A simplified proof of Serre's conjecture

Luis Victor Dieulefait[1] · **Ariel Martín Pacetti[2]**

Abstract
The purpose of the present article is to present a simplified proof of Serre's modularity conjecture using the strong modularity lifting results currently available.

Introduction

In [29] (Section 3, Problem 1) Serre posted the following question:

Question: is it true that any odd, continuous, irreducible two dimensional Galois representation of the absolute Galois group of $\mathbb{Q}$ (denoted by $\mathrm{Gal}_{\mathbb{Q}}$) defined over a finite field is obtained as the reduction of the p-adic Galois representation attached to a modular form?

Recall that a 2-dimensional representation of $\mathrm{Gal}_{\mathbb{Q}}$ is called *odd* if the image of complex conjugation has determinant -1. This question is known as "Serre's weak modularity conjecture". In his famous 1987 paper ([30]) he went further, giving a precise recipe for a level and weight (minimal in a certain sense) where the modular form should appear. This second conjecture is known as "Serre's strong modularity conjecture". A very nice reference for the precise statement of the strong modularity conjecture are the lecture notes [26]. The equivalence between Serre's strong and weak version is due to many authors, the main contributions

---

✉ Luis Victor Dieulefait
   ldieulefait@ub.edu

   Ariel Martín Pacetti
   apacetti@ua.pt

[1] Departament de Matemàtiques i Informàtica, Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain

[2] Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal

∅ Springer

on the weight reduction being due to Edixhoven (see [8] and the references therein). The level reduction is mainly due to Ribet (see [1, 23] see also [24]). For this reason, we will focus on proving Serre's weak modularity conjecture, namely.

Theorem (Serre's modularity conjecture) *Let $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be an odd continuous irreducible Galois representation. Then ρ is modular, i.e. there exists a modular form $f \in S_k(\Gamma_0(N), \varepsilon)$ such that $\rho \simeq \overline{\rho_{f,p}}$.*

Remark 1 In the present article (and following Serre's original conjecture given in [30]) we will only consider the so called "cohomological" modular forms, i.e. those whose weight k is at least 2 (so that they appear in the cohomology of a Shimura curve). However, the notion of "modularity" in the last statement could also include modular forms of weight one, since the congruence proved in [7] §6.9 shows that if the representation ρ is congruent to a representation coming from a weight 1 modular form, then it is also congruent to the representation attached to a cohomological modular form, so we can (and will) restrict to weights $k \geq 2$.

The first cases of the conjecture (small level and weight) were proved by Khare and Wintenberger ([16]), and also independently by the first named author ([5]). A complete proof of the conjecture was then given by Khare and Wintenberger in [17, 18] (a proof for the case of odd level was also given by the first named author in [6]). The proof is based on a smart inductive argument involving both a level and a weight reduction. The main issue when the proof appeared was that different modularity lifting results required many technical conditions (specially while manipulating representations whose residual image is reducible). Such conditions have been removed during the last years, which allows us to present a more elegant inductive argument. In particular, our argument becomes much simpler because the sophisticated process of weight reduction disappears. The procedure (that will be explained in detail in Sect. 2) to prove the conjecture when p is odd is the following:

- Make the representation ρ part of a compatible system $\{\rho_\ell^{(0)}\}$ (whose definition is recalled in Sect. 1.3).
- Add a large prime N to the level of the family. This prime is needed to ensure that the p-th member of the family (and the families appearing in the "chain" of congruences) has large residual image for each prime p dividing the "level" of the family (except at N itself). The prime number N is called a *good-dihedral prime* in [17].
- If a prime $p \neq N$ is in the level of the system, remove it by looking at the reduction modulo p of the p-th member of the family and taking a minimal lift (the prime $p = 2$ is handled in a similar way, with extra technicalities).
- Remove the remaining prime N from the level via the same procedure (strong modularity lifting results assure this can be done even if the residual image is reducible). Now we are left with a family of level 1 (but without any control on the weight).
- The reduction of the 5-th member of the last family is either reducible (so it is modular), or it is irreducible with Serre weight (up to twist) 2, 4 or 6. In the first two cases, taking a lift with the right weight, and moving (through a compatible family) to the prime $p = 3$ (as explained in **Paso 6**) allows us to reach the base case of level 1 proved by Serre. If Serre's weight at the prime 5 is 6, our representation is the reduction of a representation attached a semistable abelian variety unramified outside 5, corresponding to a base case proved by Schoof.

It is important to emphasize that the last step, which substitutes the "weight reduction", is only possible due to two very strong and general modularity lifting theorems: a result of

Kisin in the residually irreducible case and a recent result of Pan in the residually reducible case. The proof for $p = 2$ is based on a reduction to the odd case.

The article is organized as follows: the first section is the most technical one. It contains the main results (mostly different modularity lifting theorems) needed to prove Serre's modularity conjectures. One of the goals of the present article is to allow a non-expert reader to learn the ideas behind the proof of Serre's conjectures taking for granted the results of this section. However, we included in the first section two lemmas (1.13 and 1.14) on properties of residual Galois representations that are well known to experts, but whose detailed proof is hard to find in the literature. The second section contains the proof of Serre's conjectures in the case of odd characteristic, filling in the details of the previous sketch. The last section contains the proof in characteristic two.

1 Main results involved in the proof

1.1 Base cases

The base cases we rely on to *propagate* modularity are the following.

