are not claimed to be minimal. But for the computations to come, this is not relevant.

To compute the relative cotangent spaces, observe first that $\widehat{\Omega}_{\mathcal{R}/\mathcal{O}} \otimes_{\mathcal{R}} \mathcal{O}$ is the free $\mathcal{O}$-module of rank 8 with basis da, db, dc, de, $d\alpha$, $d\beta$, $d\gamma$, $d\delta$. We identify this module with $\mathcal{O}^8$ via this basis and regard it as a space of row vectors. We shall write $dr|_\lambda$ for the evaluation of a differential dr, $r \in \mathcal{R}$, under λ. Evaluation under λ simply means that the coefficients in $\mathcal{R}$ of dr are evaluated

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

at the origin, i.e. we set all of $a, b, c, e, \alpha, \beta, \gamma, \delta$ to zero. For $? \in \{\square, \mathrm{uni}, \mathrm{St}\}$, we denote by $\Lambda^? \subset \widehat{\Omega}_{\mathcal{R}/\mathcal{O}} \otimes_{\mathcal{R}} \mathcal{O}$ the $\mathcal{O}$-submodule spanned by elements $dr_j|_\lambda$ such that the dr_j generate $\mathcal{I}_q^?$. Then $\widehat{\Omega}_{R_q^?/\mathcal{O}} \otimes_{\mathcal{R}} \mathcal{O}$ is the quotient of $\widehat{\Omega}_{\mathcal{R}/\mathcal{O}} \otimes_{\mathcal{R}} \mathcal{O}$ by $\Lambda^?$. We shall compute an $\mathcal{O}$-basis of the $\mathcal{O}$-submodules $\Lambda^?$ of $\widehat{\Omega}_{\mathcal{R}/\mathcal{O}} \otimes_{\mathcal{R}} \mathcal{O}$. The relative cotangent spaces are the $\mathcal{O}$-modules

$$\Phi_{\lambda, R_q^?/R_q^{??}} \cong \Lambda^{??}/\Lambda^?. \tag{7.2}$$

The quotient is described by the square matrix expressing the basis we find for $\Lambda^?$ in terms of the basis we find for $\Lambda^{??}$. Finding $\Phi_{\lambda, R_q^?/R_q^{??}}$ is thereby reduced to the elementary divisor theorem.

The computation of $\Lambda^\square$ proceeds as follows: let $N = B - I$, so that $N|_\lambda = \left(\begin{smallmatrix} 0 & t \\ 0 & 0 \end{smallmatrix}\right)$ and $(N|_\lambda)^2 = 0$. Then

$$dr_1^\square = (1-a-e)\,da - (q+a)(da+de) - (b+s)\,dc - c\,db \rightsquigarrow dr_1^\square|_\lambda = da - q\,da - q\,de - s\,dc,$$

$$dr_2^\square = d\alpha + d\delta + \alpha\,d\delta + \delta\,d\alpha - (\beta+t)\,d\gamma - \gamma\,d\beta \rightsquigarrow dr_2^\square|_\lambda = d\alpha + d\delta - t\,d\gamma,$$

$$d\begin{pmatrix} r_3^\square & r_4^\square \\ r_5^\square & r_6^\square \end{pmatrix} = d\Big(AN - \sum_{i=1}^{q} \binom{q}{i} N^i A\Big) = (dA)N - \sum_{i=1}^{q} \binom{q}{i} N^i\,dA + A(dN) - \sum_{i=1}^{q} \binom{q}{i} \sum_{j=1}^{i} N^{j-1}(dN)N^{i-j}A$$

$$\rightsquigarrow \begin{pmatrix} dr_3^\square|_\lambda & dr_4^\square|_\lambda \\ dr_5^\square|_\lambda & dr_6^\square|_\lambda \end{pmatrix} = dA\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} - \binom{q}{1}\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} dA + \left(\begin{pmatrix} q & s \\ 0 & 1 \end{pmatrix} - qI\right) dN$$
$$- dN\left(\binom{q}{1}\begin{pmatrix} q & s \\ 0 & 1 \end{pmatrix} - qI\right)$$
$$- \binom{q}{2}\left(dN\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} dN\right)\begin{pmatrix} q & s \\ 0 & 1 \end{pmatrix}$$
$$- \binom{q}{3}\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} dN \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}\begin{pmatrix} q & s \\ 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} -qt\,dc + s\,d\gamma - (q^2-q)\,d\alpha & t\,da + qt\,da + qt\,de + s\,d\delta - qs\,d\alpha \\ -(q-1)\,d\gamma - (q^2-q)\,d\gamma & t\,dc - (q-1)\,d\delta - qs\,d\gamma \end{pmatrix}$$
$$- t\binom{q}{2}\begin{pmatrix} q\,d\gamma & s\,d\gamma + d\alpha + d\delta \\ 0 & d\gamma \end{pmatrix} - t^2\binom{q}{3}\begin{pmatrix} 0 & d\gamma \\ 0 & 0 \end{pmatrix}.$$

This gives six row vectors in $\mathcal{O}^8$ that span $\Lambda^\square$ and that we write as a matrix. Gauß elimination shows that the rows of the following matrix form an $\mathcal{O}$-basis of $\Lambda^\square$,

$$\begin{pmatrix} q-1 & 0 & s & q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -t & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & q^2-1 & 0 \\ 0 & 0 & t & 0 & 0 & 0 & \zeta & 1-q \\ 2t & 0 & 0 & 0 & 0 & 0 & \xi & 2s \end{pmatrix},$$

where we abbreviate $\xi = -qs^2 - st - t^2\binom{q+1}{3}$ and $\zeta = -qs - t\binom{q}{2}$.

To compute a basis of Λ^{uni}, observe first that $dr_i^{\mathrm{uni}} = dr_i^\square$, for $i = 1, 2$, and that their evaluation under λ is given by the first two rows of the last matrix. A straightforward computation

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

gives $dr_i^{\mathrm{uni}}|_\lambda$, $i = 7, \dots, 10$, as

$$dr_7^{\mathrm{uni}}|_\lambda = d\alpha + d\delta, dr_8^{\mathrm{uni}}|_\lambda = (\alpha\, de + e\, d\alpha)|_\lambda = 0, dr_9^{\mathrm{uni}}|_\lambda = \cdots = t\, de, dr_{10}^{\mathrm{uni}}|_\lambda = \cdots = 0.$$

When computing $d(\dots)|_\lambda$ for the 2×2 minors of the matrix in (7.1), observe that the evaluation under λ of the second row is identically zero. One deduces that $d(\dots)|_\lambda$ of these 2×2 minors are the 2×2 minors of the matrix

$$\begin{pmatrix} 0 & t & q-1 & s \\ d\gamma & d\delta & dc & -da - de \end{pmatrix}.$$

The 'inner' four of the resulting six minors provide us with the missing generators for Λ^{uni}. Gauß elimination on the resulting 10×8 matrix shows that a basis of Λ^{uni} over $\mathcal{O}$ is given by the rows of

$$\begin{pmatrix} q-1 & 0 & s & q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & f & 0 \\ 0 & 0 & t & 0 & 0 & 0 & 0 & 1-q \\ t & 0 & 0 & 0 & 0 & 0 & 0 & s \end{pmatrix}$$

with $f = \gcd(s, t, q-1)$. We clearly have $\Lambda^{\mathrm{uni}} \supset \Lambda^{\square}$. Writing the basis of $\Lambda^{\square}$ in terms of the basis of Λ^{uni} gives the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{-t}{f} & \frac{q^2-1}{f} & \frac{\zeta}{f} & \frac{\xi}{f} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

From the last matrix and the isomorphism (7.2), we find that

$$\Phi_{\lambda, R^{\square}/R^{\mathrm{uni}}} \cong \mathcal{O}/\frac{q^2-1}{\gcd(s,t,q-1)}\mathcal{O} = \mathcal{O}/((q^2-1)\varpi^{-n_q}).$$

Finding a set of generators of Λ^{St} over $\mathcal{O}$ is analogous to the other two cases. In fact, out of the eight spanning vectors in $\mathcal{O}^8$ from the $dr_i^{\mathrm{St}}|_\lambda$, we already computed six and the other two are easy. When performing Gauß elimination, one has to distinguish two cases.

Case I: $(q-1)$ divides s. Then a basis of Λ^{St} over $\mathcal{O}$ is given by the rows of

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & f & 0 \\ 0 & 0 & t & 0 & 0 & 0 & 0 & 1-q \\ f & 0 & f\frac{s}{q-1} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{7.3}$$

where we note that $\gcd(t, q-1) = \gcd(s, t, q-1) = f$ in case I.

Case II: s divides $(q-1)$. Then a basis of Λ^{St} over $\mathcal{O}$ is given by the first three rows of the matrix (7.3) together with the rows of

$$\begin{pmatrix} f\frac{q-1}{s} & 0 & f & 0 & 0 & 0 & 0 & 0 \\ t & 0 & 0 & 0 & 0 & 0 & 0 & s \end{pmatrix},$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

where in case II we use $\gcd(s,t)=\gcd(s,t,q-1)=f$.

In both cases we have $\Lambda^{\mathrm{St}}\supset\Lambda^{\mathrm{uni}}$ and writing the basis of Λ^{uni} in terms of the basis of Λ^{St} gives the matrices

$$\text{Case I: } \begin{pmatrix} q & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{s}{1-q} \\ \frac{q-1}{f} & 0 & 0 & 0 & \frac{t}{f} \end{pmatrix}, \quad \text{Case II: } \begin{pmatrix} q & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \frac{s}{f} & 0 & 0 & \frac{t}{f} & 0 \\ 0 & 0 & 0 & \frac{1-q}{s} & 1 \end{pmatrix}.$$

In either case, we find that

$$\Phi_{\lambda,R^{\mathrm{uni}}/R^{\mathrm{St}}}\cong\mathcal{O}/\mathcal{O}\tfrac{t}{f}=\mathcal{O}/(\varpi^{m_q-n_q}).$$

From the two expressions for $\Phi_{\lambda,R^?/R^{??}}$, the claims in the proposition are immediate. We only gave the computations for $\bar\rho$ unramified, but similar arguments allow one also to deduce 1 without 2 and 3 by directly comparing $\Lambda^{\square}$ with Λ^{St}. □

Now let $R_\infty=R_{\mathrm{loc}}[[x_1,\dots,x_g]]$ and $R_\infty^{\mathrm{St}}=R_{\mathrm{loc}}^{\mathrm{St}}[[x_1,\dots,x_g]]$ be as in § 6 and let $\lambda:R_\infty^{\mathrm{St}}\to\mathcal{O}$ be an augmentation, which then induces augmentations $R_q^{\mathrm{St}}\to\mathcal{O}$ for each $q\in Q$.

