VANISHING FOURIER COEFFICIENTS OF HECKE EIGENFORMS

FRANK CALEGARI AND NASER TALEBIZADEH SARDARI

Abstract. We prove that, for fixed level $(N,p)=1$ and $p>2$, there are only finitely many Hecke eigenforms f of level $\Gamma_1(N)$ and even weight with $a_p(f)=0$ which are not CM.

1. Introduction

Lehmer [Leh47] raised the question of whether $\tau(n)=0$ for any of the non-trivial Fourier coefficients of Ramanujan's Delta function $\Delta = q\prod_{n=1}^{\infty}(1-q^n)^{24} = \sum \tau(n)q^n$. He proved that if $\tau(n)=0$ for some n, then necessarily $\tau(p)=0$ for a prime $p|n$. Lehmer's problem remains open, as does the analogous question for *any* cuspidal eigenform f of level one. If one weakens the hypothesis further and assumes only that f has level N for some N prime to p, then there are a number of ways in which $a_p(f)=0$, including the following:

(1) If f is a modular form with CM arising from an imaginary quadratic field $F/\mathbf{Q}$ in which p is inert, then $a_p(f)=0$.
(2) If f is a weight two modular form arising from an elliptic curve $E/\mathbf{Q}$ with good supersingular reduction at p, and $p\geq 5$, then $a_p(f)=0$.

In this paper, we examine a vertical analogue of Lehmer's conjecture where p is fixed and we vary the weight. Our main theorem is as follows:

Theorem 1.0.1. *Fix a prime $p>2$ and an integer $(N,p)=1$. Then there are only finitely many non-CM Hecke eigenforms of level N and even weight with $a_p(f)=0$.*

We shall deduce from this the following:

Corollary 1.0.2. *Fix a prime p. There are only finitely many eigenforms of level 1 with $a_p(f)=0$.*

Our arguments are not effective. The existence of non-CM (modular) elliptic curves $E/\mathbf{Q}$ which are supersingular at p shows that some exceptions must be included. The assumption $p>2$ and the assumptions on the weight are not intrinsic to our method, but rather reflect the absence of certain $R=\mathbf{T}$ theorems either when $p=2$ or when the residual representation $\overline{\rho}_f|_{G(\mathbf{Q}(\zeta_p))}$ is reducible. The weight condition can be weakened to requiring either that the weight n is even or $n-1$ is not divisible by $(p+1)/2$.

We now explain two further motivations for considering this problem (in addition to the analogy with Lehmer's question).

The first author was supported in part by NSF Grants DMS-1701703 and DMS-2001097. The second author was supported in part by NSF Grant DMS-2015305.

1.1. **Analogies with counting Maass forms.** Let $\mathbf{A}_{\mathbf{Q}}^{(\infty)}$ denote the finite adeles, let $G = \mathrm{GL}(2)/\mathbf{Q}$, and let $U \subseteq G(\mathbf{A}_{\mathbf{Q}}^{(\infty)})$ denote a compact open subgroup. The problem of counting spaces of cuspidal modular forms of level U and weight $n \geq 2$ amounts to computing the sum

$$\sum \dim \pi^U \tag{1}$$

as $\pi \subset L^2_{\mathrm{cusp}}(G(\mathbf{Q})\backslash G(\mathbf{A_Q}), \chi)$ ranges over all cuspidal automorphic representations (with a fixed suitably chosen central character χ) such that π_∞ corresponds to a discrete series representation $\mathcal{D}_n$ of weight n. In contrast, the problem of counting spaces of algebraic Maass forms with eigenvalue $\lambda = 1/4$ amounts to the same sum (1) except now where π_∞ corresponds to a particular principal series representation. The philosophical explanation for why the first sum can be estimated precisely using the trace formula while the latter can not is that discrete series representations have positive measure in the Plancherel measure of the unitary dual of $\mathrm{PGL}_2(\mathbf{R})$ whereas any fixed principal series does not. (For this perspective on counting automorphic forms, see [Shi12].) Finally, consider the problem of counting modular forms of weight $n \geq 2$ and level prime to p with $a_p = 0$ (the subject of this paper). This amounts to computing the same sum (1) where once more π_∞ corresponds to the discrete series $\mathcal{D}_n$, but now one *additionally* requires that π_p is the spherical representation of $\mathrm{GL}_2(\mathbf{Q}_p)$ with given central character and with Satake parameters α and β satisfying $\alpha + \beta = 0$. The obstracle in computing this sum is the same problem as for Maass forms except now the difficulties have moved from the place ∞ to the place p, namely, the representation π_p up to twist has zero measure in the Plancherel measure of the unitary dual of $\mathrm{PGL}_2(\mathbf{Q}_p)$. From an analytic point of view, these difficulties are quite similar. Using the trace formula, one can try to estimate (1) where now π_∞ or π_p respectively are now allowed to range over some class of unitary representations of positive measure (amounting to allowing the Laplace eigenvalue λ or the Hecke eigenvalue a_p to vary in an ε ball around $\lambda = 1/4$ or $a_p = 0$ respectively) and then try to control the error as ε becomes small. The (upper) bounds one obtains in this way typically (see the discussion before Theorem 1 in [Duk95]) have the shape $O(V/\log V)$ where V is the trivial bound (which in the case of eigenforms with $a_p = 0$ amounts to $V \asymp n$). Thus one is led to ask whether the extra arithmetic structure present when considering the latter question allows one to improve upon this analytic estimate (for which the answer is clearly yes).

1.2. **Analogies with a question of Greenberg.** Consider an irreducible modular Galois representation

$$\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

of weight $n \geq 2$. If f has CM by an imaginary quadratic field $F/\mathbf{Q}$ in which p splits, then the restriction of ρ_f to $G_{\mathbf{Q}_p}$ splits into a direct sum of characters. A well-known open question (attributed to Ralph Greenberg [GV04]) asks whether the converse holds:

Question 1.2.1. Suppose that ρ_f splits after restriction to $G_{\mathbf{Q}_p}$. Does f necessarily have CM?

Equivalently, is any *local* splitting of ρ_f due to a *global* splitting over some finite extension of $\mathbf{Q}$? The condition that $a_p = 0$ for a modular form of level prime

to p turns out to imply (see Theorem 2.1.1 below) that the representation ρ_f is *induced* after restriction to $G_{\mathbf{Q}_p}$. If f has CM by an imaginary quadratic field $F/\mathbf{Q}$ in which p is inert, then ρ_f is globally induced from F and the restriction of ρ_f to $G_{\mathbf{Q}_p}$ is induced from the unramified quadratic extension $K/\mathbf{Q}_p$. Hence the problem we are considering is the analogue of Question 1.2.1 when *split* is replaced by *induced*. The analogy here is not perfect, however, since (as noted above) non-CM elliptic curves with supersingular reduction provide a negative answer to this question in general whereas the answer to Greenberg's question is expected to always be positive. Note that in addition to supersingular elliptic curves there are other isolated counterexamples, including

$$q + 4q^3 - 8q^4 - 5q^5 - 22q^6 - 11q^9 + \ldots \in S_4(\Gamma_0(95), \mathbf{Q}),$$

$$q - 2q^2 + 4q^4 + 2q^5 - 7q^7 - 8q^8 - 27q^9 + \ldots \in S_4(\Gamma_0(154), \mathbf{Q})$$

with $a_2 = 0$ and $a_3 = 0$ respectively. The listed forms are identified by their labels 95.4.a.a and 154.4.a.b in the L-functions and modular forms database [LMF20].

1.3. **Some preliminaries on Group representations.** We recall some standard facts about Galois representations. Let

$$\rho : \Gamma \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$$

be any continuous homomorphism of a compact group Γ (such as a local or global Galois group with its natural topology). As explained in [Ski09, §2], ρ takes values in $\mathrm{GL}_n(E)$ for some finite extension $E/\mathbf{Q}_p$ which comes with a fixed embedding $E \hookrightarrow \overline{\mathbf{Q}}_p$, and thus also fixed embeddings $\mathcal{O}_E \hookrightarrow \overline{\mathbf{Z}}_p$ and $k_E = \mathcal{O}_E/\pi_E \hookrightarrow \overline{\mathbf{F}}_p$. Since Γ is compact, the image of ρ preserves a lattice $\Lambda \subset E^n$ and thus gives rise to a representation

$$\rho_\Lambda : \Gamma \to \mathrm{Aut}(\Lambda) \simeq \mathrm{GL}_n(\mathcal{O}_E).$$

We obtain from this a corresponding residual representation

$$\overline{\rho}_\Lambda : \Gamma \to \mathrm{GL}_n(k_E) \subset \mathrm{GL}_n(\overline{\mathbf{F}}_p).$$

The representation $\overline{\rho}_\Lambda$ may depend on Λ. On the other hand, by the Brauer–Nesbitt theorem, the semisimplification of $\overline{\rho}_\Lambda$ does not depend on any choices [Böc13, §3.2]. We denote the correspond semisimple representation $\Gamma \to \mathrm{GL}_n(\overline{\mathbf{F}}_p)$ by $\overline{\rho}$. If $k \subset \overline{\mathbf{F}}_p$ is any field such that $\overline{\rho}$ is valued in $\mathrm{GL}_n(k)$, we say that k is a *coefficient field* for $\overline{\rho}$. By abuse of notation, we shall also let $\overline{\rho}$ denote the representation of Γ to $\mathrm{GL}_n(k)$ for any coefficient field k.