Theorem 1.1 *There are no continuous, odd, absolutely irreducible two dimensional Galois representations of the absolute Galois group of $\mathbb{Q}$ unramified outside p and with values on a finite field of characteristic p, for $p = 2$ or 3.*

Proof The result was proved by Tate ([36]) for $p = 2$, and later Serre observed that the exact same argument worked for $p = 3$ ([32, page 710]).   □

Theorem 1.2 *There do not exist non-zero semistable rational abelian varieties that have good reduction outside ℓ for $\ell = 2, 3, 5, 7, 13$.*

Proof See [28, Theorem 1.1].   □

Theorem 1.3 (Langlands–Tunnell) *Let p be an odd prime and $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ be an odd, continuous, irreducible representation with solvable image. Then ρ is modular.*

Proof By Dickson's classification, any solvable subgroup of $\mathrm{PGL}_2(\overline{\mathbb{F}_p})$ is isomorphic to a cyclic group, a dihedral group, A_4 or S_4. In all cases, the representation lifts to an odd representation of $\mathrm{GL}_2(\mathbb{C})$. The cyclic case gives a reducible representation (and can be omitted). The dihedral case was considered by Hecke, the tetrahedral case was proved by Langlands in [19] and the octahedral by Tunnell in [37]. Note that all the aforementioned results provide a weight one modular form (which is enough for our purposes by Remark 1).   □

1.2 Modularity lifting Theorems

Let $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ be a continuous Galois representation.

Problem: how can we decide whether ρ matches the representation attached to a modular form?

The term "Modularity lifting Theorem" refers to results (like the ones given in the pioneering articles [40, 41]) that provide an answer to the problem. Most modularity lifting theorems

have as a key hypothesis that the residual representation $\overline{\rho}$ (obtained as the reduction of ρ modulo p) matches the reduction of a representation coming from a modular form. If the residual representation $\overline{\rho}$ happens to be reducible (a very hard case of study), it matches the representation attached to an Eisenstein series, so it is already "residually modular".

If p is an odd prime, let $p^\star = \left(\frac{-1}{p}\right) p$, so that the extension $\mathbb{Q}(\sqrt{p^\star})$ is the unique quadratic extension of $\mathbb{Q}$ unramified outside p.

1.2.1 Modularity lifting Theorems for residually irreducible representations

Theorem 1.4 *Let p be an odd prime and $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ be a continuous, odd Galois representation ramified at finitely many primes and satisfying all the following hypotheses:*

- *The residual restriction $\overline{\rho}|_{\mathrm{Gal}_{\mathbb{Q}(\sqrt{p^\star})}}$ is absolutely irreducible,*
- *The representation $\rho|_{\mathrm{Gal}_{\mathbb{Q}_p}}$ is de Rham with Hodge-Tate weights $\{0, k-1\}$, with $k > 1$,*
- *The residual representation is modular, i.e. $\overline{\rho} \simeq \overline{\rho_f}$.*

Then ρ matches the representation of a weight k modular form.

Proof The case $k = 2$ is proven in [15] (Theorem in the second page), while the general case follows from [13] (also stated as Theorem in the second page). There are two extra hypotheses in the last result: the second one (related to a compatibility between classical and p-adic local Langlands correspondence, as explained in Hypothesis (1.2.6) of [13]) is removed in [9, Theorem 1.2.1] and in [21, Theorem 1.1]. Our precise statement incorporates the results in [11, Theorem 1.4], where the other hypothesis in Kisin's article is removed for $p \geq 5$ together with [38] (main theorem) where it is removed for $p = 3$.   □

We also need a similar result for $p = 2$.

Theorem 1.5 *Let $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_2})$ be a continuous, odd Galois representation ramified at finitely many primes and satisfying the following hypotheses:*

- *The representation $\rho|_{\mathrm{Gal}_{\mathbb{Q}_2}}$ is de Rham with Hodge-Tate weights $\{0, k-1\}$, with $k > 1$,*
- *The residual representation $\overline{\rho}$ is modular and has non-solvable image.*

Then ρ matches the representation of a weight k modular form.

Proof See [14, Theorem 0.1] for $k = 2$ and ρ potentially Barsotti-Tate, [22, Theorem 1.1] and [39, Theorem A] for the general case.   □

1.2.2 Modularity of residually reducible representations

Theorem 1.6 *Let $p \geq 5$ be a prime number and $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ be a continuous, irreducible odd Galois representation ramified at finitely many primes and satisfying the following two hypotheses:*

- *The representation $\rho|_{\mathrm{Gal}_{\mathbb{Q}_p}}$ is de Rham with Hodge-Tate weights $\{0, k-1\}$ and $k > 1$,*
- *The semisimplification of $\overline{\rho}$ is a sum of two characters $\overline{\chi}_1 \oplus \overline{\chi}_2$.*

Then ρ matches the representation of a weight k modular form.

Proof See [34](Theorem in the third page) and [20, Theorem 1.0.2].   □

Theorem 1.7 *Let $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_3})$ be a continuous, irreducible odd Galois representation ramified at finitely many primes and satisfying all the following hypotheses:*

- $\overline{\rho}^{ss} \simeq 1 \oplus \chi_3,$
- $\rho|_{D_3} \neq \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right),$
- $\rho|_{I_3} \simeq \left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right),$
- $\det(\rho) = \psi \chi_3^{k-1}$ *for some $k \geq 2$.*

Then ρ matches the representation of a weight k modular form.

Proof See [34] Theorem in the third page.   □

1.3 Existence of lifts

Let $\overline{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_q)$ be a Galois representation.

Problem: does there exist a continuous representation $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ whose residual representation is isomorphic to $\overline{\rho}$?

Note that the representation ρ (if it exists) is far from being unique. Any such representation is called a *lift* of $\overline{\rho}$. While working with deformation rings, one studies lifts into more general coefficient rings, but for our purposes it is enough to restrict to finite extensions of $\mathbb{Q}_p$.

Let $\overline{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ be an odd, continuous, irreducible Galois representation, and let $k(\overline{\rho})$ be the weight of the residual representation as defined by Serre in [30] (formulas (2.2.4), (2.3.2), (2.4.5), (2.4.8) and (2.4.9)).