Corollary 7.10. *The relative cotangent space $\Phi_{R_\infty/R_\infty^{\mathrm{St}}}$ at λ of the morphism $R_\infty\to R_\infty^{\mathrm{St}}$ is of length*

$$\sum_{q\in Q}(m_q+\mathrm{ord}_{\mathcal{O}}(q^2-1)-2n_q)$$

as an $\mathcal{O}$-module.

Proof. Write $T=[\widehat{\bigotimes}_{\ell|N}R_\ell^{\mathrm{min}}]\widehat{\otimes}R_p^{\mathrm{fl}}[[x_1,\dots,x_g]]$, so that $R_\infty=[\widehat{\bigotimes}_{q|Q}R_q]\widehat{\otimes}_{\mathcal{O}}T$ and $R_\infty^{\mathrm{St}}=[\widehat{\bigotimes}_{q|Q}R_q^{\mathrm{St}}]\widehat{\otimes}_{\mathcal{O}}T$. Lemma 7.4 now gives

$$\widehat{\Omega}_{R_\infty/\mathcal{O}}\cong\left[\bigoplus_{q|Q}R_\infty\widehat{\otimes}_{R_q}\widehat{\Omega}_{R_q/\mathcal{O}}\right]\oplus(R_\infty\widehat{\otimes}_T\widehat{\Omega}_{T/\mathcal{O}})$$

and

$$\widehat{\Omega}_{R_\infty^{\mathrm{St}}/\mathcal{O}}\cong\left[\bigoplus_{q|Q}R_\infty^{\mathrm{St}}\widehat{\otimes}_{R_q^{\mathrm{St}}}\widehat{\Omega}_{R_q^{\mathrm{St}}/\mathcal{O}}\right]\oplus(R_\infty^{\mathrm{St}}\widehat{\otimes}_T\widehat{\Omega}_{T/\mathcal{O}})$$

and so $\Phi_{\lambda,R_\infty}\cong\bigoplus_{q|Q}\Phi_{\lambda,R_q}\oplus\Phi_{\lambda,T}$ and $\Phi_{\lambda,R_\infty^{\mathrm{St}}}\cong\bigoplus_{q|Q}\Phi_{\lambda,R_q^{\mathrm{St}}}\oplus\Phi_{\lambda,T}$. But this implies that $\Phi_{\lambda,R_\infty/R_\infty^{\mathrm{St}}}=\bigoplus_{q|Q}\Phi_{\lambda,R_q/R_q^{\mathrm{St}}}$, so the claim follows by Proposition 7.9. □

7.3 Presentation of R_q^{St}

The following lemma gives a coordinate-free explanation of how to obtain the relations in [Sho16, Proposition 5.8] for R_q^{St} that were used in the above proof.

Lemma 7.11. *Let R be a ring of characteristic zero and let $q>1$ be an integer. Let A,B be 2×2 matrices over R. Then the following hold.*

1. *Suppose that A and B have characteristic polynomials $\chi_A=(T-q)(T-1)$ and $\chi_B=(T-1)^2$ in $R[T]$. Then $AB=B^qA$ if and only if $(B-I)(A-I)=0$.*
2. *For $A:=\left(\begin{smallmatrix}q+a & b\\ c & 1-e\end{smallmatrix}\right)$ and $B:=\left(\begin{smallmatrix}1+\alpha & \beta\\ \gamma & 1+\delta\end{smallmatrix}\right)$, the following are equivalent:*

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

(a) $\chi_A = (T-q)(T-1)$, $\chi_B = (T-1)^2$ *and* $(B-I)(A-I) = 0$;
(b) $e = a$, $\delta = -\alpha$ *and all* 2×2 *minors of the following matrix vanish:*

$$\begin{pmatrix} \alpha & \beta & (q-1+a) & b \\ \gamma & -\alpha & c & -a \end{pmatrix}. \tag{7.4}$$

Proof. (a) Let $N := B - I$. The conditions on χ_A and χ_B are equivalent to $\operatorname{tr} A = q+1$, $\det A = q$, $\operatorname{tr} N = 0$ and $\det N = 0$. The latter also implies that $N^2 = 0$, so that $B^q A = (I+N)^q A = (I+qN)A$ and $AB = A(I+N)$. It follows that $AB = B^q A$ is equivalent to $AN = qNA$ and hence to $(A-qI)N = qN(A-I)$. Thus, we need to show that

$$(A-qI)N = qN(A-I) \iff N(A-I) = 0. \tag{7.5}$$

Set $C := (A-qI)N$ and recall that the main involution is given by the linear map

$$D := \begin{pmatrix} w & x \\ y & z \end{pmatrix} \mapsto D^\iota := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} D^t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} z & -x \\ -y & w \end{pmatrix}.$$

It satisfies $D + D^\iota = \operatorname{tr} D \cdot I$ and $(DD')^\iota = (D')^\iota D^\iota$. Hence, $\operatorname{tr} A = q-1$ and $\operatorname{tr} N = 0$ yield $(A-I)^\iota = (qI - A)$ and $N^\iota = -N$ and the anti commutativity implies that $C^\iota = N(A-I)$, so that the left-hand side of (7.5) is equivalent to $C = qC^\iota$.

Now, if $N(A-I) = 0$, then $C^\iota = 0$ and hence $C = qC^\iota = 0 = C^\iota$. Conversely, if $C = qC^\iota$, then $(q+1)C = qC + qC^\iota = q\operatorname{tr}(C)$ and so C must be a scalar matrix because $q+1$ is a non-zero divisor. But then $C^\iota = C$ and thus $C^\iota = qC$ and $q-1$ non-zero imply that $C = 0$ and hence $N(A-I) = C^\iota = 0$.

(b) The equations $e = a$ and $d = a$ are equivalent to $\operatorname{tr}(A-I) = q-1$ and $\operatorname{tr} N = 0$. The six 2×2 minors in (ii) express the following conditions on A and N: the left-hand minor means $\det(N) = 0$, the right-hand minor that $\det(A-I) = 0$ and the remaining minors express that $N(A-I) = 0$. These are precisely the conditions obtained in (a). □

Remark 7.12. Denote by $\widetilde{R}_q^{\mathrm{St}}$ the quotient of $\mathcal{R} = \mathcal{O}[[a,b,c,e,\alpha,\beta,\gamma,\delta]]$ by the ideal $\mathcal{I}_q^{\mathrm{St}} = (r_1^{\mathrm{St}}, \dots, r_8^{\mathrm{St}})$ from the proof of Proposition 7.9; except for a slight change of notation, this is the ring described by the relations coming from Lemma 7.11. To be able to identify this ring with R_q^{St}, we need to show that it is reduced and flat over $\mathcal{O}$, because Lemma 7.11 *a priori* only gives equations cutting out the closed points on the generic fiber. We now give a proof. The arguments have much overlap with those used in [Sho16].

As the r_i^{St}, $i = 1, \dots, 6$, are the 2×2 minors of (7.1), the ring

$$\mathcal{O}[\alpha,\beta,\gamma,\delta,a,b,c,e]/(r_1^{\mathrm{St}}, \dots, r_6^{\mathrm{St}})$$

is determinantal. By [BV88, § 2], this quotient ring is a normal Cohen–Macaulay domain of dimension 6 and hence its completion, say $\mathcal{T}$, at the maximal ideal $(\alpha,\beta,\gamma,\delta,a,b,c,e)$ is Cohen–Macaulay of dimension 6. A short calculation shows that the quotient of $\mathcal{T}$ modulo the six-term sequence $r_7^{\mathrm{St}}, r_8^{\mathrm{St}}, \varpi, \gamma-\beta, c+b, \beta+b$ is finite and hence the sequence is regular. It follows that $\widetilde{R}_q^{\mathrm{St}} = \mathcal{T}/(r_7^{\mathrm{St}}, r_8^{\mathrm{St}})$ is Cohen–Macaulay and flat over $\mathcal{O}$ of relative dimension 3. To prove its reducedness, we may thus invert p and rely on results on $R_q^{\square}[1/p]$.

Because $\widetilde{R}_q^{\mathrm{St}}$ is flat over $\mathcal{O}$, it has characteristic zero and hence Lemma 7.11 shows that the matrices A, B from the proof of Proposition 7.9 satisfy $AB = B^q A$. The universal property of $R_q^{\square}$ thus yields a surjective homomorphism $R_q^{\square} \to \widetilde{R}_q^{\mathrm{St}}$ in CNL. By the previous paragraph, the target of $R_q^{\square}[1/p] \to \widetilde{R}_q^{\mathrm{St}}[1/p]$ is equidimensional of dimension 3. Adapting [BGGT14, § 1.3] to a fixed determinant situation shows that $\operatorname{Spec} R_q^{\square}[1/p]$ contains a Zariski dense set of *robustly*

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

smooth points and hence all irreducible components of $\operatorname{Spec} R_q^{\square}[1/p]$ are generically formally smooth of dimension 3. We deduce that $\widetilde{R}_q^{\mathrm{St}}[1/p]$ is reduced.

Remark 7.13. Let us also give the coordinate change between the relations found here for $R_q^{\mathrm{St}} \cong \widetilde{R}_q^{\mathrm{St}} = \mathcal{R}/(r_1^{\mathrm{St}}, \dots, r_8^{\mathrm{St}})$ and those given in [Sho16, proof of Proposition 5.8, p. 1467]: the ring $\widetilde{R}_q^{\mathrm{St}}$ is to be compared with the ring $\mathcal{S} := \mathcal{B}/(J_0 + J_1 + (T - q + 1))$ from [Sho16], where $\mathcal{B} = \mathcal{O}[[X_1, \dots, X_4, Y_1, \dots, Y_4, T]]$, J_0 are the 2 minors of the 2×4 matrix with rows $(X_1 \ \dots X_4)$ and $(Y_1 \ \dots \ Y_4)$ and $J_1 = (X_1 + Y_2, Y_3 - X_4 + 2((q-1)/(q+1)))$; note that in the definition of J_1 in [Sho16], the variables Y_3 and X_3 as well as X_4 and Y_4 are mixed up. By going through the proof of [Sho16, Proposition 5.8], the change of variables from $\mathcal{S}$ is $X_1 = \alpha$, $X_2 = \beta$, $Y_1 = \gamma$, $Y_2 = -\alpha$, $X_3 = -2b/(q+1)$, $X_4 = (a + q - 1)/(q+1)$, $Y_3 = 2a/(q+1)$, $Y_4 = 2c/(q+1)$ and $T = q - 1$. It is immediate to see that the relations here and there agree.