We now prove a few straightfoward group theory lemmas to be used in the sequel.

Lemma 1.3.1. *If H is an index 2 subgroup of G and χ is a character of H, then $\mathrm{Ind}_H^G \chi$ is reducible if and only if χ extends to a character of G.*

Proof. If $V = \mathrm{Ind}_H^G \chi$ is reducible, then $V \simeq \chi_1 \oplus \chi_2$ for two characters χ_i of G. Restricting to H, we deduce that $\chi_i|_H = \chi$ for at least one i, and hence χ extends to a character of G. Conversely, if χ extends to a character of G (which by abuse of notation we also denote by χ), then $V \simeq \chi \oplus (\chi \otimes \eta_{G/H})$ where $\eta_{G/H}$ is the the non-trivial quadratic character of G with kernel H. In particular, V is reducible. □

Lemma 1.3.2. *Suppose G acts irreducibly on a 2-dimensional vector space V over $\overline{\mathbf{F}}_p$, but this action becomes reducible after restriction to a normal subgroup H*

such that G/H is cyclic of order prime to p. Then the projective image of G is dihedral, and the representation V is induced from an index two subgroup. Moreover, either the index two subgroup is unique and contains H, or the projective image of G is $D_4 \simeq (\mathbf{Z}/2\mathbf{Z})^2$ and V is induced from exactly three index two subgroups of G, precisely one of which contains H.

Proof. We first claim that $V|_H$ is completely decomposable. Let $V' \subset V$ be a one-dimensional H-stable submodule. Since V is irreducible as a G-module, there exists $g \in G$ such that $gV' \neq V'$. But since H is normal in G, it follows that gV' is also an H-stable submodule of V and hence there is a decomposition $V \simeq V' \oplus gV'$ as H-modules. We deduce that the representation V restricted to H is the direct sum of two characters, which implies that the projective representation of H associated to V has cyclic image (given by the image of the ratio of the characters) and has order prime to p. The assumption that G/H is cyclic of order prime to p then implies that the image of G in $\mathrm{PGL}_2(\overline{\mathbf{F}}_p)$ is metacyclic of order prime to p. The finite subgroups of $\mathrm{PGL}_2(\overline{\mathbf{F}}_p)$ of order prime to p consist of the groups $\mathbf{Z}/r$, D_{2r}, A_4, S_4, and A_5 (see for example [Ser72, §2.5]). The latter three groups are not metacyclic. If the projective image of G is cyclic, then the image Γ of G on V has the property that $\Gamma/Z(\Gamma)$ is cyclic. But $\Gamma/Z(\Gamma)$ is cyclic only if it is trivial, which implies that Γ is abelian and acts reducibly on V. Thus the projective image of G is the dihedral group D_{2r} for some $r \geq 2$. If V is induced from more than one index two subgroup, then the projective image of G must contain at least two cyclic subgroups of index two, and for D_{2r} and $r \geq 2$ this happens only for $r = 2$, when there are precisely three such subgroups. Taken together, this proves the lemma. □

Lemma 1.3.3. *Let $(A, \mathfrak{m})$ be a local Artinian ring with residue field k. Let V_A be an $A[G]$-module which is free of rank 2 over A. Assume that there exists a decomposition $V_A/\mathfrak{m} =: V_k \simeq U_k \oplus U_k'$ of G-modules where $\dim U_k = \dim U_k' = 1$ and U_k is* not *isomorphic to U_k'. Then any G-equivariant decomposition:*

$$V_A \simeq U_A \oplus U_A'$$

is unique, and — possibly swapping the factors — $U_A/\mathfrak{m} = U_k$ and $U_A'/\mathfrak{m}' = U_k'$.

Proof. If U_A is any $A[G]$-module which is free of rank one as an A-module, then all the Jordan–Hölder factors of U_A are isomorphic to $U_A/\mathfrak{m}$. Assume that there exists two decompositions $V_A \simeq U_A \oplus U_A'$ and $V_A \simeq T_A \oplus T_A'$. Since U_k is not isomorphic to U_k', the decomposition $V_k = U_k \oplus U_k'$ is unique and thus (up to reordering) we may assume that $U_A/\mathfrak{m} = U_k$ and also $T_A/\mathfrak{m} = U_k$. If $T_A \neq U_A$, then $I := (T_A + U_A) \cap U_A'$ must be non-trivial. Viewing I as a subspace of $T_A + U_A$ (which is a quotient of $T_A \oplus U_A$) and U_A' respectively, we deduce that all the Jordan–Hölder factors of I are all both isomorphic to U_k and isomorphic to U_k', a contradiction if I is non-trivial. Hence $T_A = U_A$ and similarly $T_A' = U_A'$. □

1.4. **Preliminaries on Galois deformations.** If k is a finite field, we let $W(k)$ denote the Witt vectors of k, and

$$\langle \cdot \rangle : k^\times \to W(k)^\times$$

the Teichmüller map. Let $\mathscr{C}$ denote the category of local Artinian $W(k)$-algebras $(A, \mathfrak{m})$ with a fixed identification $A/\mathfrak{m} = k$.

Throughout this paper, we shall often consider 2-dimensional representations $V = V_k$ of a group Γ over a field k.

Definition 1.4.1. Assume that V_k is absolutely irreducible as a representation of Γ. A deformation V_A of V_k to $A \in \mathscr{C}$ consists of an $A[\Gamma]$-module V_A free over A such that $V_A \otimes_A A/\mathfrak{m} \simeq V_k$. Two deformations are equivalent if they are isomorphic as $A[\Gamma]$-modules.

We now explain why this notion of deformation coincides with the alternate description in terms of matrix representations in [Maz89]. A fixed choice of basis for V_k gives rise to a representation: $\overline{\rho} : \Gamma \to \mathrm{GL}_2(k)$. A choice of basis for V_A lifting the given choices of basis for V_k gives rise to a representation

$$\rho : \Gamma \to \mathrm{GL}_2(A)$$

such that the corresponding residual representation is isomorphic to $\overline{\rho}$, and any such ρ gives rise to a module V_A. If ρ and ρ' are two representations such that the underlying modules V_A and V'_A are isomorphic, then they are conjugate by some matrix M. Since we are assuming V_k is absolutely irreducible, it follows by Schur's lemma that this matrix must be scalar modulo $\mathfrak{m}$. Hence (after scaling) we may assume $M \in I + \mathfrak{m}M_2(A)$, which is the usual notion of strict equivalence in [Maz89].

The deformation functors $D : \mathscr{C} \to \mathrm{Sets}$ we consider will all be pro-represented by complete local Noetherian rings. We shall assume basic familiarity with the further theory of Galois deformations as contained in [Maz89, Gou01].

2. The Argument

2.1. **Local consequences of the condition $a_p(f) = 0$.** Suppose we have a cuspidal modular eigenform $f \in S_n(\Gamma_1(N), \overline{\mathbf{Z}}_p)$, and let

$$\rho_f : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

denote the corresponding Galois representation. By the main theorem of [Sch90, Sai97], the p-adic representation $\rho_f|_{G_{\mathbf{Q}_p}}$ is crystalline, and the characteristic polynomial of crystalline Frobenius is $x^2 - a_p(f)x + p^{n-1}\chi(p)$, where χ is the Nebentypus character of f. For irreducible 2-dimensional crystalline representations of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, the characteristic polynomial of crystalline Frobenius is enough to determine the representation uniquely by [Bre03, Prop.3.1.1]. When $a_p(f) = 0$, there is a very simple description of the corresponding local Galois representation which we now describe. Let $K/\mathbf{Q}_p$ denote the unique unramified quadratic extension. By local class field theory, there is a unique character $K^\times \to G_K^{\mathrm{ab}} \to K^\times$ which sends p to 1 and $z \in \mathcal{O}_K^\times$ to z. With respect to the two embeddings of K into $\overline{\mathbf{Q}}_p$, this gives two characters ε_2, ε_2' from G_K to $\overline{\mathbf{Q}}_p^\times$ which are permuted by the action of $\mathrm{Gal}(K/\mathbf{Q}_p)$. We have the following result which follows from [Bre03, Prop 3.1.2]:

Theorem 2.1.1 (Breuil)**.** *Suppose that $a_p(f) = 0$. Then*

$$\rho_f|_{G_{\mathbf{Q}_p}} = \left(\mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \varepsilon_2^{n-1}\right) \otimes \psi$$

for some unramified character ψ with $\psi^2 = \chi|_{G_{\mathbf{Q}_p}}$.