Remark 2 Given $\psi : \mathrm{Gal}_{\mathbb{Q}} \to \overline{\mathbb{F}_p}^{\times}$ a continuous character, it always has a lift to $\overline{\mathbb{Q}_p}^{\times}$ (for example by taking its Teichmüller lift). Then the veracity of Serre's conjecture for a representation $\overline{\rho}$ is equivalent to the veracity of the twist of $\overline{\rho}$ by ψ. Following Serre's notation, there is always a twist of our representation $\overline{\rho}$ such that in formulas $(2, 2, 4)$, $(2, 3, 2)$, $(2, 4, 5)$, $(2, 4, 8)$ and $(2, 4, 9)$ of [30] we can take $a = 0$, so we can (and will during the present article) abusing notation assume that the weight $k(\overline{\rho})$ is at most $p + 1$ if p is odd and at most 4 if $p = 2$.

The inductive argument in the proof of Serre's conjecture depends on reducing (through a combination of taking lifts, building families and creating congruences) the number of ramified primes of the representation $\overline{\rho}$ ("killing the level"), so we need to be able to impose extra conditions on the lift. Let $E/\mathbb{Q}_p$ be a finite extension and let $\mathcal{O}_E$ denote its ring of integers.

Definition 1.8 Let ℓ be a prime number, and let I_ℓ be the inertia subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}_\ell}/\mathbb{Q}_\ell)$. An *inertial type* at ℓ is a continuous representation $\tau_\ell : I_\ell \to \mathrm{GL}_2(E)$.

For a prime $\ell \neq p$, we will be mostly concerned with the following two inertial types:

- the *unramified* type is the one corresponding to the trivial representation.
- the *Steinberg* type, corresponding to the restriction to the inertial subgroup of the representation of $\mathrm{Gal}_{\mathbb{Q}_\ell}$ sending a generator of the tame inertia to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and a Frobenius element to $\left(\begin{smallmatrix} \ell & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

In order to define the inertial type at p (i.e. $\ell = p$) it is better to consider representations of the Weil-Deligne group of $\mathrm{Gal}_{\mathbb{Q}_p}$ (see [35]). A representation of the Weil-Deligne group consists of pairs (τ, N) where:

(1) $\tau : W(\mathbb{Q}_p) \to \mathrm{GL}_2(\mathbb{C})$ is a 2-dimensional complex representation of the Weil group,
(2) N is a nilpotent endomorphism of $\mathbb{C}^2$ such that

$$wNw^{-1} = \omega_1(w)N, \text{ for all } w \in W(\mathbb{Q}_p),$$

where ω_1 is the unramified quasi-character giving the action of $W(\mathbb{Q}_p)$ on roots of unity.

The unramified type at p is defined to be the p-adic Galois representation whose Weil-Deligne representation equals $(\omega_1^{k-1} \oplus 1, \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right))$, while the Steinberg type at p is defined to be a twist of $(\omega_1 \oplus 1, \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right))$. The unramified type at p corresponds to the notion of a *crystalline* representation in p-adic Hodge theory.

Let Σ be a finite set of primes containing p and the primes where $\overline{\rho}$ is ramified. We want to impose an inertial type condition on deformations of $\overline{\rho}$ at each prime of Σ. For that purpose, for each $\ell \in \Sigma$, $\ell \neq p$, let τ_ℓ be an inertial type compatible with $\overline{\rho}$, i.e. such that there exists an $\mathcal{O}_E$-lattice Λ_ℓ in E^2 which is stable by τ_ℓ (so the choice of a basis for Λ_ℓ provides a representation $\tau_\ell : I_\ell \to \mathrm{GL}_2(\mathcal{O}_E)$) such that $\overline{\tau_\ell} = \overline{\rho}|_{I_\ell}$.

For the purposes of the present article, a "minimal lift" is a lift that is unramified at all primes $\ell \neq p$ where the representation $\overline{\rho}$ is unramified.

Theorem 1.9 *Let $\overline{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be an odd, continuous, representation whose restriction to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$ is absolutely irreducible. Assume furthermore that when $p = 2$ $\overline{\rho}$ has non-solvable image. Then there exists a lift $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(E)$ (for some finite extension $E/\mathbb{Q}_p$) with any of the following prescribed properties:*

(1) *If $p = 2$ and $k(\overline{\rho}) = 2$, then ρ is a minimal crystalline lift with Hodge-Tate weights $\{0, 1\}$.*
(2) *If $p = 2$ and $k(\overline{\rho}) = 4$, then ρ is a lift with Hodge-Tate weights $\{0, 1\}$, minimally ramified outside 2 and the inertial Weil-Deligne parameter at 2 is given by $(\omega_1 \oplus 1, \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right))$.*
(3) *If $p > 2$, then ρ is a minimal crystalline lift with Hodge-Tate weights $\{0, k(\overline{\rho}) - 1\}$.*
(4) *If $p > 2$, then ρ is a lift with Hodge-Tate weights $\{0, 1\}$, any inertial type τ_ℓ compatible with $\overline{\rho}$ at primes $\ell \neq p$ in Σ and unramified outside Σ. Furthermore, if $k(\overline{\rho}) = 2$ the lift can be taken to be crystalline at p and if $k(\overline{\rho}) = p + 1$ the lift can be taken to be Steinberg at p, i.e., its inertial Weil-Deligne parameter at p is given by $(\omega_1 \oplus 1, \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right))$*

Proof The first three cases are due to Khare-Wintenberger ([17, Theorem 5.1], its proof given in [18]). Partial results of the last case are also proven in Khare-Wintenberger's article (same Theorem), the more general case is due to Gee and Snowden ([10]; [33, Theorem 7.2.1]). □

1.4 Existence of almost strictly compatible systems

One important property of Galois representations coming from modular forms is that they come in "families". More concretely, if $f \in S_k(\Gamma_0(N), \varepsilon)$, $k \geq 2$, is a newform, then (by [3]) for every prime number p, there exists a continuous odd Galois representation $\rho_{f,p} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ unramified outside Np, such that for any prime number $\ell \nmid Np$, the characteristic polynomial of $\rho_{f,p}(\mathrm{Frob}_\ell)$ equals $x^2 - a_\ell(f)x + \varepsilon(\ell)\ell^{k-1}$, where a_ℓ denotes the eigenvalue of f for the action of the Hecke operator T_ℓ. Furthermore, one can obtain information at primes ℓ dividing Np as well.