7.4 Patching and global cotangent spaces

Consider a global augmentation $\lambda : R^{\mathrm{St}} \to \mathcal{O}$ arising from a newform in $f \in S_2(\Gamma_0(NQ))$ which is new at Q, and the augmentations it induces by pull back to other global deformation rings and local deformation rings, which by abuse of notation we denote by λ again. We assume that the residual representation $\bar{\rho}$ of $\rho_{f,\iota}$ satisfies the conditions of § 2. Further, if q is -1 mod p, such that $\bar{\rho}$ is unramified at q and $q \in Q$, and $\rho_{f,\iota}|_{G_q}$ is

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$$

up to twist by an unramified character of order dividing 2, then we choose β_q of § 4 to be 1 or -1 accordingly.

The main result of this section is as follows.

Theorem 7.14. *There is an isomorphism* $\Phi_{\lambda, R_\infty/R_\infty^{\mathrm{St}}} \xrightarrow{\sim} \Phi_{\lambda, R/R^{\mathrm{St}}}$.

Proof. We begin by observing that Theorem 6.3 shows that there exist rings R_∞, R_∞^{St} and $S_\infty = \mathcal{O}[[y_1, \dots, y_d]]$ with $\dim R_\infty = \dim R_\infty^{\mathrm{St}} = \dim S_\infty = d + 1$ and a commutative diagram

$$\begin{array}{ccccccc} S_\infty & \hookrightarrow & R_\infty & \twoheadrightarrow & R & & \\ \downarrow{=} & & \downarrow & & \downarrow & & \\ S_\infty & \hookrightarrow & R_\infty^{\mathrm{St}} & \twoheadrightarrow & R^{\mathrm{St}} & \xrightarrow{\lambda} & \mathcal{O} \end{array}$$

where $R = R_\infty \otimes_{S_\infty} \mathcal{O} = R_\infty/(y_1, \dots, y_d)$ and $R^{\mathrm{St}} = R_\infty^{\mathrm{St}} \otimes_{S_\infty} \mathcal{O} = R_\infty^{\mathrm{St}}/(y_1, \dots, y_d)$. We will also use λ to denote the maps $\lambda : R \twoheadrightarrow \mathcal{O}$, $\lambda : R_\infty \twoheadrightarrow \mathcal{O}$ and $\lambda : R_\infty^{\mathrm{St}} \twoheadrightarrow \mathcal{O}$ induced by $\lambda : R^{\mathrm{St}} \twoheadrightarrow \mathcal{O}$.

Now, by Lemma 7.5, the local ring maps $\mathcal{O} \to S_\infty \to R_\infty$ give rise to an exact sequence

$$\widehat{\Omega}_{S_\infty/\mathcal{O}} \otimes_{S_\infty} R_\infty \to \widehat{\Omega}_{R_\infty/\mathcal{O}} \to \widehat{\Omega}_{R_\infty/S_\infty} \to 0.$$

Applying $__ \otimes_{S_\infty} \mathcal{O} = __ \otimes_{R_\infty} R$ to this sequence gives

$$\widehat{\Omega}_{S_\infty/\mathcal{O}} \otimes_{S_\infty} R \to \widehat{\Omega}_{R_\infty/\mathcal{O}} \otimes_{R_\infty} R \to \widehat{\Omega}_{R_\infty/S_\infty} \otimes_{S_\infty} \mathcal{O} \to 0.$$

Now we have $\widehat{\Omega}_{S_\infty/\mathcal{O}} = \Omega_{\mathcal{O}[y_1,\dots,y_d]/\mathcal{O}} \otimes_{\mathcal{O}[y_1,\dots,y_d]} S_\infty \cong \mathcal{O}[y_1, \dots, y_d]^d \otimes_{\mathcal{O}[y_1,\dots,y_d]} S_\infty \cong S_\infty^d$. Also, by Lemma 7.2 and [Sta19, Lemma 00RV],

$$\widehat{\Omega}_{R_\infty/S_\infty} \otimes_{S_\infty} \mathcal{O} \cong \Omega_{R_\infty/S_\infty} \otimes_{S_\infty} \mathcal{O} \cong \Omega_{R/\mathcal{O}},$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

so the above exact sequence becomes

$$R^d \to \widehat{\Omega}_{R_\infty/\mathcal{O}} \otimes_{R_\infty} R \to \Omega_{R/\mathcal{O}} \to 0.$$

Finally, applying $__ \otimes_\lambda \mathcal{O}$ to this and noting that $\left(\widehat{\Omega}_{R_\infty/\mathcal{O}} \otimes_{R_\infty} R\right) \otimes_\lambda \mathcal{O} = \widehat{\Omega}_{R_\infty/\mathcal{O}} \otimes_\lambda \mathcal{O} = \Phi_{\lambda,R_\infty}$, we get the exact sequence

$$\mathcal{O}^d \to \Phi_{\lambda,R_\infty} \to \Phi_{\lambda,R} \to 0.$$

By an analogous argument with R^{St}_∞ and R^{St} in place of R_∞ and R, we now get a commutative diagram

$$\begin{array}{ccccccc}
 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \\
 & & \Phi_{\lambda,R_\infty/R^{\mathrm{St}}_\infty} & \longrightarrow & \Phi_{\lambda,R/R^{\mathrm{St}}} & & \\
 & & \downarrow & & \downarrow & & \\
\mathcal{O}^d & \longrightarrow & \Phi_{\lambda,R_\infty} & \longrightarrow & \Phi_{\lambda,R} & \longrightarrow & 0 \\
\| & & \downarrow & & \downarrow & & \\
\mathcal{O}^d & \longrightarrow & \Phi_{\lambda,R^{\mathrm{St}}_\infty} & \longrightarrow & \Phi_{\lambda,R^{\mathrm{St}}} & \longrightarrow & 0
\end{array}$$

with exact rows and columns.

Recall from Proposition 4.3 that for any $q \in Q$ there is a finitely generated $\mathcal{O}$-algebra $\mathcal{R}^{\mathrm{St}}_q$ for which R^{St}_q is the completion of $\mathcal{R}^{\mathrm{St}}_q$ at a maximal ideal. More specifically, we may take $\mathcal{R}^{\mathrm{St}}_q = \mathcal{O}[\alpha,\beta,\gamma,a,b,c]/I_q$, where I_q is generated by the 2×2 minors of the matrix in (7.4) if q is a trivial prime and $\mathcal{R}^{\mathrm{St}}_q = \mathcal{O}[a,b,c]$ otherwise. Now let

$$\mathcal{R}^{\mathrm{St}}_\infty = \left[\bigotimes_{q|Q} \mathcal{R}^{\mathrm{St}}_q\right] \otimes \left[\bigotimes_{\ell|N} \mathcal{O}[X_1,X_2,X_3]\right] \otimes \mathcal{O}[X_1,X_2,X_3,X_4] \otimes \mathcal{O}[x_1,\dots,x_g],$$

so that Proposition 4.3 gives that R^{St}_∞ is the completion of $\mathcal{R}^{\mathrm{St}}_\infty$ at a maximal ideal $\mathfrak{m}$.

Similarly, by the computations in [Sho16, § 5], there is also a natural choice of a finitely generated $\mathcal{O}$-algebra $\mathcal{R}_\infty$ such that $\mathcal{R}^{\mathrm{St}}_\infty$ is a quotient of $\mathcal{R}_\infty$ (making $\operatorname{Spec}\mathcal{R}^{\mathrm{St}}_\infty$ an irreducible component of $\operatorname{Spec}\mathcal{R}_\infty$) and R_∞ is the completion of $\mathcal{R}_\infty$ at the same maximal ideal $\mathfrak{m}$.

Now notice that

$$\begin{aligned}\Phi_{\lambda,R_\infty} \otimes_{\mathcal{O}} E &= \Omega_{\mathcal{R}_\infty/\mathcal{O}} \otimes_\lambda E = \left(\Omega_{\mathcal{R}_\infty/\mathcal{O}} \otimes_{\mathcal{R}_\infty} \mathcal{R}_\infty[1/\varpi]\right) \otimes_\lambda E \\ &= \Omega_{\mathcal{R}_\infty[1/\varpi]/\mathcal{O}} \otimes_\lambda E = \Omega_{\mathcal{R}_\infty[1/\varpi]/E} \otimes_\lambda E,\end{aligned}$$

by Lemma 7.3 and [Sta19, Lemma 00RT], and similarly

$$\begin{aligned}\Phi_{\lambda,R^{\mathrm{St}}_\infty} \otimes_{\mathcal{O}} E &= \Omega_{\mathcal{R}^{\mathrm{St}}_\infty[1/\varpi]/E} \otimes_\lambda E, \\ \Phi_{\lambda,R} \otimes_{\mathcal{O}} E &= \Omega_{R[1/\varpi]/E} \otimes_\lambda E, \\ \Phi_{\lambda,R^{\mathrm{St}}} \otimes_{\mathcal{O}} E &= \Omega_{R^{\mathrm{St}}[1/\varpi]/E} \otimes_\lambda E.\end{aligned}$$

Now observe that R and R^{St} are finite over $\mathcal{O}$ and reduced (by Theorem 6.3). This implies that $R[1/\varpi]$ and $R^{\mathrm{St}}[1/\varpi]$ are both finite direct products of finite extensions of E. It follows

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

that $\Phi_{\lambda,R} \otimes_{\mathcal{O}} E = \Phi_{\lambda,R^{\mathrm{St}}} \otimes_{\mathcal{O}} E = 0$ and so $\Phi_{\lambda,R}$ and $\Phi_{\lambda,R^{\mathrm{St}}}$ are both (finitely generated) torsion $\mathcal{O}$-modules.

As computed in [Sho16, Proposition 5.8], for each q, $\operatorname{Spec} \mathcal{R}_q^{\mathrm{St}}[1/\varpi]$ is a smooth, irreducible variety of dimension 3 over E (specifically, this follows from a computation identical to the computation that $\operatorname{Spec}(R^{\square}(\bar{\rho}, \tau_{1,\mathrm{ns}}) \otimes E)$ is formally smooth in [Sho16, Proposition 5.8]). It follows that $\operatorname{Spec} \mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]$ is a smooth, irreducible variety of dimension $d = \dim_{\mathcal{O}} R_\infty^{\mathrm{St}}$ over E. By [Sta19, Lemma 02G1], it follows that $\Omega_{\mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]/E}$ is locally free of rank d over $\mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]$ and so $\Phi_{\lambda,R_\infty} \otimes_{\mathcal{O}} E = \Omega_{\mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]/E} \otimes_\lambda E \cong E^d$.