Note that ψ is *a priori* only uniquely defined up to the unramified quadratic character $\eta_{K/\mathbf{Q}_p}$, but since $\rho_f|_{G_{\mathbf{Q}_p}} \otimes \eta_{K/\mathbf{Q}_p} \simeq \rho_f|_{G_{\mathbf{Q}_p}}$, either choice is correct.

Following §1.3, the characters $\overline{\varepsilon}_2$ and $\overline{\varepsilon}_2'$ from G_K to $\overline{\mathbf{F}}_p^\times$ correspond by local class field theory to the maps $K^\times \to G_K^{\mathrm{ab}} \to k_K^\times$ sending p to 1 and $z \in \mathcal{O}_K^\times$ to $z \bmod p$ composed with the two embeddings of $k_K \simeq \mathbf{F}_{p^2}$ into $\overline{\mathbf{F}}_p$. Moreover, we also have a corresponding identification

$$\overline{\rho}_f\big|_{G_{\mathbf{Q}_p}} = \left(\mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \overline{\varepsilon}_2^{n-1}\right) \otimes \overline{\psi} \tag{2}$$

For any $\overline{\rho}_f$, we fix a coefficient field $k = k_E \subset \overline{\mathbf{F}}_p$ as above such that $\overline{\rho}_f$ is valued in $\mathrm{GL}_2(k)$. For the remainder of §2 (with the exception of of §2.6) we assume that $p > 2$. We now show that $\overline{\rho}_f$ is absolutely irreducible after restriction to $G_{\mathbf{Q}_p}$.

Lemma 2.1.2. *Assume that $a_p(f) = 0$.*

1. *The representation $\overline{\rho}_f|_{G_{\mathbf{Q}_p}}$ is absolutely irreducible if $n-1$ is not divisible by $(p+1)$, and in particular absolutely irreducible whenever n is even.*
2. *The representation $\overline{\rho}_f|_{G_{\mathbf{Q}(\zeta_p)}}$ is absolutely irreducible if $n-1$ is not divisible by $(p+1)/2$. In particular, if $\overline{\rho}_f|_{G_{\mathbf{Q}(\zeta_p)}}$ is reducible and n is even, then $p \equiv 1 \bmod 4$ and $(n-1) \equiv (p+1)/2 \bmod (p+1)$.*

Proof. The representation $\overline{\rho}_f|_{G_{\mathbf{Q}_p}}$ is induced from a character of an index 2 subgroup. Thus, by Lemma 1.3.1, to prove part (1) it suffices (via the description of $\overline{\rho}_f|_{G_{\mathbf{Q}_p}}$ in equation (2)) to show that the character $\overline{\varepsilon}_2^{n-1}$ of G_K does not extend to $G_{\mathbf{Q}_p}$. Assume otherwise. Then $\overline{\varepsilon}_2^{n-1}$ coincides with its $\mathrm{Gal}(K/\mathbf{Q}_p)$ conjugate. By local class field theory, the action of $\mathrm{Gal}(K/\mathbf{Q}_p)$ on G_K^{ab} coincides with the action of $\mathrm{Gal}(K/\mathbf{Q}_p)$ on $K^\times$ under the Artin map. Since the non-trivial element of $\mathrm{Gal}(K/\mathbf{Q}_p)$ acts on the residue field k of K by Frobenius, it follows that the conjugate of $\overline{\varepsilon}_2$ is $\overline{\varepsilon}_2' = \overline{\varepsilon}_2^p$. Thus we may assume that $\overline{\varepsilon}_2^{n-1} = \overline{\varepsilon}_2^{p(n-1)}$ and hence

$$\overline{\varepsilon}_2^{(p-1)(n-1)} = 1.$$

Since $\overline{\varepsilon}_2$ has order $|k^\times| = p^2 - 1$, this forces $(n-1)$ to be divisible by $(p+1)$. This proves part (1).

Now suppose that $\overline{\rho}_f$ is irreducible as a representation of $G = G_{\mathbf{Q}_p}$ but becomes reducible over $H = G_{\mathbf{Q}_p(\zeta_p)}$. Since $G/H = \mathrm{Gal}(\mathbf{Q}_p(\zeta_p)/\mathbf{Q}_p)$ is cyclic of order prime to p, it follows by Lemma 1.3.2 that $\overline{\rho}_f|_{G_{\mathbf{Q}_p}}$ is induced from an index two subgroup of $G_{\mathbf{Q}_p}$ containing $G_{\mathbf{Q}_p(\zeta_p)}$. In particular, it must be induced from G_L where $L/\mathbf{Q}_p$ is the ramified quadratic extension inside $\mathbf{Q}_p(\zeta_p)$. On the other hand, $\overline{\rho}_f|_{G_{\mathbf{Q}_p}}$ is also induced from the unramified extension $K/\mathbf{Q}_p$, and so $\overline{\rho}|_{G_{\mathbf{Q}_p}}$ is irreducible and induced from at least two distinct quadratic fields. By Lemma 1.3.2, it follows that the projective image of $G_{\mathbf{Q}_p}$ is isomorphic to $(\mathbf{Z}/2\mathbf{Z})^2$, and that the projective image restricted to G_K has order 2. Since the representation $\overline{\rho}_f$ restricted to G_K is (up to twist) the direct sum $\overline{\varepsilon}_2^{n-1} \oplus \overline{\varepsilon}_2^{p(n-1)}$, it follows that the ratio of the characters $\overline{\varepsilon}_2^{n-1}$ and $\overline{\varepsilon}_2^{p(n-1)}$ is quadratic, or equivalently that

$$\overline{\varepsilon}_2^{2(p-1)(n-1)} = 1.$$

It follows that $(n-1)$ is divisible by $(p+1)/2$, which proves part (2) after noting that $(p+1)/2$ is even if $p \equiv 3 \bmod 4$. □

The proof of Lemma 2.1.2 is the only place where we use the assumption that the weight n is even. The proof otherwise proceeds without any further changes required under the weaker assumption that $(n-1)$ is not divisible by $(p+1)$ and $(n-1)$ is divisible by $(p+1)/2$ only if $p \equiv 1 \mod 4$.

2.2. **Some reductions.** By a theorem of Jochnowitz [Joc82], there are only finitely many irreducible modular residual representations of level N. By class field theory, there are only finitely many Nebentypus characters $\chi : (\mathbf{Z}/N\mathbf{Z})^\times \to \overline{\mathbf{Z}}_p^\times$ of level N. Moreover, if

$$\langle\overline{\chi}\rangle : (\mathbf{Z}/N\mathbf{Z})^\times \to k^\times \to W(k)^\times \to \overline{\mathbf{Z}}_p^\times$$

is the Teichmüller lift of the mod $\mathfrak{m}_{\overline{\mathbf{Z}}_p}$ reduction of χ, then $\chi/\langle\overline{\chi}\rangle$ is valued in $1+\mathfrak{m}_{\overline{\mathbf{Z}}_p}$. Any such character is the square of a unique character valued in $1+\mathfrak{m}_{\overline{\mathbf{Z}}_p}$, because $\sqrt{1+x} = 1 + x/2 + \ldots$ converges in $\overline{\mathbf{Z}}_p$ for $p > 2$. Thus, after a finite global twist, we may assume that the Nebentypus characters of f are fixed and valued in $W(k)^\times$.

In particular, to prove Theorem 1.0.1, we may assume the following:

Assumption 2.2.1. There exist an infinite number of cuspidal eigenforms f of level $\Gamma_1(N)$ and even weight n with $a_p(f) = 0$ satisfying the following further assumptions:

(1) All such f have the same fixed residual representation
$$\overline{\rho} = \overline{\rho}_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(k).$$
(2) There is an isomorphism
$$\overline{\rho}|_{G_{\mathbf{Q}_p}} = \left(\mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \overline{\varepsilon}_2^{n-1}\right) \otimes \overline{\psi}$$
for fixed n and fixed $\overline{\psi} : G_{\mathbf{Q}_p} \to k^\times$.
(3) If ψ denotes the character $\langle\overline{\psi}\rangle : G_{\mathbf{Q}_p} \to W(k)^\times$, then the Nebentypus character χ of f restricted to $G_{\mathbf{Q}_p}$ is equal to ψ^2. By Theorem 2.1.1, this is equivalent to saying that the representation $\det(\overline{\rho}_f)|_{G_{\mathbf{Q}_p}^{\mathrm{ab}}}$ evaluated at p considered as an element of $\mathbf{Q}_p^\times \to G_{\mathbf{Q}_p}^{\mathrm{ab}}$ via the Artin map is $\psi^2(p)$.

2.3. **Local deformation rings.** In this section, we define some local deformation rings associated to $\overline{\rho}$.