There is a notion of compatible families for abstract representations (as given by Serre in [31], I-11). Recall the definition of a strictly compatible system and an almost strictly compatible system of Galois representations given in [17].

Definition 1.10 A rank 2 *strictly compatible system* of Galois representations $\mathcal{R}$ of $\mathrm{Gal}_{\mathbb{Q}}$ defined over K is a 5-tuple

$$\mathcal{R} = (K, S, \{Q_\ell(x)\}, \{\rho_{\mathfrak{p}}\}, k),$$

where

(1) K is a number field.
(2) S is a finite set of primes.
(3) for each prime $\ell \notin S$, $Q_\ell(x)$ is a degree 2 polynomial in $K[x]$.
(4) For each prime ideal $\mathfrak{p}$ of K, the representation

$$\rho_{\mathfrak{p}} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(K_{\mathfrak{p}}),$$

is a continuous semisimple representation such that:

- If $\ell \notin S$ and $\ell \nmid \mathcal{N}(\mathfrak{p})$ (the norm of $\mathfrak{p}$), then $\rho_{\mathfrak{p}}$ is unramified at ℓ and $\rho_{\mathfrak{p}}(\mathrm{Frob}_\ell)$ has characteristic polynomial $Q_\ell(x)$.
- If $\ell \mid \mathcal{N}(\mathfrak{p})$, then $\rho|_{\mathrm{Gal}_{\mathbb{Q}_\ell}}$ is de Rham and furthermore crystalline if $\ell \notin S$.

(5) The Hodge-Tate weights $\mathrm{HT}(\rho_{\mathfrak{p}}) = \{0, k-1\}$.
(6) For each prime ℓ there exists a Weil-Deligne representation $\mathrm{WD}_\ell(\mathcal{R})$ of $W_{\mathbb{Q}_\ell}$ over $\overline{K}$ such that for each place $\mathfrak{p}$ of K and every K-linear embedding $\iota : \overline{K} \hookrightarrow \overline{K}_{\mathfrak{p}}$, the push forward $\iota\,\mathrm{WD}_\ell(\mathcal{R}) \simeq \mathrm{WD}(\rho_{\mathfrak{p}}|_{\mathrm{Gal}_{\mathbb{Q}_\ell}})^{K\text{-ss}}$.

Remark 3 Comparing to the case of representations coming from a newform $f \in S_k(\Gamma_0(N), \varepsilon)$, the set S consists of the primes dividing N. Deligne's result implies that $\rho_{f,p}$ satisfies the third hypothesis and the first item of the fourth one, while the last one is the compatibility at the primes dividing Np.

An *almost strictly compatible system* is a 5-tuple satisfying the first five properties, and also condition (6) but with some exceptions: for a prime λ whose residual characteristic is equal to the prime p, if the residual representation $\bar\rho_\lambda$ is reducible, then we only impose the compatibility as in condition (6) if this prime p is odd and the representation $\mathrm{WD}_p(\mathcal{R})$ is unramified.

Theorem 1.11 *Let $\rho : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(K_\lambda)$ be an odd, irreducible, continuous Galois representation ramified at finitely many places and de Rham at p with Hodge-Tate weights $\{0, k-1\}$, with $k > 1$ such that the restriction of $\bar\rho$ to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$ is absolutely irreducible. Assume furthermore that when $p = 2$ $\bar\rho$ has non-solvable image. Then ρ is part of a rank 2 almost strictly compatible system of Galois representations.*

Proof See [4, Theorem 1.1]. ∎

Remark 4 If ρ is part of a compatible system of Galois representations $\{\rho_{\mathfrak{p}}\}$, then ρ is modular if and only if any given member of the family is. The reason is the following: if $\rho \simeq \rho_f$, where f is a newform, then by Deligne's theorem, there exists a strong compatible system $\{\rho_{f,\mathfrak{p}}\}$ containing ρ_f. Then for any prime $\mathfrak{p}$, the representations $\rho_{\mathfrak{p}}$ and $\rho_{f,\mathfrak{p}}$ have the same trace and determinant at the Frobenius element Frob_ℓ, for all prime numbers ℓ belonging to a density one set of primes (actually all primes but finitely many), so by the Brauer-Nesbitt theorem, they are indeed isomorphic.

1.5 Some lemmas on the image of Galois representations

Definition 1.12 A residual representation $\bar{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ which is irreducible but becomes reducible while restricted to $\mathrm{Gal}_{\mathbb{Q}(\sqrt{p^\star})}$ is called *bad dihedral*.

Some of the previous theorems are stated under the hypothesis that the restriction of the residual representation to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$ is absolutely irreducible. Let us show that this is in fact equivalent to the condition (introduced by Wiles) of not being bad dihedral.

Lemma 1.13 *Let p be an odd prime and $\bar{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ be an odd continuous representation. Then the following are equivalent:*

(1) $\bar{\rho}|_{\mathrm{Gal}_{\mathbb{Q}(\sqrt{p^\star})}}$ *is irreducible,*
(2) $\bar{\rho}|_{\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}}$ *is irreducible.*

Proof We can assume that $p \neq 3$ as otherwise the statement is trivial. Clearly the second condition implies the first one. For the converse, suppose that the restriction of $\bar{\rho}$ to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$ is reducible. In particular, the image of $\bar{\rho}$ is a solvable group. Then either our representation is contained in a Borel group (hence it is reducible), it lies in the normalizer of a split Cartan group (as our coefficient field is algebraically closed) or its projective image is one of the exceptional groups A_4 or S_4 (it cannot be A_5 because it is solvable).