Recall that $\operatorname{Spec} \mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]$ is an irreducible component of $\operatorname{Spec} \mathcal{R}_\infty[1/\varpi]$. Treating $\lambda : \mathcal{R}_\infty \twoheadrightarrow \mathcal{R}_\infty^{\mathrm{St}} \xrightarrow{\lambda} \mathcal{O} \hookrightarrow E$ as an E point of $\operatorname{Spec} \mathcal{R}_\infty[1/\varpi]$, we have that λ is not contained in any irreducible components of $\operatorname{Spec} \mathcal{R}_\infty[1/\varpi]$ other than $\operatorname{Spec} \mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]$ (as $\operatorname{Spec} \mathcal{R}_\infty[1/\varpi]$ must be smooth at λ, by Lemma 6.1). It follows from this that

$$\Phi_{\lambda,R_\infty} \otimes_{\mathcal{O}} E \cong \Omega_{\mathcal{R}_\infty[1/\varpi]/E} \otimes_\lambda E \cong \Omega_{\mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]/E} \otimes_\lambda E \cong E^d$$

(the second equality follows from [Sta19, Lemma 02G1] again, after noting that $\mathcal{R}_\infty[1/\varpi]_{(J)} \cong \mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]_{(J)}$, where $J = \ker(\mathcal{R}_\infty[1/\varpi] \xrightarrow{\lambda} E)$, and so $\left(\Omega_{\mathcal{R}_\infty[1/\varpi]/E}\right)_{(J)} \cong \left(\Omega_{\mathcal{R}_\infty^{\mathrm{St}}[1/\varpi]/E}\right)_{(J)}$).

Thus, Φ_{λ,R_∞} and $\Phi_{\lambda,R_\infty^{\mathrm{St}}}$ both have rank d as $\mathcal{O}$-modules.

But now Φ_{λ,R_∞} and $\mathcal{O}^d$ have the same rank as $\mathcal{O}$-modules and $\Phi_{\lambda,R} = \operatorname{coker}(\mathcal{O}^d \to \Phi_{\lambda,R_\infty})$ is a torsion $\mathcal{O}$-module. Hence, $\ker(\mathcal{O}^d \to \Phi_{\lambda,R_\infty})$ is a torsion $\mathcal{O}$-module and is thus 0. Therefore, the sequence

$$0 \to \mathcal{O}^d \to \Phi_{\lambda,R_\infty} \to \Phi_{\lambda,R} \to 0$$

from above is exact. By identical reasoning,

$$0 \to \mathcal{O}^d \to \Phi_{\lambda,R_\infty^{\mathrm{St}}} \to \Phi_{\lambda,R^{\mathrm{St}}} \to 0$$

is also exact. The above commutative diagram, together with the snake lemma, now gives that the natural map $\Phi_{\lambda,R_\infty/R_\infty^{\mathrm{St}}} \to \Phi_{\lambda,R/R^{\mathrm{St}}}$ is an isomorphism. □

Corollary 7.15.

$$\ell_{\mathcal{O}}(\Phi_{\lambda,R/R^{\mathrm{St}}}) = \ell_{\mathcal{O}}(\Phi_{\lambda,\mathbf{T}/\mathbf{T}^{\mathrm{St}}}) = \sum_{q \in Q} (m_q + \operatorname{ord}_{\mathcal{O}}(q^2 - 1) - 2n_q).$$

Proof. This is a consequence of Theorem 6.3, Corollary 7.10 and Theorem 7.14. □

Remark 7.16. The above result proves that the size of the relative global cotangent space at an augmentation of a global deformation ring is the same as the size of the relative local cotangent space. We prove this using a combination of level-raising results (as used in the proof of Lemma 6.2) and patching. Similar local–global results were proved in [Wil95] when the change in the local deformation condition is from unramified to unrestricted ramification. The arguments are different caused by the fact that the rings $R^{\mathrm{St}} = \mathbf{T}^{\mathrm{St}}$ are not complete intersections and hence have non-trivial Wiles defect with respect to our fixed augmentation λ (while in [Wil95] the minimal deformation and Hecke rings are complete intersections and hence satisfy the Wiles numerical criterion). Wiles computed an upper bound on change of cotangent space (when allowing ramification at a prime) and a matching lower bound on change in congruence module, which together imply an exact formula for change of cotangent space and congruence module. In contrast, in our work here to compute the Wiles defect for $R^{\mathrm{St}}, \mathbf{T}^{\mathrm{St}}$, we have to compute the exact change of cotangent spaces (between R and R^{St}) and congruence modules (between $\mathbf{T}$ and $\mathbf{T}^{\mathrm{St}}$) independently of each other.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

8. Relating congruence modules

We recall that the main result of [Man21] explicitly determines the structure of M_∞^{St} as an R_∞^{St}-module (or at least of $M_\infty^{\mathrm{St}}/\varpi$ as an $R_\infty^{\mathrm{St}}/\varpi$-module). In particular, provided that Q is divisible by at least one trivial prime for $\bar\rho$, M_∞^{St} is not free over R_∞^{St}. We will not need to use the full description of M_∞^{St} in this paper; however, we will use the following consequence of these computations (see [Man21, Theorem 1.2] or more specifically Propositions 4.13 and 4.14).

THEOREM 8.1. *In both the definite and indefinite cases, the map $R^{\mathrm{St}} \to \mathrm{End}_{R^{\mathrm{St}}}(M^{\mathrm{St}}(N))$ is an isomorphism.*

Remark 8.2. There is a small subtlety here in that [Man21] excludes the case where D ramifies at a prime which is $-1 \pmod p$, whereas we have not excluded this case. Fortunately the restriction in [Man21] is purely for the sake of convenience, so this does not present an issue. Under the notation of this paper we have $R_q^{\mathrm{St}(\beta_q)} \cong \mathcal{O}[[A,B,C]]$ at all primes for which $q \equiv -1 \pmod p$, which means that the ring R_∞^{St} considered in this paper will coincide with the ring R_∞ considered in [Man21] in all cases. Thus, the proof given in [Man21] will establish Theorem 8.1 in all cases considered in this paper, without modifications.

COROLLARY 8.3. *For an augmentation $\lambda : R^{\mathrm{St}} \to \mathcal{O}$ as above, we have $|\Psi_\lambda(R^{\mathrm{St}})| = |\Psi_\lambda(M^{\mathrm{St}}(N))|$ and $\delta_{\lambda,R^{\mathrm{St}}}(R^{\mathrm{St}}) = \delta_{\lambda,R^{\mathrm{St}}}(M^{\mathrm{St}}(N)) = \delta_{\lambda,R^{\mathrm{St}}}\big(S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}\big)$.*

Proof. This follows upon combining Theorems 3.12 and 8.1. □

Thus, to prove our main result, it will suffice to compute $\delta_{\lambda,\mathbf{T}^{\mathrm{St}}}\big(S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}\big)$, which will be our goal in the remainder of this paper.

9. Computation of change of the η-invariant

We put ourselves in the setting of § 5. Consider a quaternion algebra D over $\mathbf{Q}$ that is ramified at a finite set of primes Q (and hence is definite if $|Q|$ is odd and indefinite if $|Q|$ is even). We have the augmentation $\lambda_f : \mathbf{T}^{\mathrm{St}} \to \mathcal{O}$ which corresponds (by Jacquet–Langlands) to a newform $f \in S_2(\Gamma_0(NQ))$ with residual representation $\bar\rho$. Recall that $N|N(\bar\rho)|NQ$. We recall from § 5 that $\mathbf{T}^{\mathrm{St}}$ acts faithfully on $H^1(X_0^Q(N),\mathcal{O})_{\mathfrak{m}'_Q}$, with $\mathfrak{m}'_Q$ characterized by the property that it is the maximal ideal of the full Hecke algebra acting on $H^1(X_0^Q(N),\mathcal{O})$ that contains the kernel of the augmentation arising from f. We also have from § 5 the Hecke algebra $\mathbf{T}$ that acts faithfully on $S(NQ^2)_{\mathfrak{m}_Q} = H^1(X_0(NQ^2),\mathcal{O})_{\mathfrak{m}_Q}$, the oldform f^Q of level NQ^2 that gives rise to the maximal ideal $\mathfrak{m}_Q$ of the full Hecke algebra acting on $S(NQ^2) = H^1(X_0(NQ^2),\mathcal{O})$ and $S(NQ^2)^*_{\mathfrak{m}_Q} = M(NQ^2)^{\oplus 2}$.

Let $\mathcal{A}_f$ stand for the isogeny class of the abelian variety A_f (which is an optimal quotient of $J_0(NQ)$). The residual representations arising from the class $\mathcal{A}_f$ with respect to the fixed embedding $K_f \hookrightarrow \bar{\mathbf{Q}}_p$ are all isomorphic to our fixed absolutely irreducible $\bar\rho$. We have the following result, which is deduced from the methods of [RT97].

PROPOSITION 9.1. *Assume that Q contains a prime t that divides $N(\bar\rho)$ and we are in the indefinite case (i.e. that Q is of even cardinality). We have the equality of lengths of $\mathcal{O}$-modules:*

$$\ell_{\mathcal{O}}(\Psi_\lambda(M(NQ^2))) = \ell_{\mathcal{O}}(\Psi_\lambda(M^{\mathrm{St}}(N))) + \sum_{q\in Q}(m_q + \mathrm{ord}_{\mathcal{O}}(q^2-1)).$$

Proof. Let $Q = \{q_1, q_2, \dots, q_{2r}\}$ with $q_1 = t$ such that $t|N(\bar\rho)$ and, for $0 \le s \le r$, set $Q_s = \{q_1, \dots, q_{2r-2s}\}$ and J_s to be the Jacobian of $X_s = X_0^{Q_s}(Nq_{2r-2s+1}\cdots q_{2r})$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

To study the congruence modules of $H^1(X_s,\mathcal{O})_{\mathfrak{m}}$ using the work of § 3, we will need a Hecke equivariant perfect pairing on $H^1(X_s,\mathcal{O})_{\mathfrak{m}}$. In general, the perfect pairing $\langle\,,\rangle'_s : H^1(X_s,\mathcal{O}) \times H^1(X_s,\mathcal{O}) \to \mathcal{O}$ arising from Poincaré duality will not be Hecke equivariant (instead for any double coset operator T it will satisfy $\langle T^*x, y\rangle'_s = \langle x, T_*y\rangle'_s$, where T^* and T_* are induced respectively by Picard and Albanese functoriality, respectively). As in Lemma 5.5 of [Hel07], one may modify this to obtain a Hecke equivariant perfect pairing $\langle\,,\rangle_s : H^1(X_s,\mathcal{O}) \times H^1(X_s,\mathcal{O}) \to \mathcal{O}$ defined by $\langle x, y\rangle_s = \langle x, wy\rangle'_s$, where w is an appropriately chosen Atkin–Lehner involution. We will also use $\langle\,,\rangle_s$ to denote the localized perfect pairing $\langle\,,\rangle_s : H^1(X_s,\mathcal{O})_{\mathfrak{m}} \times H^1(X_s,\mathcal{O})_{\mathfrak{m}} \to \mathcal{O}$. There is an analogous Hecke equivariant perfect pairing $\langle\,,\rangle : H^1(X_0(NQ^2),\mathcal{O})_{\mathfrak{m}_Q} \times H^1(X_0(NQ^2),\mathcal{O})_{\mathfrak{m}_Q} \to \mathcal{O}$. The $\mathcal{O}$-modules $H^1(X_0,\mathcal{O})[\ker(\lambda_f)], H^1(X_r,\mathcal{O})[\ker(\lambda_f)], H^1(X_0(NQ^2),\mathcal{O})[\ker(\lambda_{f^Q})]$ are all free of rank 2, and we choose $\mathcal{O}$-bases $\{A_0, B_0\}$, $\{A_r, B_r\}$ and $\{A, B\}$, respectively.