Recall that we have fixed a coefficient field k for $\overline{\rho}$. After increasing k if necessary, we may assume that the eigenvalue of any element in the image of $\overline{\rho}$ lands in k, and moreover that $\mathcal{O}_K \subseteq W(k)$, where $W(k)$ is the Witt vectors of k.

Associated to $\overline{\rho}|_{G_{\mathbf{Q}_p}}$ is a local universal deformation ring R^{loc} which is a complete local Noetherian $W(k)$-algebra (it represents the functor D recalled below). We now construct a quotient of this ring corresponding to deformations which are "induced" from K. Let V_k denote the underlying representation of $\overline{\rho}$ over k. After restricting to G_K, there is a canonical splitting $V_k = U_k \oplus U_k'$ such that G_K acts on U_k and U_k' by $\overline{\varepsilon}_2^{n-1} \otimes \overline{\psi}$ and $(\overline{\varepsilon}_2')^{n-1} \otimes \overline{\psi}$ respectively, and U_k is not isomorphic to U_k'.

Definition 2.3.1 (Locally induced deformations)**.** For a local Artinian $W(k)$-algebra $(A, \mathfrak{m}) \in \mathscr{C}$, let $D^{\mathrm{loc}}(A)$ denote the deformations V_A of V_k to A. Let $D^{\mathrm{ind}}(A)$ denote the subset of deformations V_A which admit a splitting $V_A = U_A \oplus U_A'$ into free A-modules of rank one such that U_A and U_A' are G_K-modules which are G_K-deformations of U_k and U_k' respectively.

Let $D^{\mathrm{ind},\psi}(A)$ denote the subset of $D^{\mathrm{ind}}(A)$ such that the action of $p \in K^\times \to G_K^{\mathrm{ab}}$ on U_A is given by $\psi(p)$, where $\psi : G_{\mathbf{Q}_p} \to W(k)^\times$ is the Teichmüller lift of $\overline{\psi}$ as in Assumption 2.2.1(3).

Finally, let D_U denote the deformations U_A of U_k.

Let $\mathcal{O}_K(p) = 1 + \mathfrak{m}_K$ denote the units in $K^\times$ which are $1 \pmod p$. Since $p > 2$, the exponential map gives an isomorphism $\exp : \mathfrak{m}_K \to \mathcal{O}_K(p)$. The group $\mathfrak{m}_K$ may be further identified as a topological group with $(\mathbf{Z}_p)^2$ via an arbitrary choice of a basis for $\mathfrak{m}_K$ over $\mathbf{Z}_p$.

Lemma 2.3.2. *D^{ind} is pro-represented by a complete $W(k)$-algebra $R^{\mathrm{loc,ind}}$ which is isomorphic to the universal deformation ring of*

$$\overline{\varepsilon}_2^{n-1} \otimes \overline{\psi} : G_K \to k^\times.$$

In particular, $R^{\mathrm{loc,ind}} \simeq W(k)[\![\mathcal{O}_K(p) \oplus \mathbf{Z}_p]\!]$ is smooth of relative dimension 3 over $W(k)$. $D^{\mathrm{ind},\psi}$ is pro-represented by the quotient $R^{\mathrm{loc,ind},\psi} \simeq W(k)[\![\mathcal{O}_K(p)]\!]$ of $R^{\mathrm{loc,ind}}$.

Proof. It suffices to identify the functors D^{ind} and D_U. Given an element of D^{ind}, there exists a corresponding free A-submodule $U_A \subset V_A$ deforming U_k by definition, and a decomposition $V_A = U_A \oplus U_A'$. This decomposition is unique by Lemma 1.3.3. Conversely, given a deformation U_A of U_k, then $V_A = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}}(U_A)$ gives an element of $D^{\mathrm{ind}}(A)$, and this gives the desired identification (one easily checks that these maps are mutual inverses). Hence the two deformation functors coincide.

Let Γ_K denote the Galois group of the maximal pro-p abelian extension of K. By local class field theory, we have $G_K^{\mathrm{ab}} \simeq \widehat{K^\times} \simeq \mathcal{O}_K(p) \oplus k_K^\times \oplus \widehat{\mathbf{Z}}$ and then (since $p > 2$) we have an isomorphism $\Gamma_K \simeq \mathcal{O}_K(p) \oplus \mathbf{Z}_p$. It follows that the 1-dimensional deformation ring associated to any character $\overline{\chi} : G_K \to k^\times$ is isomorphic to

$$W(k)[\![\Gamma_K]\!] \simeq W(k)[\![\mathcal{O}_K(p) \oplus \mathbf{Z}_p]\!],$$

where the $\mathbf{Z}_p$ factor on the right hand side is topologically generated by $p \in K^\times$. The corresponding Galois representation

$$K^\times \to G_K^{\mathrm{ab}} \to W(k)[\![\Gamma_K]\!]^\times$$

is given explicitly as follows:

$$x \mapsto \langle \overline{\chi}(x) \rangle [\varphi(x)], \tag{3}$$

where $\varphi(x) \in \Gamma_K$ is the image of x under the natural map $K^\times \to G_K^{\mathrm{ab}} \to \Gamma_K$. In particular, when $\overline{\chi} = \overline{\varepsilon}_2^{n-1}\overline{\psi}$, we have the desired isomorphism (noting that $\overline{\varepsilon}_2(p) = 1$ and ψ is the Teichmüller lift of $\overline{\psi}$)

$$R^{\mathrm{loc,ind},\psi} \simeq R^{\mathrm{loc,ind}}/(\langle \overline{\psi}(p) \rangle [\varphi(p)] - \psi(p)) \simeq R^{\mathrm{loc,ind}}/([\varphi(p)] - 1) \simeq W(k)[\![\mathcal{O}_K(p)]\!].$$

□

We now define local deformation rings $R^{\mathrm{loc,unr}}$ and $R^{\mathrm{loc,split}}$.

Definition 2.3.3 (Unramified and split local deformation rings)**.** Let $M/\mathbf{Q}_p$ denote the fixed field of $\ker(\overline{\rho}_f|_{G_{\mathbf{Q}_p}})$. Let M^{unr} denote the maximal unramified extension of M. We define subfunctors D^{unr} and D^{split} of D as follows:

(1) $D^{\mathrm{unr}}(A) \subset D^{\mathrm{loc}}(A)$ consists of deformations V_A such that the action of $G_{\mathbf{Q}_p}$ on V_A factors through $\mathrm{Gal}(M^{\mathrm{unr}}/\mathbf{Q}_p)$.

(2) $D^{\mathrm{split}}(A) \subset D^{\mathrm{unr}}(A) \subset D^{\mathrm{loc}}(A)$ consists of deformations V_A such that the action of $G_{\mathbf{Q}_p}$ on V_A factors through $\mathrm{Gal}(M/\mathbf{Q}_p)$.

Let $D^{\mathrm{unr},\psi}(A)$ and $D^{\mathrm{split},\psi}(A)$ denote the subsets of D^{unr} and D^{split} respectively such that the action of $p \in K^\times \to G_K^{\mathrm{ab}}$ on $\wedge^2 V_A$ is given by $\psi^2(p)$. (Equivalently, the determinant character of V_A evaluated on p is $\psi^2(p)$.)

Lemma 2.3.4. *The functor D^{unr} is a subfunctor of D^{ind}, and the functor $D^{\mathrm{unr},\psi}$ is a subfunctor of $D^{\mathrm{ind},\psi}$. The functors D^{unr}, D^{split} and $D^{\mathrm{unr},\psi}$, $D^{\mathrm{split},\psi}$ are pro-representable by quotients of $R^{\mathrm{loc,ind}}$ and $R^{\mathrm{loc,ind},\psi}$ respectively. There are isomorphisms and surjections as follows:*

$$\begin{array}{ccc}
W(k)[\![\mathcal{O}_K(p) \oplus \mathbf{Z}_p]\!] \simeq R^{\mathrm{loc,ind}} & \longrightarrow & R^{\mathrm{loc,ind},\psi} \simeq W(k)[\![\mathcal{O}_K(p)]\!] \\
\downarrow & & \downarrow \\
W(k)[\![\mathbf{Z}_p]\!] \simeq R^{\mathrm{loc,unr}} & \longrightarrow & R^{\mathrm{loc,unr},\psi} \simeq W(k) \\
\downarrow & & \| \\
W(k) \simeq R^{\mathrm{loc,split}} & = & R^{\mathrm{loc,split},\psi} \simeq W(k)
\end{array}$$

where the maps in the first column send all elements of $\mathcal{O}_K(p)$ to 1 for the first map and all elements of $\mathbf{Z}_p$ to 1 for the second.