Let G denote the image of $\bar{\rho}$ and suppose it lies in N, the normalizer of a Cartan group. Recall that N fits into the short exact sequence

$$1 \longrightarrow T \longrightarrow N \overset{\phi}{\longrightarrow} \mathbb{Z}/2 \longrightarrow 1$$

where T is a torus (corresponding to matrices of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$), and $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ can be taken as a lift of the generator of $\mathbb{Z}/2$. If we intersect each term of the exact sequence with the subgroup G, we get a similar sequence. Note that the image of G by ϕ is non-trivial as otherwise the group G would be abelian, and $\bar{\rho}$ would not be irreducible. Also note that since G is not abelian, it must contain at least one matrix of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$ with $a \neq b$.

Let H be the image of the restriction of $\bar{\rho}$ to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$, a normal subgroup of G. Since the restriction of $\bar{\rho}$ to $\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}$ is reducible, its image must be an abelian group. The reason is that in some chosen basis, its image lies in a Borel subgroup, but the normalizer of a Cartan group does not have elements of order p (as its order is not divisible by p). The only abelian subgroups of N are the ones contained in T, or subgroups of the form $\left\langle \left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \right\rangle$ (where a lies in $\overline{\mathbb{F}_p}^{\times}$). The latter are not normal subgroups of G (since any such group is not preserved under conjugation by any matrix of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$ when $a \neq b$), so H must be a subgroup of T.

Let $\sigma \in \mathrm{Gal}_{\mathbb{Q}}$ be such that it generates the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$. Then $G = \langle H, \bar{\rho}(\sigma) \rangle$. The image of the restriction of $\bar{\rho}$ to $\mathrm{Gal}_{\mathbb{Q}(\sqrt{p^\star})}$ equals $\langle H, \bar{\rho}(\sigma^2) \rangle$ which also lies in T so the restriction of $\bar{\rho}$ to $\mathrm{Gal}_{\mathbb{Q}(\sqrt{p^\star})}$ is also a reducible representation.

The other two cases actually cannot occur when $p > 3$. The reason is the following: suppose that the restriction $\bar{\rho}|_{\mathrm{Gal}_{\mathbb{Q}(\zeta_p)}}$ is reducible. This implies that the image of this restriction lies in a Borel subgroup. Since $p > 3$ and A_4 nor S_4 have elements of order p, this image must actually be an abelian group. Moreover, the image of its projectivization is a cyclic group.

This implies that the projective image of the residual representation $\overline{\rho}$ must contain a cyclic normal subgroup with cyclic quotient. There is no such a subgroup for the groups S_4 nor A_4. $\qquad\square$

Let us state (and give a detailed proof of) a result that is well-known to experts and will be needed later. As explained in Remark 1 we assume that (possible after twisting) all Serre weights lie in the range $[2, p + 1]$.

Lemma 1.14 *Let p be an odd prime and $\bar{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous odd bad dihedral representation. Then either $p = 2k(\bar{\rho}) - 3$ (niveau 2 case) or $p = 2k(\bar{\rho}) - 1$ (niveau 1 case).*

Proof This is Lemma 6.2 (ii) of [17]. The proof lacks some details, hence the difference between the niveau 1 and the niveau 2 case is hard to see (which justifies the present proof); our argument follows the lines of [25] (see Proposition 2.2). The fact that the representation has irreducible image together with the projective representation being dihedral (which corresponds to the image of the representation lying in the normalizer of a Cartan group) and $p \geq 3$ imply that there are no elements of order p in the image of $\bar{\rho}$. In particular, all matrices are semisimple and the image of the p-th inertia subgroup I_p factors through the tame part, so it is a cyclic group.

We claim that the projective image of $\bar{\rho}(I_p)$ has order at most 2. Suppose on the contrary that it has order n greater than two and that our representation $\bar{\rho}$ is bad dihedral. Let L denote the field extension fixed by the projective residual image, so $\mathrm{Gal}(L/\mathbb{Q})$ is a dihedral group D_{2m} of order $2m$ where $n \mid m$. Since our representation is bad dihedral, the restriction of the projectivization of $\bar{\rho}$ to the subgroup $\mathrm{Gal}(L/\mathbb{Q}(\sqrt{p^\star}))$ is reducible and decomposable (as all elements are semisimple). In particular, the group $\mathrm{Gal}(L/\mathbb{Q}(\sqrt{p^\star}))$ is a normal abelian subgroup of order m, i.e. it is the cyclic subgroup C_m of rotations, of index 2 in D_{2m}. Note that there is a unique such subgroup because we are assuming $n > 2$, thus $m > 2$.

Since the projective image of inertia is a cyclic group of order greater than 2, it must lie in the subgroup C_m of D_{2m} of rotations, so the field fixed by the rotations subgroup is on the one hand $\mathbb{Q}(\sqrt{p^\star})$ and on the other one a quadratic unramified at p extension of $\mathbb{Q}$, which is a contradiction. We have thus established that the projective image of $\bar{\rho}(I_p)$ has order at most 2.

Start considering the case of a "niveau 1" character, i.e, the image of the inertia group I_p is of the form $\begin{pmatrix} \chi_p^{k(\bar{\rho})-1} & 0 \\ 0 & 1 \end{pmatrix}$ with $k(\bar{\rho}) \leq p$ (recall that we have already shown that this image is abelian). Observe that in this case the order of this group agrees with its projective order. Then we know that $\chi^{k(\bar{\rho})-1}$ has order at most 2 and $k(\bar{\rho}) \leq p$, so either $2k(\bar{\rho}) - 2 = p - 1$ or $k(\bar{\rho}) - 1 = p - 1$. In the second case, we have $k(\bar{\rho}) = p$, but then the image of inertia is trivial, hence the Galois extension corresponding to $\bar{\rho}$ is disjoint from $\mathbb{Q}(\sqrt{p^\star})$. In particular, the representation is not bad dihedral.