- Step 1: we prove the equality $(\langle A_r, B_r\rangle_r) = (\prod_{q\in Q} \varpi^{m_q})(\langle A_0, B_0\rangle_0)$ of ideals of $\mathcal{O}$. This is done by using [RT97] and its extension in [Kha03].
- Step 2: we prove the equality $(\langle A, B\rangle) = (\prod_{q\in Q}(q^2-1))(\langle A_r, B_r\rangle_r)$ of ideals of $\mathcal{O}$. This uses the arguments of [Wil95, § 2], in particular [Wil95, Lemma 2.5] (see also [DR97, § 4]).

We go into more of the details of the two steps.

Step 1. In [RT97], the authors study the change in degree of parametrizations of (an isogeny class of) an elliptic curve by Shimura curves as the Shimura curves vary. We use the extension of [RT97] in [Kha03] to the case of parametrizations of A_f when A_f is not an elliptic curve. Thus, we first briefly recall some of the work of [Kha03].

Let $\xi_s : J_s \to A_s$ with $A_s \in \mathcal{A}_f$ be an optimal quotient: the map ξ_s is equivariant for the Hecke action and its kernel is connected. Analogously, we have the optimal quotient $\xi_{s+1} : J_{s+1} \to A_{s+1}$. (The optimal quotients that we consider are all isogenous to each other via isogenies defined over $\mathbf{Q}$ and which are Hecke equivariant.) We consider the maps $\mathrm{Ta}_p(J_s) \to \mathrm{Ta}_p(A_s)$ and $\mathrm{Ta}_p(J_{s+1}) \to \mathrm{Ta}_p(A_{s+1})$ that ξ_s and ξ_{s+1} induce on the corresponding Tate modules. To ease the notation, we set $J := J_s$ and $J' := J_{s+1}$, $A_s := A$ and $A_{s+1} := A'$, $\xi := \xi_s$ and $\xi' := \xi_{s+1}$ and also set $q_{2(r-s)} = q$ and $q_{2(r-s)-1} = q'$.

We have induced maps $\xi'^* : \mathrm{Ta}_p((A')^d)_{\mathfrak{m},\mathcal{O}} \to \mathrm{Ta}_p(J')_{\mathfrak{m},\mathcal{O}}$ and $\xi'_* : \mathrm{Ta}_p(J')_{\mathfrak{m},\mathcal{O}} \to \mathrm{Ta}_p(A')_{\mathfrak{m},\mathcal{O}}$ (injective and surjective respectively by optimality and their analogs for ξ), where for instance $\mathrm{Ta}_p((A')^d)_{\mathfrak{m},\mathcal{O}} := (\mathrm{Ta}_p((A')^d) \otimes_{\mathbf{Z}_p} \mathcal{O})_{\mathfrak{m}}$ is the localization at $\mathfrak{m}$ of p-adic Tate module of the dual abelian variety $(A')^d := \mathrm{Pic}^0(A')$ of A' tensored with $\mathcal{O}$ and the other symbols are defined analogously. Observe that $\mathrm{Ta}_p((A')^d)_{\mathfrak{m},\mathcal{O}}$ is free of rank 2 over $\mathcal{O}$. The module $\mathrm{Ta}_p((A')^d)_{\mathfrak{m},\mathcal{O}}$ is (non-canonically) isomorphic to $\mathrm{Ta}_p(A')_{\mathfrak{m},\mathcal{O}}$ as a $\mathcal{O}[G_{\mathbf{Q}}]$ module by [Car94] using the irreducibility of the residual representation $\bar\rho$. We identify these by choosing any such isomorphism: the ambiguity is only up to elements of $\mathcal{O}^*$, which is immaterial in the calculations below. Analogously, we have the maps induced by ξ involving a choice of a $\mathcal{O}[G_{\mathbf{Q}}]$-isomorphism between localizations at $\mathfrak{m}$ of the Tate module of A and its dual. The maps $\xi'_*\xi'^*$ and $\xi_*\xi^*$ of $\mathrm{Ta}_p(A')_{\mathfrak{m},\mathcal{O}}$ and $\mathrm{Ta}_p(A)_{\mathfrak{m},\mathcal{O}}$ commute with the $\mathcal{O}[G_{\mathbf{Q}}]$-action and by the irreducibility of this action can be regarded as given by multiplication by elements of $\mathcal{O}$. We denote the corresponding ideals of $\mathcal{O}$ by $(\xi'_*\xi'^*)$ and $(\xi_*\xi^*)$.

The method of [RT97] and [Kha03, Theorem 4] gives the key relation

$$(\xi'_*\xi'^*) = (\omega^{m_{q_{2r-2s}} + m_{q_{2r-2s-1}}})(\xi_*\xi^*)$$

for $0 \le s \le r-1$, where recall from Definition 7.7 that for $q \in Q$, m_q is defined to be the largest integer such that $\rho_f|_{I_q}$ is the identity mod ω^{m_q}. We justify how this result follows from

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

our hypotheses. Lemma 2 and Corollary on page 11113 of [RT97], and [Kha03, Proposition 3], using the vanishing of $\Phi_t(A)_{\mathcal{O}}$, the component group at t of A which is a consequence of $t|N(\bar{\rho})$, and the fact that t divides the discriminant of the quaternion algebra from which J arises, imply that the map $\xi_* : \phi_{q'}(J)_{\mathfrak{m},\mathcal{O}} \to \phi_{q'}(A)_{\mathfrak{m},\mathcal{O}}$ induced by ξ, with $\phi_{q'}(J)_{\mathfrak{m},\mathcal{O}}, \phi_{q'}(A)_{\mathfrak{m},\mathcal{O}}$ component groups at q' of J and A localized at $\mathfrak{m}$, is surjective. Using this and Ribet exact sequences [Kha03, Equation (2), p. 209], the asserted relation is deduced in [Kha03, Theorem 4].

This relation may be applied to compute change of congruence modules as follows. Let A_s, B_s be a basis of the free $\mathcal{O}$-module $H^1(X_s, \mathcal{O})_{\mathfrak{m}}[\ker \lambda]$ of rank 2 for $0 \leq s \leq r$. Then $(\langle A_s, B_s \rangle_s) = (\xi_{s*}\xi_s^*)$ as $(\langle \xi_s^*(x), \xi_s^*(y) \rangle_s) = (\langle x, \xi_{s*}\xi_s^*(y) \rangle)$, where $\{x, y\}$ is an $\mathcal{O}$-basis of $\mathrm{Ta}_p(A)_{\mathfrak{m},\mathcal{O}}$ and $(\langle x, y \rangle)$ is $\mathcal{O}$, where we use the perfect pairing on $\mathrm{Ta}_p(A)_{\mathfrak{m},\mathcal{O}}$ induced by the perfect pairings above. Using [Kha03, Theorem 4], applied successively for $0 \leq s \leq r-1$, we deduce that $(\langle A_r, B_r \rangle_r) = (\omega^{\Sigma_{q \in Q} m_q})(\langle A_0, B_0 \rangle_0)$.

Step 2. Let $q \in Q$ and consider the usual degeneracy map

$$(\delta_0, \delta_1) : H^1(X_0(NQ), \mathcal{O})^2 \to H^1(X_0(NQq), \mathcal{O}),$$

where δ_i is induced by the map of the upper half plane which sends $z \to q^i z$. Wiles's result [Wil95, Lemma 2.5] shows that the torsion in the cokernel of (δ_0, δ_1) is supported only at Eisenstein maximal ideals (i.e. the corresponding residual Galois representations are reducible) of the Hecke algebra acting on $H^1(X_0(NQq), \mathcal{O})$.

We claim that the map $\beta : H^1(X_0(NQ), \mathcal{O}) \to H^1(X_0(NQq), \mathcal{O})$ given by $\delta_0 - q^{-1}U_q\delta_1$ induces an isomorphism (that we denote by the same symbol) $\beta : H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)] \to H^1(X_0(NQq), \mathcal{O})[\ker(\lambda_{f^q})]$ (see [DR97, § 4] for related arguments in a similar situation). Here λ_{f^q} is the homomorphism of the full Hecke algebra acting on $S_2(\Gamma_0(NQq), \mathcal{O})$ associated to an oldform f^q in it with associated newform f such that f^q has the same eigenvalues as f for Hecke operators T_r (or U_r for $r|NQ$) for positive integers r that are prime to q and $f^q|U_q = 0$. To see the claim, we first note that as f is a newform of level NQ, the map $\beta : H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)] \to H^1(X_0(NQq), \mathcal{O})$ is injective (using an easy implication of [Wil95, Lemma 2.5] that can be seen by considering the map induced by β on $H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)] \otimes E$) and hence its image is a free $\mathcal{O}$-module of rank 2. Then the fact that $\beta : H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)] \to H^1(X(\Gamma_0(NQq), \mathcal{O})[\ker(\lambda_{f^q})]$ is actually an isomorphism follows from [Wil95, Lemma 2.5] and the fact that ρ_f is residually irreducible.

Let β' be the map $H^1(X_0(NQq), \mathcal{O}) \to H^1(X_0(NQ), \mathcal{O})$ dual to β. Then we see that the induced map (that we denote by the same symbol) $\beta'\beta : H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)] \to H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)]$ is (up to a unit) given by multiplication by $q^2 - 1$ (see also [DR97, § 4]).