Proof. We start by proving that $D^{\mathrm{unr}}(A) \subset D^{\mathrm{ind}}(A)$. Let $G = \mathrm{Gal}(M^{\mathrm{unr}}/\mathbf{Q}_p)$, let $H = \mathrm{Gal}(M^{\mathrm{unr}}/K)$, and let $I \subset H \subset G$ denote the inertia group. There is an isomorphism $M^{\mathrm{unr}} = M.\mathbf{Q}_p^{\mathrm{unr}}$. We deduce that I is cyclic of order prime to p and H is abelian. It follows that the action of I on V_A is diagonalizable. Since the action of I on V_k decomposes as the direct sum $U_k \oplus U_k'$ of distinct characters, it follows that the action of I on V_A decomposes as $U_A \oplus U_A'$ where U_A and U_A' are free rank one A-modules which reduce to U_k and U_k' modulo $\mathfrak{m}$ respectively. Since H is abelian and contains I, it follows that hU_A for $h \in H$ is also a free rank one A-module which is preserved by I, and thus $V_A = hU_A \oplus hU_A'$. By Lemma 1.3.3, we deduce that $hU_A = U_A$ or U_A', and the former is ruled out by noting that $hU_k = U_k$. Hence the decomposition $V_A \simeq U_A \oplus U_A'$ extends to a decomposition of H-modules. Since the image of G_K in G is H, this says precisely that $V_A \in D^{\mathrm{ind}}(A)$.

Having shown that $R^{\mathrm{loc,unr}}$ is a quotient of $R^{\mathrm{loc,ind}}$, we can reinterpret the functors D^{unr}, D^{split} in terms of the deformations of the one dimensional character $\overline{\varepsilon}_2^{n-1} \otimes \overline{\psi} : G_K \to k^\times$. The unramified lifts correspond to deformations of the unramified character times the Teichmüller lift of the ramified character, and the split lift corresponds to the Teichmüller lift of the entire representation. Since the action of $\mathrm{Gal}(K/\mathbf{Q}_p)$ on $K^\times$ fixes p, the action of $p \in K^\times \to G_K^{\mathrm{ab}}$ on U_A for any element of $D^{\mathrm{ind}}(A)$ coincides with the action of p on U_A', and thus the subsets $D^{\mathrm{unr},\psi}(A) \subset D^{\mathrm{unr}}(A)$ and $D^{\mathrm{split},\psi}(A) \subset D^{\mathrm{split}}(A)$ are given by imposing the condition that the image of p of any deformation of our character acts by $\psi(p)$. The explicit descriptions of these rings can then be read off from the explicit description of R^{ind} in Lemma 2.3.2 and from equation (3). □

2.4. **Global deformation rings.** We now consider some global deformation rings associated to $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$.

Definition 2.4.1. Let $D(A)$ denote the deformations of V_k which are unramified outside Np.

The functor $D(A)$ is represented by the universal global deformation ring R^{glob}. The ring R^{glob} is an R^{loc}-algebra by Yoneda's lemma. This allows us to define the key deformation ring R of interest, together with auxiliary rings R^{unr} and R^{split}.

Definition 2.4.2. Let $R = R^{\mathrm{glob}} \otimes_{R^{\mathrm{loc}}} R^{\mathrm{loc,ind},\psi}$, let $R^{\mathrm{unr}} = R^{\mathrm{glob}} \otimes_{R^{\mathrm{loc}}} R^{\mathrm{loc,unr}}$, and let $R^{\mathrm{split}} \simeq R^{\mathrm{glob}} \otimes_{R^{\mathrm{loc}}} R^{\mathrm{loc,split}}$.

The ring R represents the functor $D^{\mathrm{glob,ind},\psi}(A)$ of deformations V_A unramified outside Np such that $V_A|_{G_{\mathbf{Q}_p}} \in D^{\mathrm{ind},\psi}(A)$. Similarly (following Lemma 2.3.4) the rings R^{unr} and R^{split} represent the subfunctor of $D^{\mathrm{glob,ind}}(A)$ for which the corresponding representations locally factor through $\mathrm{Gal}(M^{\mathrm{unr}}/\mathbf{Q}_p)$ and $\mathrm{Gal}(M/\mathbf{Q}_p)$ respectively, where M is the fixed field of $\overline{\rho}|_{G_{\mathbf{Q}_p}}$. Because $R^{\mathrm{split,unr},\psi} \simeq R^{\mathrm{split},\psi}$ by Lemma 2.3.4, the ring R^{split} is a quotient of R. On the other hand, R^{unr} need not be a quotient of R. The construction of R guarantees that all of our eigenforms (satisfying Assumption 2.2.1) give rise to $\overline{\mathbf{Q}}_p$-valued points of R.

The last global deformation ring we consider parametrizes deformations of $\overline{\rho}$ which are *globally* induced from a quadratic field F. If $\overline{\rho}$ is induced from G_F, then there is a decomposition $V_k \simeq U_k \oplus U_k'$ as G_F-modules. Since $G_K \subset G_F$, this decomposition must coincide with the unique such decomposition of G_K-modules for the unramified quadratic extension $K/\mathbf{Q}_p$.

Definition 2.4.3 (CM deformation rings)**.** Suppose that $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ is induced from a quadratic field $F/\mathbf{Q}$. Let $D^{\mathrm{CM},F}(A)$ denote the subset of deformations $D(A)$ of V_A which admit a splitting $V_A = U_A \oplus U_A'$ into free A-modules of rank one such that U_A and U_A' are G_F-modules which are G_F-deformations of U_k and U_k' respectively, and such that the image of $p \in \mathbf{Q}_p^\times \to G_{\mathbf{Q}_p}^{\mathrm{ab}}$ on $\wedge^2 V_A$ is given by $\psi^2(p)$.

Lemma 2.4.4. *There exist at most three quadratic fields $F/\mathbf{Q}$ such that $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ is induced from F. For any such F, the functor $D^{\mathrm{CM},F}(A) \subset D(A)$ is pro-representable by a ring R_F^{CM} which is a quotient of R.*

Proof. The first claim follows from Lemma 1.3.2. The second claim follows from the fact that any element of $D^{\mathrm{CM},F}(A)$ gives an element of $D^{\mathrm{ind},\psi}(A)$ by restriction to $G_K \subset G_F$. □

Remark 2.4.5. If $\overline{\rho}$ is induced from a quadratic extension $F/\mathbf{Q}$, the field $F/\mathbf{Q}$ is either real or imaginary. There is some abuse of notation to call R_F^{CM} a CM deformation ring when F is real. This will not cause any issues since we use only the fact that all CM deformations of $\overline{\rho}$ of level N lie on R_F^{CM} for some F (which will be imaginary). When $F/\mathbf{Q}$ is a real quadratic field, the maximal pro-p extension of F unramified outside Np is the compositum of the $\mathbf{Z}_p$-cyclotomic extension with a finite extension of F, and so $R_F^{\mathrm{CM}}(\overline{\mathbf{Q}}_p)$ consists of a finite number of Artin representations up to twist.

2.5. **The ring R is small.** We have now constructed a deformation ring R which captures the eigenforms F with $a_p(f) = 0$. The first key step is establish a finiteness result for this ring.

Lemma 2.5.1. *R is finite as a module over $R^{\mathrm{loc,ind},\psi} = W(k)[\![\mathcal{O}_K(p)]\!]$.*

Proof. Let $\mathfrak{m}_{R^{\mathrm{loc,ind},\psi}}$ denote the maximal ideal of $R^{\mathrm{loc,ind},\psi}$. By Nakayama's lemma, to show that R is finite over $R^{\mathrm{loc,ind},\psi}$ it suffices to show that $R/\mathfrak{m}_{R^{\mathrm{loc,ind},\psi}}$ is finite over $R^{\mathrm{loc,ind},\psi}/\mathfrak{m}_{R^{\mathrm{loc,ind},\psi}} \simeq k$. Now by Lemma 2.3.4, we have isomorphisms

$$k = R^{\mathrm{loc,ind},\psi}/\mathfrak{m}_{R^{\mathrm{loc,ind},\psi}} \simeq R^{\mathrm{loc,split}}/\mathfrak{m}_{R^{\mathrm{loc,split}}},$$

and thus

$$R/\mathfrak{m}_{R^{\mathrm{loc,ind},\psi}} \simeq R^{\mathrm{split}}/\mathfrak{m}_{R^{\mathrm{loc,split}}}.$$

If R^{split} is finite over $R^{\mathrm{loc,split}} \simeq W(k)$, then $R^{\mathrm{split}}/\mathfrak{m}_{R^{\mathrm{loc,split}}}$ is certainly finite over $R^{\mathrm{loc,split}}/\mathfrak{m}_{R^{\mathrm{loc,split}}} \simeq k$, and we would be done. The finiteness of R^{split} follows immediately from [AC14, Thm.1(2)] under the additional Taylor–Wiles hypothesis that $\overline{\rho}|_{G_{\mathbf{Q}(\zeta_p)}}$ is absolutely irreducible. Indeed, that reference proves the stronger claim that R^{unr} is finite over $W(k)$, and R^{split} is a quotient of R^{unr}.