In the case of a character of niveau 2, the image of the inertia groups I_p is of the form $\psi^{k(\bar{\rho})-1} \oplus \psi'^{k(\bar{\rho})-1}$, hence the projective order equals the order of $\psi^{(k(\bar{\rho})-1)(p-1)}$, which is the $k(\bar{\rho}) - 1$-st power of a character of order $p + 1$. It has order at most 2 when $\frac{p+1}{2} \mid k(\bar{\rho}) - 1$ and the condition $k(\bar{\rho}) \leq p$ gives that $p = 2k(\bar{\rho}) - 3$. $\qquad \square$

Remark 5 Over $\mathbb{Q}$, a Galois representation with Serre's level 1 cannot be bad dihedral in the niveau 2 case due to a Lemma of Wintenberger (see [12, Lemma 6.2] part (i)), hence the bad dihedral case can only occur for $p = 2k(\bar{\rho}) - 1$ in level 1. But in this case $k(\bar{\rho})$ is even, hence a bad dihedral representation of Serre's level 1 can only occur when $p \equiv 3 \pmod 4$.

The following lemma will also be needed in the proof.

Lemma 1.15 *Let p be a prime that is congruent to 1 modulo 4 and let $\overline{\rho} : \mathrm{Gal}_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_p)$ be an odd representation with non-solvable image. Denote by $\overline{\rho}_{proj}$ the projectivization of $\overline{\rho}$*

and by $c \in \mathrm{Gal}_{\mathbb{Q}}$ a complex conjugation. Then there exists a set of primes $\{q\}$ of positive density that are unramified for $\overline{\rho}$ and such that

(1) *$\overline{\rho}_{proj}(\mathrm{Frob}_q)$ and $\overline{\rho}_{proj}(c)$ define the same conjugacy class in $\overline{\rho}_{proj}(\mathrm{Gal}_{\mathbb{Q}})$,*
(2) *q is congruent to 1 modulo all primes $\leq p - 1$ and q is congruent to 1 modulo 8,*
(3) *q is congruent to -1 modulo p.*

Proof See [17, Lemma 8.2]. □

2 The proof of Serre's modularity conjecture for $p \neq 2$.

We can assume without loss of generality that ρ has non-solvable image, as otherwise modularity follows from Theorem 1.3. Let $\rho^{(0)}$ be a minimal crystalline lift of ρ, which exists by Theorem 1.9 (3) and let $\{\rho_\ell^{(0)}\}$ be an almost strictly compatible system passing through $\rho^{(0)}$ (Theorem 1.11). Let $k = k(\rho)$ be the weight of the system (recall that taking a suitable twist we always assume that $k \leq p + 1$ as explained in Remark 2), i.e., the Hodge-Tate weights of the representation $\rho_\ell^{(0)}$ are $\{0, k - 1\}$. Note that by Remark 4 it is enough to prove that any member of the family is modular.

Paso 1: change to a weight two system. Let w be a sufficiently large prime, so that it does not belong to the ramification set of the compatible system and such that $w > 2k$. Then $\rho_w^{(0)}$ is in what is called the *Fontaine-Laffaille* case (i.e. the representation is crystalline and the weight is smaller than the residual characteristic). Consider the reduction $\overline{\rho_w^{(0)}}$ of $\rho_w^{(0)}$. Then Serre's weight of $\overline{\rho_w^{(0)}}$ coincides with the weight k of the compatible system. Lemma 1.14 implies that $\rho_w^{(0)}$ is not residually bad dihedral. If its image happens to be solvable, we claim that $\rho_w^{(0)}$ is modular, and then so is $\rho_\ell^{(0)}$ (by Remark 4). The reason is that either the residual image of $\rho_w^{(0)}$ is reducible (in which case modularity follows from Theorem 1.6) or otherwise, it is irreducible but solvable, in which case the residual representation is modular by Theorem 1.3 and $\rho_w^{(0)}$ is modular by Theorem 1.4. In this case the proof ends here (this argument is crucial and will be used later on).

Suppose on the contrary, that the residual representation has non-solvable image. Take a weight 2 lift $\rho_w^{(1)}$ of $\overline{\rho_w^{(0)}}$, with ramification set equal to the one of $\overline{\rho_w^{(0)}}$ (whose existence is guaranteed by Theorem 1.9 (4)). This representation in general will not be crystalline at w. Now we can focus on proving modularity of the representation $\rho_w^{(1)}$, since Theorem 1.4 implies that $\rho_w^{(0)}$ is modular if and only if $\rho_w^{(1)}$ is. This is the idea of *propagating* modularity, one can move via a suitable congruence from one representation to another one until we hit a modular representation.

Using Theorem 1.11, make $\rho_w^{(1)}$ part of an almost strictly compatible system of Galois representations $\{\rho_\ell^{(1)}\}$ and let $p_1, \ldots, p_r$ be the primes in the ramification set of the system (set that we denote by S_1).

Paso 2: add what is called a "good-dihedral prime". This implies adding an extra prime N to the ramification set, such that the local type of the resulting system at N is dihedral, i.e. the induction of a character from a quadratic extension.