Thus, if A_r, B_r as above is a basis of the free $\mathcal{O}$-module $H^1(X_0(NQ), \mathcal{O})[\ker(\lambda_f)]$ of rank 2, then $(\langle \beta(A_r), \beta(B_r) \rangle) = (\langle A_r, \beta'\beta(B_r) \rangle) = ((q^2-1)\langle A_r, B_r \rangle)$, where the first pairing is the twisted Poincaré pairing on the cohomology of $X_0(NQq)$. Iterating this argument $2r$ times, using the degeneracy maps $H^1(X_0(NQq_1 \dots q_t), \mathcal{O})^2 \to H^1(X_0(NQq_1 \dots q_t q_{t+1}), \mathcal{O})$, yields the equality of ideals $(\langle A, B \rangle) = (\prod_{q \in Q}(q^2-1))(\langle A_r, B_r \rangle_r)$ of $\mathcal{O}$, finishing Step 2.

Now, using $S(NQ^2)^*_{\mathfrak{m}_Q} = M(NQ^2)^{\oplus 2}$ and invoking Lemmas 3.5, 3.6 and 3.7 finishes the proof of Proposition 9.1. □

COROLLARY 9.2. *Assume that Q contains a prime that divides $N(\bar{\rho})$ and that Q is of even cardinality. We have the equality $\eta_\lambda(\mathbf{T}) = (\prod_{q \in Q}(q^2-1)\varpi^{m_q})\eta_\lambda(\mathbf{T}^{\mathrm{St}})$.*

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Proof. The result follows from Proposition 9.1, on using Corollary 8.3 (which gives that $|\Psi_\lambda(\mathbf{T}^{\mathrm{St}})| = |\Psi_\lambda(M^{\mathrm{St}}(N))|$) and Theorem 5.2 (which gives that $M(NQ^2)$ is free of rank 1 over $\mathbf{T}$ and hence $|\Psi_\lambda(\mathbf{T})| = |\Psi_\lambda(M(NQ^2))|$). □

Corollary 9.3.

1. *Assume that we are in the definite case (i.e. that $|Q|$ is odd) and in the case that $N = 1$ assume that $N(\bar\rho)$ is divisible by at least two primes. We have the equality*
$$\eta_\lambda(\mathbf{T}) = \left(\prod_{q\in Q}(q^2-1)\varpi^{m_q}\right)\eta_\lambda(\mathbf{T}^{\mathrm{St}}).$$
2. *Assume that we are in the indefinite case (i.e. that $|Q|$ is even) and in the case that $(Q, N(\bar\rho)) = 1$ that $N(\bar\rho)$ is divisible by at least two primes. We have the equality*
$$\eta_\lambda(\mathbf{T}) = \left(\prod_{q\in Q}(q^2-1)\varpi^{m_q}\right)\eta_\lambda(\mathbf{T}^{\mathrm{St}}).$$

Proof. Just for this proof, in which we vary Q and N, we denote by $\mathbf{T}^{\mathrm{St}}_{Q,N}$ the Hecke algebra previously denoted $\mathbf{T}^{\mathrm{St}}$ that acts faithfully on $H^1(X_0^Q(N), \mathcal{O})_{\mathfrak{m}'_Q}$. We reduce to the case proved in Corollary 9.2 by using the Jacquet–Langlands correspondence as follows.

1. Assume that $|Q|$ is odd.
   - If $N = 1$ and $t|(Q, N(\bar\rho))$, by the Jacquet–Langlands isomorphism, we have that $\mathbf{T}^{\mathrm{St}}_{Q,1} = \mathbf{T}^{\mathrm{St}}_{Q\setminus\{t\},t}$, and $(N(\bar\rho), Q/t) > 1$ by assumption.
   - If $N \neq 1$ and $t|N$ for a prime t, by the Jacquet–Langlands isomorphism, we have that $\mathbf{T}^{\mathrm{St}}_{Q,N} = \mathbf{T}^{\mathrm{St}}_{Q\cup\{t\},N/t}$.
2. Assume that Q is of even cardinality and that Q is coprime to $N(\bar\rho)$ (as the case when Q and $N(\bar\rho)$ are not coprime and Q is of even order is covered by Proposition 9.1). Thus, $N = N(\bar\rho)$, by assumption N is divisible by two primes t_1, t_2 and the Jacquet–Langlands isomorphism gives that $\mathbf{T}^{\mathrm{St}}_{Q,N} = \mathbf{T}^{\mathrm{St}}_{Q\cup\{t_1,t_2\},N/t_1t_2}$. □

Remark 9.4.

- We see from Corollary 9.2 that the change of the congruence module when dropping trivial primes from the discriminant is qualitatively the same as when dropping primes that are non-trivial, in contrast to the change of cotangent spaces in Corollary 7.15.
- One expects the general case of Proposition 9.1, without assuming that Q contains a prime t that divides $N(\bar\rho)$ and Q is of even cardinality, to be true, but we do not know how to modify its proof to yield this. One may remark (at least in the indefinite case, i.e. that Q is of even cardinality), that Corollary 9.3 can be established without the conditions assumed in it if we knew that the top exact sequence
$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Hom}(\mathcal{X}(J,q')_{\mathfrak{m},\mathcal{O}}, \mathcal{O}(1)) & \longrightarrow & \mathrm{Ta}_p(J)_{\mathfrak{m},\mathcal{O}} & \longrightarrow & \mathcal{X}(J,q')_{\mathfrak{m},\mathcal{O}} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathrm{Hom}(\mathcal{X}(A,q')_{\mathfrak{m},\mathcal{O}}, \mathcal{O}(1)) & \longrightarrow & \mathrm{Ta}_p(A)_{\mathfrak{m},\mathcal{O}} & \longrightarrow & \mathcal{X}(A,q')_{\mathfrak{m},\mathcal{O}} & \longrightarrow & 0 \end{array} \tag{9.1}$$
splits. This would imply that the map of component groups at $\phi(J,q')_{\mathfrak{m}} \to \phi(A,q')_{\mathfrak{m}}$ is surjective. Further, if q' is not 1 mod p, it is easy to see that the top exact sequence indeed splits. On the other hand, if q' is a trivial prime, then, not knowing this, we instead rely on the trick in [RT97], which uses the assumption that there is a prime t at which A has trivial

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

component group and which furthermore divides the discriminant of the quaternion algebra which gives rise to J.

- In forthcoming work we hope to be able to compute the change in congruence modules as in Proposition 9.1 without its superfluous assumptions, by computing the Wiles defect without using the work of [RT97].

10. Computing the Wiles defect

We can now prove our main theorem.

THEOREM 10.1. *Let N and Q be relatively prime squarefree integers. Let $p > 2$ be a prime not dividing NQ and let $E/\mathbf{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$, uniformizer ϖ and residue field k. Let $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O})$ be a Galois representation arising from a newform $f \in S_2(\Gamma_0(NQ))$ and let $\bar{\rho}_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ be the residual representation. Assume that $\bar{\rho}_f$ is irreducible and $N | N(\bar{\rho}_f)$.*

Let R^{St} be the Galois deformation ring of $\bar{\rho}_f$ defined in § 4, parameterizing lifts of $\bar{\rho}_f$ of fixed determinant which are Steinberg at each prime dividing Q, finite flat at p and minimal at all other primes.

Let D be the quaternion algebra with discriminant Q and let $\Gamma_0^Q(N)$ be the congruence subgroup for D defined in § 5. Let $\mathbf{T}^Q(N)$ and $S^Q(\Gamma_0^Q(N))$ be the Hecke algebra and cohomological Hecke module at level $\Gamma_0^Q(N)$ and let $\mathfrak{m} \subseteq \mathbf{T}^Q(N)$ be the maximal ideal corresponding to $\mathfrak{m}$. Let $\mathbf{T}^{\mathrm{St}} = \mathbf{T}^Q(N)_{\mathfrak{m}}$ and let $\lambda : \mathbf{T}^{\mathrm{St}} \to \mathcal{O}$ be the augmentation corresponding to f.

Assume that at least one of the following holds:

1. *Q is a product of an even number of primes (i.e. D is indefinite) and $(N(\bar{\rho}), Q) > 1$;*
2. *Q is a product of an odd number of primes (i.e. D is definite) and $N > 1$;*
3. *$N(\bar{\rho})$ is divisible by at least two primes.*

Then the Wiles defects (defined in Definition 3.10) of $\mathbf{T}^{\mathrm{St}}$ and $S^Q(\Gamma_0^Q(N))$ with respect to the map $R^{\mathrm{St}} \twoheadrightarrow \mathbf{T}^{\mathrm{St}}$ and the augmentation λ are

$$\delta_\lambda(\mathbf{T}^{\mathrm{St}}) = \delta_\lambda(S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}) = \delta_\lambda(M^{\mathrm{St}}(N)) = \sum_{q|Q} \frac{2n_q}{e},$$

where e is the ramification index of $\mathcal{O}$ and, for each $q|Q$, n_q is the largest integer for which $\rho_f|_{G_{\mathbf{Q}_q}}$ $(\mathrm{mod}\ \varpi^{n_q})$ is unramified and $\rho_f(\mathrm{Frob}_q) \equiv \pm \mathrm{Id}\ (\mathrm{mod}\ \varpi^{n_q})$.

Proof. The first two equalities $\delta_\lambda(\mathbf{T}^{\mathrm{St}}) = \delta_\lambda(S^Q(\Gamma_0^Q(N))_{\mathfrak{m}'_Q}) = \delta_\lambda(M^{\mathrm{St}}(N))$ follow from Corollary 8.3. For the remaining equality

$$\delta_\lambda(\mathbf{T}^{\mathrm{St}}) = \sum_{q|Q} \frac{2n_q}{e},$$

we study the map $R^{\mathrm{St}} \to \mathbf{T}^{\mathrm{St}}$, with augmentation $\lambda = \lambda_f$, in relation to the isomorphism $R \to \mathbf{T}$ of complete intersections of Theorem 5.2, which arises when we relax the ramification conditions at Q, and the induced augmentation $\lambda : \mathbf{T} \to \mathcal{O}$ as in § 5. From Theorems 5.2 and 1.1, we deduce that $|\Phi_\lambda(R)| = |\Phi_\lambda(\mathbf{T})| = |\mathcal{O}/\eta_\lambda(\mathbf{T})|$. Combining this with:

- Corollary 7.15, which gives

$$\ell_{\mathcal{O}}(\Phi_{\lambda,\mathbf{T}/\mathbf{T}^{\mathrm{St}}}) = \sum_{q \in Q} (m_q + \mathrm{ord}_{\mathcal{O}}(q^2 - 1) - 2n_q);$$

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

- Corollary 9.3, which gives $\eta_\lambda(\mathbf{T}) = (\prod_{q\in Q}(q^2-1)\varpi^{m_q})\eta_\lambda(\mathbf{T}^{\mathrm{St}})$

proves our theorem. □

Remark 10.2. The Wiles defect in the situation studied in this paper (say when $\mathcal{O} = \mathbf{Z}_p$) is always an even number. It would be interesting to have a conceptual explanation for this. We hope to return to this question in a future work which will give a computation of the Wiles defect that has a stronger local–global flavor.