It suffices to consider the remaining case when the Taylor–Wiles hypothesis fails, or equivalently that $\overline{\rho}|_{G_{\mathbf{Q}(\zeta_p)}}$ is reducible. This certainly implies that $\overline{\rho}|_{G_{\mathbf{Q}_p(\zeta_p)}}$ is reducible, and hence, by Lemma 2.1.2, we may assume that $n-1 \equiv (p+1)/2 \mod (p+1)$ and that $p \equiv 1 \mod 4$. But now we may invoke Theorem A.0.1 of the appendix. □

Remark 2.5.2. An alternative approach to proving finiteness is to specialize to a height one prime ideal $\mathfrak{p}$ of $R^{\mathrm{loc,ind},\psi}$ corresponding to a representation of the form $\varepsilon_1^m \otimes (\mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \varepsilon_2^{k-1}) \otimes \psi$, where $2 \le k \le p-1$ and ε_1 is the cyclotomic character. In order for the corresponding residual representation to agree with $(\mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \overline{\varepsilon}_2^{n-1}) \otimes \overline{\psi}$, it suffices (using the identity $\overline{\varepsilon}_1 = \overline{\varepsilon}_2^{(p+1)}$) to chose m and k such that the following congruence is satisfied:

$$m(p+1) + k - 1 \equiv (n-1) \text{ or } (n-1)p \mod (p^2-1).$$

We may take $k \equiv n \mod (p+1)$ unless $n \equiv 0 \mod (p+1)$, and we may take $k \equiv (p+3-n) \equiv 2-n \mod (p+1)$ unless $n \equiv 2 \mod (p+1)$, and so m and k exist as long as $p \ge 5$. The corresponding deformation ring $R^{\mathrm{loc,ind},\psi}/\mathfrak{p}$ is a quotient of the crystalline local deformation ring with Hodge–Tate weights $[m, m+k-1]$, and is thus a twist of a crystalline deformation ring of weight $[0, k-1]$ which is in the Fontaine–Laffaille range. Since one expects to be able to prove $R = \mathbf{T}$ theorems in this context (exploiting the fact that the corresponding local deformation rings are Cohen–Macaulay), this leads to explicit bounds on $\dim_{\mathbf{Q}_p}(R/\mathfrak{p})[1/p]$ in terms of dimensions of spaces of modular forms of weight at most $p-1$, although in this approach one would also need to deal separately with the case when $\overline{\rho}|_{G(\mathbf{Q}(\zeta_p))}$ is reducible.

We now turn to the study of the finite $\Lambda := W(k)[\![\mathcal{O}_K(p)]\!]$-module R. The eigenforms with $a_p(f) = 0$ captured by R give rise to a map from R to $\overline{\mathbf{Q}}_p$ and thus a prime of R. Any such prime is contained inside a minimal prime of R, and since R is Noetherian, there are only finitely many minimal prime ideals, and thus we may assume that there are infinitely many non-CM points which lie inside a fixed minimal prime $\mathfrak{P}$, or equivalently lie on a fixed irreducible component $R/\mathfrak{P}$ of R.

Lemma 2.5.3. *Suppose that there are infinitely many modular Galois representations ρ_f of level dividing N giving rise to points of $R/\mathfrak{P}$. Then the support of $R/\mathfrak{P}$ is all of Λ.*

Proof. The ring $R/\mathfrak{P}$ is a finite Λ module by Lemma 2.5.1. The support of a finite module is closed, and thus it suffices to show that the support includes a Zariski dense subset of Λ. Since there are only finitely many Galois representations of any fixed weight, we may assume that $R/\mathfrak{P}$ has points ρ_f for modular eigenforms f of infinitely many different weights. If f has weight n, then we may explicitly write down the corresponding point of Λ. We make the explicit identification:

$$\mathcal{O}_K(p) \simeq (1+p)\mathbf{Z}_p \oplus (1+p)^\eta \mathbf{Z}_p,$$

where $\eta = \sqrt{u}$ for any fixed non-quadratic residue u. We may then take $X = [1+p]-1$ and $Y = [(1+p)^\eta]-1$. The classical modular forms we are considering all correspond to specializations where $z \in \mathcal{O}^\times$ maps to z^{n-1}, or equivalently to

$$X \mapsto (1+p)^{n-1} - 1, Y \mapsto (1+p)^{\eta(n-1)} - 1. \tag{4}$$

It suffices to show that any infinite collection of these points are Zariski dense in Λ. The problem is that the Zariski closure is *trying to be* given by the equation

$$H = \eta \log(1+X) - \log(1+Y) = 0, \tag{5}$$

but this is not an element of $\Lambda \otimes \mathbf{Q}_p$ because the denominators grow without bound. (Note that we have chosen the field k so that $\eta \in \mathcal{O}_K \subseteq W(k)$.) Alternatively, the Zariski closure wants to be $(1+X)^\eta - (1+Y) = 0$, although the corresponding formal power series (unlike H) doesn't even converge for all $|X|, |Y| < 1$ as we shall see shortly in Sublemma 1 below.

Suppose the Zariski closure of these points is given by the vanishing set of $F(X,Y)$ in $W(k)[\![X,Y]\!]$. Choose a primitive p^mth root of unity ζ_m for each m and let $\pi_m = 1-\zeta_m$. There is an inclusion

$$F(X,Y) \in W(k)[\![X,Y]\!] \subset W(k)[\pi_m][\![X,Y]\!] \subset W(k)[\pi_m][\![X/\pi_m, Y/\pi_m]\!],$$

which amounts to considering the restriction of functions bounded by 1 on the open unit ball $B(1)$ to functions bounded by 1 on the open ball $B(\pi_m)$. (Here $B(r)$ denotes the open ball centered at the origin with radius $|r|$.)

Sublemma 1. *Suppose that $v(\eta) = 0$ but $\eta \notin \mathbf{Z}_p$. Then*

$$H_m := (1+X)^{\eta p^{m-1}} - (1+Y)^{p^{m-1}}. \tag{6}$$

is an element of $W(k)[\pi_m, Y][\![X/\pi_m]\!] \subset W(k)[\pi_m][\![X/\pi_m, Y/\pi_m]\!]$, but is not an element of $W(k)[\pi_{m+1}][\![X/\pi_{m+1}, Y/\pi_{m+1}]\!] \otimes \mathbf{Q}_p$.

Proof. It suffices to analyze the growth of the coefficient of X^n as n varies. The coefficient is explicitly given by a binomial coefficient, and hence its valuation is

$$v\left(\binom{\eta p^{m-1}}{n}\right) = v\left(\prod_{i=0}^{n-1}(\eta p^{m-1} - i)\right) - v(n!)$$

Our assumptions on η imply that the valuation of $\eta p^{m-1} - i$ is $v(i)$ when $v(i) \leq m-1$ and $m-1$ otherwise. Hence the valuation of the X^n coefficient is equal to

$$\sum_{k=1}^{m-1}\left(1 + \left\lfloor \frac{n-1}{p^k} \right\rfloor\right) - \sum_{k=1}^{\infty}\left\lfloor \frac{n}{p^k} \right\rfloor. \tag{7}$$

A lower bound for this expression is given by

$$\sum_{k=1}^{m-1}\left\lfloor \frac{n}{p^k} \right\rfloor - \sum_{k=1}^{\infty}\left\lfloor \frac{n}{p^k} \right\rfloor = -\sum_{k=m}^{\infty}\left\lfloor \frac{n}{p^k} \right\rfloor \geq -\sum_{k=m}^{\infty}\frac{n}{p^k} = -\frac{n}{p^{m-1}(p-1)},$$

whereas an upper bound when $n = p^r$ and $r \geq m-1$ is given by

$$\sum_{k=1}^{m-1}\left(1+\left\lfloor\frac{p^r-1}{p^k}\right\rfloor\right)-\sum_{k=1}^{\infty}\left\lfloor\frac{p^r}{p^k}\right\rfloor=-\sum_{k=m}^{r}\frac{p^r}{p^k}=-\frac{n}{(p-1)p^{m-1}}\left(1-\frac{1}{p^{r-m-1}}\right).$$

Since $v(\pi_m) = 1/(p^{m-1}(p-1))$, the lower bound implies that $H_m \in W(k)[\pi_m, Y][\![X/\pi_m]\!]$. On the other hand, the upper bound shows that this cannot be improved. □

Now H_m also vanishes at all the weights $z \mapsto z^{n-1}$. We claim that H_m is irreducible in $W(k)[\pi_m][\![X/\pi_m, Y/\pi_m]\!]$. Viewing H_m as an element inside the larger ring $W(k)[\pi_m][\![X/\pi_1, Y/\pi_1]\!]$, we *do* have the factorization

$$H_m = \prod_{i=1}^{p^{m-1}} \left(1 + Y - (1+X)^{\eta}\zeta_{m-1}^{i}\right).$$

Certainly any factorization over the smaller ring is thus promoted to a factorization over $W(k)[\pi_m, Y][\![X/\pi_m]\!]$. But if there exists a factor with $r < p^{m-1}$ terms, then the constant term considered as a polynomial in $(Y+1)$ will be a non-zero multiple of $(1+X)^{\eta r}$, which does not lie in this ring by Sublemma 1, a contradiction.