Let K be the coefficient field of the compatible system, and let $q \equiv 1 \pmod 4$ be a prime number greater than the primes in S_1 and also greater than 5 which splits in K. By Lemma 1.14, the residual representation $\overline{\rho_q^{(1)}}$ is not bad dihedral (we are using again the fact that we are in the Fontaine-Laffaille situation, now with a system of weight $k = 2$). If the

residual image of $\rho_q^{(1)}$ happens to be solvable, then once again $\rho_q^{(1)}$ is modular (because of Theorem 1.3 and 1.4 in the irreducible case and Theorem 1.6 in the reducible case). Then we can assume that the image of the residual representation $\overline{\rho_q^{(1)}}$ is non-solvable.

Since a continuous representation always fixes a lattice, after a possible conjugation, we can assume that the representation $\rho_q^{(1)}$ takes values in $\mathrm{GL}_2(\mathcal{O}_K)$ (the ring of integers of K), so its residual representation has image lying in $\mathrm{GL}_2(\mathbb{F}_q)$ (recall that q splits in K). Let N be a prime number which does not belong to the ramification set S_1 as in Lemma 1.15. The first hypothesis implies that $\mathrm{Tr}(\overline{\rho_q^{(1)}}(\mathrm{Frob}_N)) = 0$, and the last one implies that $\chi_q(\mathrm{Frob}_N) + 1$ is also congruent to zero, so the restriction of the residual representation $\overline{\rho_q^{(1)}}$ to the decomposition group at N (up to a twist) is of the form

$$\overline{\rho_q^{(1)}}|_{D_N} \simeq \begin{pmatrix} \overline{\chi_q} & 0 \\ 0 & 1 \end{pmatrix}.$$

Let $\mathbb{Q}_{N^2}$ denote the quadratic unramified extension of $\mathbb{Q}_N$, so that the unit group of its residue field has order $(N-1)(N+1)$. Since $q \mid (N+1)$, there exists (by class field theory) a character $\varkappa : \mathrm{Gal}_{\mathbb{Q}_{N^2}} \to \overline{\mathbb{Z}_q}^{\times}$ of order q (whose restriction to the inertia group at q is precisely a character of *niveau* 2). Note that the composition of $\varkappa$ with the residue map is trivial (as $\overline{\mathbb{F}_q}^{\times}$ does not have elements of order q), so the induction $\mathrm{Ind}_{\mathrm{Gal}_{\mathbb{Q}_{N^2}}}^{\mathrm{Gal}_{\mathbb{Q}_N}} \varkappa$ is a dihedral 2-dimensional representation of $\mathrm{Gal}_{\mathbb{Q}_N}$, whose residual representation is unramified, and has zero trace at a Frobenius element. In particular, we can take $\mathrm{Ind}_{\mathrm{Gal}_{\mathbb{Q}_{N^2}}}^{\mathrm{Gal}_{\mathbb{Q}_N}} \varkappa$ as our local type condition at the prime N.

Take a crystalline lift $\rho_q^{(2)}$ of weight 2 with such a dihedral image of order $2q$ at the prime N (whose existence is guaranteed by Theorem 1.9 (4)). Theorem 1.4 implies that the representation $\rho_q^{(1)}$ is modular if and only if the representation $\rho_q^{(2)}$ is modular. Make $\rho_q^{(2)}$ part of an almost strictly compatible system $\{\rho_\ell^{(2)}\}$, whose ramification set equals $S = S_1 \cup \{N\}$.

Lemma 2.1 *Let $\{\rho_\ell\}$ be an almost strictly compatible system of Galois representations, whose ramification set is contained in $S \cup \{2, 3\}$, and such that the local inertial type at the prime N matches the type of $\mathrm{Ind}_{\mathrm{Gal}_{\mathbb{Q}_{N^2}}}^{\mathrm{Gal}_{\mathbb{Q}_N}} \varkappa$ defined before. Then for any prime $p \in S_1 \cup \{2, 3\}$ the residual representation $\overline{\rho_p}$ has non-solvable image.*

Proof This result is taken from [17, Lemma 6.3]. Let $p \in S_1 \cup \{2, 3\}$ be a prime, and let ρ_p be the p-th representation of the family. The strong compatibility condition implies that its restriction to the decomposition group D_N is given as the induction of a ramified character of order q from the quadratic extension $\mathbb{Q}_{N^2}$. Since the prime q is larger than p (by construction), the kernel of the reduction map $\overline{\mathbb{Z}_p} \to \overline{\mathbb{F}_p}$ does not have elements of order q, hence the restriction of the residual representation $\overline{\rho_p}$ to the decomposition group D_N is irreducible (in particular, the residual representation is irreducible).

If the representation has solvable image, by Dickson's classification (and the fact that its projective image has order divisible by $q > 5$), its projective image must be a dihedral group. Suppose this is the case, so the projectivization of our residual representation is induced from a quadratic extension K of $\mathbb{Q}$. Note that K can only ramify at primes where our family does, namely some primes of $S \cup \{2, 3\}$ (recall that $p \in S \cup \{2, 3\}$ so the ramification at p is also considered). Since $N \equiv 1 \pmod{p}$ for all odd primes $p \in S_1 \cup \{3\}$ (by Lemma 1.15 and the fact that $p < q$), and $N \equiv 1 \pmod 8$, either N splits in K or it is ramified. If N splits in $K/\mathbb{Q}$, then the restriction of $\overline{\rho_p}$ to the decomposition group D_N would be reducible

(since $\overline{\rho_p}$ restricted to Gal_K is reducible), contradicting what we proved in the first paragraph. Otherwise, if N ramifies in $K/\mathbb{Q}$, the inertia group at the prime N has even order contradicting the fact that our local type at N is the induction from a quadratic unramified extension of an odd order character.   □

Definition 2.2 A continuous representation $\rho : \mathrm{Gal}_\mathbb{Q} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ has *large image* if its residual representation has non-solvable image.