11. Wiles defect for semistable elliptic curves and tame regulators

The Wiles defect associated to an augmentation $\lambda_f : \mathbf{T}^{\mathrm{St}} \to \mathcal{O}$ is a global invariant and as defined seems hard to compute (it is an 'f-extrinsic' invariant in Mazur's suggestive terminology, see [Maz12]), while the n_q, which our theorem relates to the Wiles defect of λ_f, are local invariants of the local Galois representation $\rho_f|_{G_q}$ (and are 'f-intrinsic' in the sense of Mazur).

In this section, we illustrate how Theorem 10.1 makes the Wiles defect rather easily computable in many examples. Let $\mathcal{E}$ be a semistable elliptic curve over $\mathbf{Q}$ of conductor N. Let p be a prime such that the mod p representation $\bar\rho$ arising from $\mathcal{E}$ is irreducible and assume that $(N,p) = 1$.

11.1 Inertial invariants n_q for Tate elliptic curves over $\mathbf{Q}_q$

Let $\rho_{\mathcal{E},p} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_p)$ be the representation on the p-adic Tate module of $\mathcal{E}$. Then the local invariant n_q at $q|N$ is the maximal integer such that $\rho_{\mathcal{E},p} \bmod p^{n_q}$ is unramified and sends Frob_q to $\pm\mathrm{Id}$. The quantity n_q may be computed using Tate's uniformization of $\mathcal{E}_{\mathbf{Q}_q}$ and Kummer theory.

Write $v_q\colon \mathbf{Q}_q^\times \to \mathbf{Z}$ and $|\cdot|_q\colon \mathbf{Q}_q^\times \to q^{\mathbf{Z}}$ for the normalized valuation and absolute value at q. Let $q_{\mathcal{E}_{\mathbf{Q}_q}} \in \mathbf{Q}_q^\times$ be the Tate period of $\mathcal{E}_{\mathbf{Q}_q}$, define m_q and t_q so that $p^{m_q}||v_q(q_{\mathcal{E}_{\mathbf{Q}_q}})$ and $p^{t_q}||q-1$, set $k_q = \min(m_q, t_q)$ and write $\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}} = q_{\mathcal{E}_{\mathbf{Q}_q}} \cdot |q_{\mathcal{E}_{\mathbf{Q}_q}}|_q$ for the unit part of the Tate period. Consider the map $\log_{q,p^{k_q}} : (\mathbf{Z}/q\mathbf{Z})^* \twoheadrightarrow \mathbf{Z}/p^{k_q}\mathbf{Z}$ and denote for an element $x \in (\mathbf{Z}/q\mathbf{Z})^*$ by $\mathrm{ind}_{q,p^{k_q}}(x)$ the index of the subgroup generated by $\log_{q,p^{k_q}}(x)$ in $\mathbf{Z}/p^{k_q}\mathbf{Z}$ (tame regulator á la [MT87]).

Proposition 11.1. *We have the formula* $n_q = \log_p\left(\mathrm{ind}_{q,p^{k_q}}(\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}})\right)$.

Proof. This follows from Kummer theory and Tate's uniformization of $\mathcal{E}$ at q. □

By the modularity theorem, we know that $\mathcal{E}$ arises from $f \in S_2(\Gamma_0(N))$. Assume that $N(\bar\rho)$ is not prime. Let Q be a set of primes which divide N, such that N factors as $N = N'Q$ with $N'|N(\bar\rho)$, let $\mathbf{T}^{\mathrm{St}} = \mathbf{T}^Q(N')$ and consider the augmentation $\lambda_{\mathcal{E}} : \mathbf{T}^{\mathrm{St}} \to \mathbf{Z}_p$ arising from f. (Our future work will address removing the condition that $N(\bar\rho)$ is not prime, as it will avoid dependence on the results of [RT97].)

Corollary 11.2.

$$\delta_{\lambda_{\mathcal{E}}}(\mathbf{T}^{\mathrm{St}}) = \Sigma_{q\in Q} 2\log_p(\mathrm{ind}_{q,p^{k_q}}(\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}})).$$

Proof. This follows from Theorem 10.1. □

Remark 11.3. Let q be a prime of multiplicative reduction of $\mathcal{E}$ with Tate period $q_{\mathcal{E}_{\mathbf{Q}_q}}$ and let $j_{\mathcal{E}}$ denote the j-invariant of $\mathcal{E}$. It takes some effort to compute $q_{\mathcal{E}_{\mathbf{Q}_q}}$ to some non-trivial precision. However, it is rather elementary to compute its unit part from $j_{\mathcal{E}}$. For this, recall from [Sil94, Theorem V.3.1(b)] that $v_q(q_{\mathcal{E}_{\mathbf{Q}_q}}) = -v_q(j_{\mathcal{E}}) > 0$ and that

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

$j_{\mathcal{E}} = 1/q_{\mathcal{E}_{\mathbf{Q}_q}} + 744 + O(q)$. Hence, $j_{\mathcal{E}}|j_{\mathcal{E}}| = |j_{\mathcal{E}}|(1/q_{\mathcal{E}_{\mathbf{Q}_q}}) + O(q)$ is a unit and we deduce that

$$\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}} \equiv j_{\mathcal{E}}^{-1}|j_{\mathcal{E}}|_q^{-1} \in (\mathbf{Z}_q/q\mathbf{Z}_q)^\times = \mathbf{F}_q^\times.$$

11.2 Numerical computations

It seemed natural to carry out computations for the example given in the introduction, where the initial level is $\Gamma_0(11)$, the prime is $p = 3$ and a trivial prime mod p occurs at $q = 193$. However, as indicated in Remark 1.3, this leads to a form f in level $11 \cdot 193$ with $[K_f : \mathbf{Q}] = 46$.[6] Even if in principle one can compute n_{193} using Mumford–Raynaud uniformization of A_f at 193 and Kummer theory, as in Proposition 11.1, to us the condition $[K_f : \mathbf{Q}] = 46$ made computations impossible.

The importance of Proposition 11.1 is that it allows practical computations for a different class of examples. Namely, one starts with an elliptic curve $\mathcal{E}$ of a 'large' squarefree level $\mathcal{N}$ with associated cusp form f; think of $\mathcal{N} = NQ$ with N, Q as in Theorem 10.1. Then one searches for prime divisors q of $\mathcal{N}$ at which $\rho_{f,p}$ admits level lowering modulo a prime p. Proposition 11.1 allows one to then compute the local invariant attached to $(\mathcal{E}, q, p)$. We eliminate cases where $p|\mathcal{N}$, so that $\rho_{f,p}$ is finite at p, and we check that $\bar{\rho}_{f,p}$ is irreducible, so that level-lowering results apply. We also ensure that the minimal conductor $N(\bar{\rho}_{f,p})$ of the reduction $\bar{\rho}_{f,p}$ has at least two prime divisors, so that Theorem 10.1 is applicable unconditionally for any $Q > 1$. Using the data in [LMF19], it is also often possible to determine the form g in level $N(\bar{\rho}_{f,p})$. This is not needed though to compute the invariant n_q or the Wiles defect for Q.

Concretely, we did an incremental search through all elliptic curves of squarefree conductor $\mathcal{N} < 400\,000$ in the Cremona tables. For each prime divisor q of $\mathcal{N}$, we make a list of all odd primes p that divide $q - 1$ but not $\mathcal{N}$. Next we compute for each elliptic curve $\mathcal{E}$ of conductor $\mathcal{N}$ the q-valuation $v_q(q_{\mathcal{E}_{\mathbf{Q}_q}})$ and the unit part $\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}}$ of its Tate period at q. Finally, we compute the quantities $t_q = \mathrm{val}_p(q-1)$, $m_q = v_p(v_q(q_{\mathcal{E}_{\mathbf{Q}_q}}))$ and $u_q = v_p(|\mathbf{F}_q^*/\langle\widetilde{q_{\mathcal{E}_{\mathbf{Q}_q}}}\rangle|)$. Then $n_q = \min\{t_q, m_q, u_q\}$ is the local invariant from Proposition 11.1 (it also depends on p and $\mathcal{E}$). For all primes p in our list, found for $\mathcal{E}$ and some q, we determine if the Galois representation $\bar{\rho}_{\mathcal{E},p}$ on the p-torsion points of $\mathcal{E}$ is absolutely irreducible. Then level lowering is possible if and only if $m_q > 0$. Using this, we can compute $N(\bar{\rho}_{f,p})$ and single out those with at least two prime divisors. The Wiles defect can then be computed as $2\sum_{q\in Q} n_q$ over the local invariants. All computations were carried out using SageMath.

For the primes $p = 3, 5, 7, 11$, the smallest examples $(\mathcal{E}, q, p)$, small in terms of size of conductor $\mathcal{N}$, with $n_q \geq 1$, are the following:

$$(805b1, 7, 3), \quad (5673a1, 61, 5), \quad (4171a1, 43, 7), \quad (27186m1, 23, 11),$$

where we use Cremona notation for the labeling of elliptic curves $\mathcal{E}$; the number to the left of the letter is the conductor of $\mathcal{E}$. For $\mathcal{N} < 400\,000$, we found the following:

Prime p	3	5	7	11
$\#\{(\mathcal{E}, q) \mid n_q > 0\}$	8346	950	43	3

[6] There are eight eigenspaces (over $\mathbf{Q}$) in level 2123. Two of them have the same Atkin–Lehner sign at 11 as the unique eigenform in level 11. The eigenspaces have dimension 33 and 46, and we call representing forms f_{33} and f_{46}. The eigenspace in level 11 is $\mathbf{Q}$-rational. Hence, the mod 3 reduction of the Hecke algebra of the wanted form in level 2123 needs to have a factor $\mathbf{F}_3$. By [Edi92, Theorem 2.5], this implies for the mod 3 Hecke eigenvalue a_3 that $a_3^2 + 1$ lies in $\mathbf{F}_3$. It turns out that no such root exists for the mod 3 reduction of T_3 for f_{33}; for f_{46}, exactly one such root (with multiplicity 7) lies in $\mathbf{F}_3$.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

For $p = 5, 7, 11$, we found no example where $n_q \geq 2$ and no example with $n_q = 1$ at two prime divisors q of $\mathcal{N}$ (for a fixed $\mathcal{E}$). For $13 \leq p \leq 61$, we found no $(\mathcal{E}, q)$ with $n_q > 0$. From this, we suspect that for $\mathcal{N} < 400\,000$ and $p \geq 5$ the Wiles defect of $\mathcal{E}$ is either 0 or 2.