Because $H_m \in W(k)[\pi_m][\![X/\pi_m, Y/\pi_m]\!]$ is irreducible, we deduce that $F(X,Y)$ must vanish at all points in $B(\pi_m)$ where H_m vanishes. But then $F(X,Y)$ must vanish at the finitely many pairs $(\xi_1 - 1, \xi_2 - 1)$ of p-power roots of unity with $v(\xi_i - 1) > v(\pi_m)$. But repeating this with m arbitrarily large implies that $F(X,Y)$ vanishes at all such pairs of p-power roots of unity, which is impossible because they are Zariski dense in Λ. □

We note in passing that (formally) $\lim_{m\to\infty} \dfrac{H_m(X,Y)}{p^{m-1}} = H(X,Y)$.

Lemma 2.5.4. *If a component $R/\mathfrak{P}$ has infinitely many points which correspond to modular Galois representations which are not CM, then the support of CM points on $R/\mathfrak{P}$ is either empty or lies on a proper closed subscheme of Λ.*

Proof. If there are no CM points the result is immediate. If $R/\mathfrak{P}$ has a single point which admits CM by $F/\mathbf{Q}$, then certainly $\overline{\rho}$ is induced from F. By Lemma 2.4.4, there are at most three such fields F, and it suffices to prove the lemma for any given F. In particular, the points with CM by F all give rise to points on R_F^{CM} and hence on the intersection $R/\mathfrak{P} \otimes_R R_F^{\mathrm{CM}}$. Either this is all of $R/\mathfrak{P}$, which contradicts the assumption, or, because $R/\mathfrak{P}$ is irreducible, it has positive co-dimension. But since R is finite over Λ, the dimension of R and any of its quotients coincides with the dimension of its support, and hence we are done. □

To complete the proof of Theorem 1.0.1, it suffices to show that if the support of $R/\mathfrak{P}$ is all of Λ, then $R/\mathfrak{P}$ contains a Zariski dense set of points which are CM. To this end, we use a variation of the idea of Ghate–Vatsal [GV04] to prove local indecomposability of non-CM Hida families by specializations in weight one. Consider points in Λ corresponding to maps $\mathcal{O}_K(p) \to \overline{\mathbf{Q}}_p^{\times}$ with finite image and such that the ratio of any such map to its $\mathrm{Gal}(K/\mathbf{Q}_p)$-conjugate has order greater than 60. These points are clearly Zariski dense in Λ. The assumption that $R/\mathfrak{P}$ has full support means that for any such specialization we obtain corresponding global Galois representations:

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

which locally at p has finite image on inertia at p. Thus, by [PS16, Thm.0.2] when $\overline{\rho}|_{G(\mathbf{Q}(\zeta_p))}$ is irreducible and by [Sas, Thm.1] otherwise (noting that we may assume that $p \equiv 1 \mod 4$ and hence $\overline{\rho}$ is induced from a real quadratic extension), it follows that ρ is modular of weight one. By construction, the image of inertia in the corresponding projective representation has order greater than 60 (given our choice of point of Λ), and hence the global projective representation also has order greater than 60. This ensures that the projective image is not of exceptional type (A_4, S_4, or A_5). It follows that ρ must be of dihedral type. Exactly as in [GV04], all but finitely many of these forms must additionally be of CM type. But then we have produced a Zariski dense set of CM points on $R/\mathfrak{P}$, a contradiction, and we are done.

Remark 2.5.5. The fact that any infinite set of characters of the form $z \mapsto z^n$ are Zariski dense in some irreducible component of $\Lambda = \mathrm{Spec}(\mathbf{Z}_p[\![\mathcal{O}_K^\times]\!])$ can be viewed as a special case of a local p-adic analogue of Lang's conjecture [Lan83] (See also [Ser18]). The classical analogue of our example is the statement that any infinite set of points $(\exp(\eta x), \exp(x))$ are Zariski dense in $(\mathbf{C}^\times)^2$ whenever $\eta \notin \mathbf{Q}$.

2.6. **Proof of Corollary 1.0.2.** If f is a form of level 1 which is CM, then the corresponding automorphic representation is induced from a character on some imaginary quadratic field $K/\mathbf{Q}$. But then the level of f will be divisible by any prime dividing the discriminant of K, a contradiction. Hence Corollary 1.0.2 follows immediately for all $p > 2$. For $p = 2$, we prove directly that if $a_2(f) = 0$ then ρ_f is dihedral, and so f is CM, from which the result follows by the argument above. Note that for $N = 1$ and $p = 2$ the representation $\overline{\rho}$ will have trivial semi-simplification (see [Che08, Lemme 1.7]), and it follows that the image of ρ_f factors through the maximal pro-2 extension of $\mathbf{Q}$ unramified outside 2. By [Che08, Prop 1.8], it follows that the image of ρ_f is isomorphic to the image of ρ_f restricted to inertia at 2. But the assumption that $a_2(f) = 0$ implies that ρ_f is locally induced, which now implies it is also globally induced, and thus CM.

3. Acknowledgments

This paper owes its genesis to two conversations between the authors during their respective number theory seminars, one in Wisconsin in February of 2018, and the second in Chicago in January of 2020. The first author would like to thank George Boxer from whom he learned the surprising fact that the characters $z \to z^n$ are Zariski dense in $\mathbf{Z}_p[\![\mathcal{O}_K^\times]\!]$, and both authors would like to thank Patrick Allen and Toby Gee for comments. The second author would like to thank Professors Nigel Boston, Jordan Ellenberg, Lue Pan and Richard Taylor for their comments on the earlier versions of this work and also he would like to thank the hospitality of the department of mathematics of University of Chicago.

Appendix A. Finiteness of unramified deformation rings

Let $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ be an absolutely irreducible odd Galois representation of the form $\mathrm{Ind}_{G_L}^{G_{\mathbf{Q}}}\chi$, where $L/\mathbf{Q}$ is the quadratic subfield $L \subset \mathbf{Q}(\zeta_p)$. Suppose that, up to unramified twist:

$$\overline{\rho}|_{G_{\mathbf{Q}_p}} = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}_p}} \overline{\varepsilon}_2^{n-1}, \quad n - 1 \equiv \frac{p+1}{2} \bmod p+1. \tag{8}$$

The main theorem of this section is the following.

Theorem A.0.1. *Assume that $p \equiv 1 \mod 4$, so $L/\mathbf{Q}$ is real. Let $(N,p)=1$, and let R^{split} denote the universal deformation ring of $\overline{\rho}$ consisting of representations which are unramified outside N and totally split at p defined in Definition 2.4.2. Then R^{split} is finite over $W(k)$.*

Before proving this, we begin with a preliminary lemma:

Lemma A.0.2. *There exists a finite extension $F/\mathbf{Q}$ with the following properties*

1. *F is totally real.*
2. *If $v|N$, then $\overline{\rho}|_{F_v}$ is trivial.*
3. *$\overline{\rho}|_{G_F}$ is absolutely irreducible.*
4. *If $v|p$, then $F_v \simeq K$, the unique unramified quadratic extension of $\mathbf{Q}_p$.*

Proof. The existence of F follows immediately from [Cal12, Proposition 3.2] (see also [MB90]). For example, one may can choose $G = \mathrm{GL}_2(k)$, and then make the following choices:

1. ϕ_v for $v|N$ is the map $\overline{\rho}|_{G_v} \to G$,
2. ϕ_v for $v = p$ is any injective map $\mathrm{Gal}(K/\mathbf{Q}_p) \to G$,
3. c_∞ is trivial.

The irreducibility of $\overline{\rho}|_{G_F}$ is then guaranteed by choosing $F/\mathbf{Q}$ to be linearly disjoint from the fixed field M of $\ker(\overline{\rho})$ by [Cal12, Lemma 3.2] (2). □

Proof of Theorem A.0.1. We begin with some reductions. If F/E is any finite extension so that $\overline{\rho}|_{G_F}$ remains absolutely irreducible, and R_F and R_E denote the universal deformation rings of $\overline{\rho}|_{G_F}$ and $\overline{\rho}|_{G_E}$ respectively, then the map $R_F \to R_E$ is always finite. As a consequence, to prove the finiteness of R^{split}, it suffices to replace $\mathbf{Q}$ by any totally real field in which $\overline{\rho}$ remains absolutely irreducible. We replace $\mathbf{Q}$ by the field $F/\mathbf{Q}$ constructed in Lemma A.0.2, so that $\overline{\rho}|_{G_v}$ is trivial for each $v|N$. By Lemma A.0.2, the field F_v for $v|p$ is precisely the unramified extension $K/\mathbf{Q}_p$ for all $v|p$. In particular,

$$\overline{\rho}|_{F_v} \simeq \overline{\varepsilon}_2^{n-1} \oplus \overline{\varepsilon}_2^{p(n-1)}$$

is reducible and p-distinguished. Make an arbitrary choice of one of these characters for each $v|p$, which gives a distinguished choice U_k of the decomposition $V_k = U_k \oplus U_k'$ as G_{F_v} modules for each $v|p$.

We now recall the deformation ring $R_{\mathcal{D}}$ defined in [SW01, §2] with respect to $\mathcal{D} = (W(k), \Sigma, \emptyset)$ where Σ is the set of primes dividing N. Note that $\overline{\rho}|_{G_F}$ is absolutely irreducible and satisfies all the conditions of [SW01] by construction. The ring $R_{\mathcal{D}}$ is global deformation ring of $\overline{\rho}$ unramified outside Np subject to the following condition: for all $v|p$, there exists a short exact sequence

$$0 \to U_A \to V_A \to U_A' \to 0$$

of G_{F_v}-modules where U_A and U_A' are free over A of rank one and $U_A/\mathfrak{m} = U_k$.

After extending F if neccessary we may assume that $[F : \mathbf{Q}]$ is even and thus the hypothesis ($\mathrm{H}_{\mathrm{even}}$) of [SW01, §3] holds. By [All14, Lemma 5.1.2], there exists a lift ρ_0 of $\overline{\rho}$ giving rise to a point on $R_{\mathcal{D}}$ relative to our choices both over F and any totally real extension of F in which $\overline{\rho}|_{G_F}$ remains irreducible. This implies that condition ($\mathrm{H}_{\mathrm{def}}$) of [SW01, §3] holds for any such F. In this situation we have a corresponding Hecke ring $\mathbf{T}_{\mathcal{D}}$ as defined in [SW01, p.196], and a surjection $R_{\mathcal{D}} \to \mathbf{T}_{\mathcal{D}}$ of Λ-modules where Λ is an Iwasawa algebra [SW01, p.191] (denoted $\Lambda_{\mathcal{O}}$).

There is a surjection $R_{\mathcal{D}} \to R^{\text{split}}$. Any deformation coming from R^{split} locally has the form $U_A \oplus U'_A$ where the action of G_K on U_A and U'_A is via the Teichmüller lifts of $\overline{\varepsilon}_2^{n-1}$ and $\overline{\varepsilon}_2^{p(n-1)}$ respectively. In particular, the map

$$\Lambda \to R_{\mathcal{D}} \to R^{\text{split}}$$

factors through a quotient of the form $W(k)$, as can be seen from the formulas in [SW01] on the last line of p.191 and the first line of p.192 respectively. A more intrinsic way to see this is that the ring Λ represents weight space and all split representations lie in the same fixed unramified weight. Hence to prove that R^{split} is finite over $W(k)$ it suffices to show that $R_{\mathcal{D}}$ is finite over Λ.

By taking the compositum of F with a suitably large totally subfield of a cyclotomic extension (exactly as in the first two lines at the top of page 204 of [SW01]) we may further ensure that the pair (F, ρ_0) is *good* in the sense of [SW01, §4]. It folows from [SW01, Prop 4.1] and [SW01, Prop 8.2] that all primes of $R_{\mathcal{D}}$ are promodular. If $\mathbf{T}_{\mathcal{D}}$ is the Hecke ring defined in [SW01, p.196], it follows that there is an isomorphism $R_{\mathcal{D}}/\mathfrak{p} \simeq \mathbf{T}_{\mathcal{D}}/\mathfrak{p}$ for every prime $\mathfrak{p}$ of $R_{\mathcal{D}}$, which implies immediately that $(R_{\mathcal{D}})^{\text{red}} = (\mathbf{T}_{\mathcal{D}})^{\text{red}}$ and hence $(R_{\mathcal{D}})^{\text{red}}$ is finite over Λ (since $\mathbf{T}_{\mathcal{D}}$ is finite over Λ by [SW01, Lemma 3.3]). Since $R_{\mathcal{D}}$ is Noetherian, it follows that $R_{\mathcal{D}}$ is also finite over Λ. But $R_{\mathcal{D}}$ surjects onto R^{split}/p, and the kernel contains the image of the maximal ideal of Λ. It follows that R^{split}/p is finite over k and hence R^{split} is also finite over $W(k)$, as claimed. □

References

[AC14] Patrick B. Allen and Frank Calegari. Finiteness of unramified deformation rings. *Algebra Number Theory*, 8(9):2263–2272, 2014.

[All14] Patrick B. Allen. Modularity of nearly ordinary 2-adic residually dihedral Galois representations. *Compos. Math.*, 150(8):1235–1346, 2014.

[Böc13] Gebhard Böckle. Galois representations. In *Travaux mathématiques. Vol. XXIII*, volume 23 of *Trav. Math.*, pages 5–35. Fac. Sci. Technol. Commun. Univ. Luxemb., Luxembourg, 2013.

[Bre03] Christophe Breuil. Sur quelques représentations modulaires et p-adiques de $\mathrm{GL}_2(\mathbf{Q}_p)$. II. *J. Inst. Math. Jussieu*, 2(1):23–58, 2003.

[Cal12] Frank Calegari. Even Galois representations and the Fontaine–Mazur conjecture. II. *J. Amer. Math. Soc.*, 25(2):533–554, 2012.

[Che08] Gaëtan Chenevier. Quelques courbes de Hecke se plongent dans l'espace de Colmez. *J. Number Theory*, 128(8):2430–2449, 2008.

[Duk95] W. Duke. The dimension of the space of cusp forms of weight one. *Internat. Math. Res. Notices*, (2):99–109, 1995.

[Gou01] Fernando Q. Gouvêa. Deformations of Galois representations. In *Arithmetic algebraic geometry (Park City, UT, 1999)*, volume 9 of *IAS/Park City Math. Ser.*, pages 233–406. Amer. Math. Soc., Providence, RI, 2001. Appendix 1 by Mark Dickinson, Appendix 2 by Tom Weston and Appendix 3 by Matthew Emerton.

[GV04] Eknath Ghate and Vinayak Vatsal. On the local behaviour of ordinary Λ-adic representations. *Ann. Inst. Fourier (Grenoble)*, 54(7):2143–2162 (2005), 2004.

[Joc82] Naomi Jochnowitz. Congruences between systems of eigenvalues of modular forms. *Trans. Amer. Math. Soc.*, 270(1):269–285, 1982.

[Lan83] Serge Lang. *Fundamentals of Diophantine geometry*. Springer-Verlag, New York, 1983.

[Leh47] D. H. Lehmer. The vanishing of Ramanujan's function $\tau(n)$. *Duke Math. J.*, 14:429–433, 1947.

[LMF20] The LMFDB Collaboration. The L-functions and modular forms database, 2020. [Online; accessed 31 July 2020].

[Maz89] B. Mazur. Deforming Galois representations. In *Galois groups over* **Q** *(Berkeley, CA, 1987)*, volume 16 of *Math. Sci. Res. Inst. Publ.*, pages 385–437. Springer, New York, 1989.

[MB90] Laurent Moret-Bailly. Extensions de corps globaux à ramification et groupe de Galois donnés. *C. R. Acad. Sci. Paris Sér. I Math.*, 311(6):273–276, 1990.

[PS16] Vincent Pilloni and Benoît Stroh. Surconvergence, ramification et modularité. *Astérisque*, (382):195–266, 2016.

[Sai97] Takeshi Saito. Modular forms and p-adic Hodge theory. *Invent. Math.*, 129(3):607–620, 1997.

[Sas] Shu Sasaki. Integral models of Hilbert modular varieties in the ramified case, deformations of modular Galois representations, and weight one forms II. preprint.

[Sch90] A. J. Scholl. Motives for modular forms. *Invent. Math.*, 100(2):419–430, 1990.

[Ser72] Jean-Pierre Serre. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. *Invent. Math.*, 15(4):259–331, 1972.

[Ser18] Vlad Serban. An infinitesimal p-adic multiplicative Manin-Mumford conjecture. *J. Théor. Nombres Bordeaux*, 30(2):393–408, 2018.

[Shi12] Sug Woo Shin. Automorphic Plancherel density theorem. *Israel J. Math.*, 192(1):83–120, 2012.

[Ski09] Christopher Skinner. A note on the p-adic Galois representations attached to Hilbert modular forms. *Doc. Math.*, 14:241–258, 2009.

[SW01] C. M. Skinner and Andrew J. Wiles. Nearly ordinary deformations of irreducible residual representations. *Ann. Fac. Sci. Toulouse Math. (6)*, 10(1):185–215, 2001.

Email address: fcale@math.uchicago.edu

The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA

Email address: ntalebiz@ias.edu

The Institute For Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540 USA