For $p = 3$, we also found pairs $(\mathcal{E}, q)$ with $n_q = 2$ (none with $n_q > 2$); there were 25 such examples with $\mathcal{N} < 400\,000$; they all had a single prime divisor $q|\mathcal{N}$ at which the local invariant was non-zero. Hence, for $Q \supseteq \{q\}$, the Wiles defect is 4 in these cases. The two examples $(\mathcal{E}, q)$ of smallest conductor and with $n_q = 2$ are

$$(36613a1, 19), \quad (104710l1, 37).$$

They have $N(\bar{\rho}_{\mathcal{E},q}) = 1927 = 41 \cdot 47$ and $N(\bar{\rho}_{\mathcal{E},q}) = 2830 = 2 \cdot 5 \cdot 283$, respectively.

For $p = 3$ only, we also found elliptic curves $\mathcal{E}$ which at two primes $q \neq q'$ dividing $\mathcal{N}$ had invariants $n_q = n_{q'} = 1$. There were 15 of these. For these and for $Q \supseteq \{q, q'\}$, the Wiles defect is again 4. In all other cases the Wiles defect was 2 when it is non-zero. The two examples $(\mathcal{E}, q, q')$ of smallest conductor and with $n_q = n_{q'} = 1$ are

$$(149149b1, 7, 13), \quad (149149c1, 7, 13)$$

with $N(\bar{\rho}_{\mathcal{E},q}) = 1639 = 11 \cdot 149$.

We also determined in many of the above cases the form g congruent to f mod p of minimal level using the database [LMF19]. One very helpful invariant when determining g were the Atkin–Lehner signs. The following table displays our findings. In the second row we use the labeling of forms given in [LMF19]; note that as in the case of elliptic forms, the label of the cusp form g contains its level as the left-hand number.

Triple $(\mathcal{E}, q, p)$	$(805b1, 7, 3)$	$(5673a1, 61, 5)$	$(4171.a1, 43, 7)$	$(27186.m1, 23, 11)$	$(104710l1, 37, 3)$
Label of g	$115.a$	$93.2.a.a$	$1731.a.f$	$1182.2.a.d$	$2830g1$

The last example gives a rare example of a congruence of elliptic curves, since also g is $\mathbf{Q}$-rational.

Acknowledgements

G.B. received support from the DFG grant FG 1920. We would like to thank the referees for a critical reading that led to an improvement of the exposition and the correction of a few inaccuracies of an earlier version.

References

All16 P. Allen, *Deformations of polarized automorphic Galois representations and adjoint Selmer groups*, Duke Math. J. **165** (2016), 2407–2460.

BGGT14 T. Barnet-Lamb, T. Gee, D. Geraghty and R. Taylor, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), 501–609; MR 3152941.

BGHT11 T. Barnet-Lamb, D. Geraghty, M. Harris and R. Taylor, *A family of Calabi–Yau varieties and potential automorphy II*, Publ. Res. Inst. Math. Sci. **47** (2011), 29–98; MR 2827723.

BV88 W. Bruns and U. Vetter, *Determinantal rings*, Lecture Notes in Mathematics, vol. 1327 (Springer, Berlin, 1988); MR 953963.

CEG+16 A. Caraiani, M. Emerton, T. Gee, D. Geraghty, V. Paškūnas and S. W. Shin, *Patching and the p-adic local Langlands correspondence*, Camb. J. Math. **4** (2016), 197–287; MR 3529394.

Car94 H. Carayol, *Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet*, in *p-adic monodromy and the Birch and Swinnerton-Dyer conjecture (Boston, MA, 1991)*, Contemporary Mathematics, vol. 165 (American Mathematical Society, Providence, RI, 1994), 213–237; MR 1279611.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

CHT08 L. Clozel, M. Harris and R. Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Publ. Math. Inst. Hautes Études Sci. **108** (2008), 1–181, with Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras; MR 2470687.

DDT97 H. Darmon, F. Diamond and R. Taylor, *Fermat's last theorem*, in *Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993)* (International Press, Cambridge, MA, 1997), 2–140; MR 1605752.

Dia97 F. Diamond, *The Taylor–Wiles construction and multiplicity one*, Invent. Math. **128** (1997), 379–391; MR 1440309.

DR97 F. Diamond and K. Ribet, *ℓ-adic deformations and Wiles's 'main conjecture'*, in *Modular forms and Fermat's last theorem (Boston, MA, 1995)* (Springer, New York, 1997), 357–371; MR 1638487.

DT94a F. Diamond and R. Taylor, *Lifting modular mod l representations*, Duke Math. J. **74** (1994), 253–269; MR 1272977.

DT94b F. Diamond and R. Taylor, *Nonoptimal levels of mod l modular representations*, Invent. Math. **115** (1994), 435–462; MR 1262939.

Edi92 B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), 563–594; MR 1176206.

Eis95 D. Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics, vol. 150 (Springer, New York, 1995), with a view toward algebraic geometry; MR 1322960.

EGS15 M. Emerton, T. Gee and D. Savitt, *Lattices in the cohomology of Shimura curves*, Invent. Math. **200** (2015), 1–96; MR 3323575.

FKP19 N. Fakhruddin, C. Khare and S. Patrikis, *Relative deformation theory and lifting irreducible Galois representations*, Duke Math.J., to appear. Preprint (2019), arXiv:1904.02374.

Fuj06 K. Fujiwara, *Galois deformations and arithmetic geometry of Shimura varieties*, International Congress of Mathematicians, vol. II (European Mathematical Society, Zürich, 2006), 347–371; MR 2275601.

HR08 S. Hamblen and R. Ramakrishna, *Deformations of certain reducible Galois representations. II*, Amer. J. Math. **130** (2008), 913–944; MR 2427004.

Hel07 D. Helm, *On maps between modular Jacobians and Jacobians of Shimura curves*, Israel J. Math. **160** (2007), 61–117; MR 2342491.

Hid81a H. Hida, *Congruence of cusp forms and special values of their zeta functions*, Invent. Math. **63** (1981), 225–261; MR 610538.

Hid81b H. Hida, *On congruence divisors of cusp forms as factors of the special values of their zeta functions*, Invent. Math. **64** (1981), 221–262; MR 629471.

Kha03 C. Khare, *On isomorphisms between deformation rings and Hecke rings*, Invent. Math. **154**(1) (2003), 199–222, with an appendix by Gebhard Böckle; MR 2004460.

KW09 C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture. I*, Invent. Math. **178** (2009), 485–504; MR 2551763 (2010k:11087).

Kis09 M. Kisin, *Moduli of finite flat group schemes, and modularity*, Ann. of Math. (2) **170** (2009), 1085–1180; MR 2600871.

Len95 H. W. Lenstra Jr., *Complete intersections and Gorenstein rings*, in *Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993)*, Series in Number Theory, vol. I International Press, Cambridge, MA, 1995, 99–109; MR 1363497.

LMF19 The LMFDB Collaboration, *The l-functions and modular forms database*, 2019, http://www.lmfdb.org [online; accessed 8 September 2019].

Man21 J. Manning, *Patching and multiplicity 2^k for Shimura curves*, Algebra Number Theory **15** (2021), 387–434.

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press

Maz12 B. Mazur, *A brief introduction to the work of Haruzo Hida* (2012), http://www.math.harvard.edu/mazur/papers/Hida.August11.pdf.

MT87 B. Mazur and J. Tate, *Refined conjectures of the 'Birch and Swinnerton-Dyer type'*, Duke Math. J. **54** (1987), 711–750; MR 899413.

Rib84 K. A. Ribet, *Congruence relations between modular forms*, in *Proceedings of the International Congress of Mathematicians, vols. 1, 2 (Warsaw, 1983)* (PWN, Warsaw, 1984), 503–514; MR 804706.

Rib90a K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\bar{\boldsymbol{Q}}/\boldsymbol{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), 431–476; MR 1047143.

Rib90b K. A. Ribet, *Multiplicities of Galois representations in Jacobians of Shimura curves*, in *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part II (Ramat Aviv, 1989)*, Israel Mathematical Conference Proceedings, vol. 3 (Weizmann, *Jerusalem*, 1990), 221–236; MR 1159117.

RT97 K. A. Ribet and S. Takahashi, *Parametrizations of elliptic curves by Shimura curves and by classical modular curves*, Proc. Natl. Acad. Sci. USA **94**(21) (1997), 11110–11114, *Elliptic curves and modular forms* (Washington, DC, 1996); MR 1491967.

Sho16 J. Shotton, *Local deformation rings for* GL_2 *and a Breuil–Mézard conjecture when* $\ell \neq p$, Algebra Number Theory **10** (2016), 1437–1475; MR 3554238.

Sho18 J. Shotton, *The Breuil–Mézard conjecture when* $l \neq p$, Duke Math. J. **167** (2018), 603–678; MR 3769675.

Sil94 J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 151 (Springer, New York, 1994); MR 1312368.

Sno18 A. Snowden, *Singularities of ordinary deformation rings*, Math. Z. **288** (2018), 759–781; MR 3778977.

Sta19 The Stacks Project Authors, *Stacks project* (2019), http://stacks.math.columbia.edu.

Tay08 R. Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations. II*, Publ. Math. Inst. Hautes Études Sci. **108** (2008) 183–239; MR 2470688.

Tho16 J. A. Thorne, *Automorphy of some residually dihedral Galois representations*, Math. Ann. **364** (2016), 589–648; MR 3451399.

TU18 J. Tilouine and E. Urban, *Integral period relations and congruences*, Preprint (2018), arXiv:1811.11166.

WITO69 K.-I. Watanabe, T. Ishikawa, S. Tachibana and K. Otsuka, *On tensor products of Gorenstein rings*, J. Math. Kyoto Univ. **9** (1969), 413–423; MR 257062.

Wil95 A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), 443–551; MR 1333035.

Gebhard Böckle gebhard.boeckle@iwr.uni-heidelberg.de
Interdisciplinary Center for Scientific Computing, Universität Heidelberg, INF 205, 69120 Heidelberg, Germany

Chandrashekhar B. Khare shekhar@math.ucla.edu
Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA

Jeffrey Manning jmanning@math.ucla.edu
Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA

https://doi.org/10.1112/S0010437X21007454 Published online by Cambridge University Press