All the congruences in Paso 3 and Paso 4 are modulo primes in $S_1 \cup \{2, 3\}$, and the ramification sets of the involved representations are contained in $S \cup \{2, 3\}$ so the hypotheses of Lemma 2.1 will always be satisfied. In particular, the good-dihedral prime N provides the large image hypothesis needed to apply the results of the first section.

Paso 3: killing the odd part of the level. Let $\{p_1, \ldots, p_r, N\}$ denote the set of primes where the system is ramified, p_1 being the smallest one (probably equal to 2). For each $i = 2, \ldots, r$, apply the following procedure:

- consider the reduction $\overline{\rho_{p_i}^{(i)}}$ of the p_i-th entry of the compatible system $\{\rho_\ell^{(i)}\}$,
- Take $\rho_{p_i}^{(i+1)}$ to be a minimal lift of $\overline{\rho_{p_i}^{(i)}}$ unramified at p_i (which exists by Theorem 1.9 (3), noting that the hypothesis on the restriction of the residual representation to $\mathrm{Gal}_{\mathbb{Q}(\zeta_{p_i})}$ being absolutely irreducible is fulfilled by Lemma 2.1),
- Make $\rho_{p_i}^{(i+1)}$ part of an almost strictly compatible system $\{\rho_\ell^{(i+1)}\}$ using Theorem 1.11. Now the ramification set of the new family does not contain the prime p_i (but is still contained in $S_1 \cup \{2, 3\}$ so Lemma 2.1 applies to the new family).

By Theorem 1.4 the representation $\rho_{p_i}^{(i)}$ is modular if and only if the representation $\rho_{p_i}^{(i+1)}$ is modular. In particular, modularity of the compatible system $\{\rho_\ell^{(i)}\}$ is equivalent to modularity of the compatible system $\{\rho_\ell^{(i+1)}\}$.

If p_1 is odd, we apply the same procedure at p_1, and end with a system unramified outside $\{N\}$. In such a case, go to Paso 5.

Paso 4: removing 2 from the level. This step follows the method of [17]. The following lemma will prove crucial. Recall that a local type at p is *Steinberg* if its inertial Weil-Deligne parameter is given by $(\omega_1 \oplus 1, \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix} \right))$.

Lemma 2.3 *Let $\{\rho_\ell\}$ be an almost strictly compatible system with Hodge-Tate weights $\{0, 1\}$ which is unramified at 3. If the Weil-Deligne representation at the prime 2 is (a twist of) Steinberg and ρ_3 has non-solvable residual image then either:*

- *there exists another almost strictly compatible system $\{\rho_\ell'\}$ with Hodge-Tate weights $\{0, 1\}$, having the same ramification set as $\{\rho_\ell\}$ such that the 3-adic members of the two systems are congruent, and whose Weil-Deligne type at the prime 2 is given by an order 3 character (in particular has trivial monodromy), or*
- *there exists another almost strictly compatible system $\{\rho_\ell'\}$ with Hodge-Tate weights $\{0, 1\}$, having the same set of odd ramified primes as $\{\rho_\ell\}$, unramified at the prime 2, such that the 3-adic members of the two systems are congruent.*

Furthermore, the modularity of one of the systems is equivalent to modularity of the other one.

Proof Under our hypotheses, $\rho_3|_{D_2} \simeq \left(\begin{smallmatrix} \chi_3 & * \\ 0 & 1 \end{smallmatrix} \right)$ (up to twist). The case when the image of inertia I_2 of the residual representation is trivial corresponds to the second claim of the statement. In

this case, take a lift of our representation which has the same inertial type at all odd primes, but which is unramified at 2 (whose existence is guaranteed by Theorem 1.9 (4)).

Suppose then that the residual image of the inertia group I_2 is non-trivial (so its image is generated by the matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ modulo 3). We follow the same strategy as we did to incorporate the good-dihedral prime to our family. We take a *niveau* 2 character χ' of order 3 (it corresponds to the cubic extension obtained by adding $\sqrt[3]{2}$ to the quadratic unramified extension $\mathbb{Q}_4$ of $\mathbb{Q}_2$) and consider its induction to $\mathrm{Gal}_{\mathbb{Q}_2}$. This corresponds to a local Galois representation $\tilde{\rho}$ of $\mathrm{Gal}_{\mathbb{Q}_2}$. Since the Galois group $\mathbb{Q}_4(\sqrt[3]{2})/\mathbb{Q}_2$ is isomorphic to S_3, the representation $\tilde{\rho}$ is just the standard 2-dimensional representation of S_3 (a representation having integral coefficients). The inertia group I_2 corresponds precisely to the subgroup generated by a 3-cycle, so the image by $\tilde{\rho}$ of I_2 is generated by the matrix $M = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$. The reduction modulo 3 of M has a unique eigenvector (namely the vector $(1, -1)$) and its cube is the identity, so the reduction of M is conjugate to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Then both representations ($\rho_3|_{D_2}$ and $\tilde{\rho}$) have the same reduction when restricted to the inertia subgroup I_2 and the traces of both reductions at a Frobenius element are zero.

Let ρ_3' be a crystalline weight 2 minimal lift of $\overline{\rho_3}$ whose local type at 2 matches that of $\tilde{\rho}$ (such a lift exists by Theorem 1.9 (4)). Let $\{\rho_\ell'\}$ be an almost strictly compatible system containing ρ_3' (whose existence is warranted by Theorem 1.11). The modularity of the system $\{\rho_\ell'\}$ is equivalent to that of $\{\rho_\ell\}$ by Theorem 1.4.   □

Remark 6 The previous lemma is needed because the existence of a crystalline lift of a Galois representation into $\mathrm{GL}_2(\overline{\mathbb{F}_2})$ is only proved when Serre's weight is 2 (see the hypotheses of Theorem 1.9 (1) and (2)). As explained in the proof of Theorem 9.1 of [17], if we start with an almost strictly compatible system of Galois representations, such that the residual representation at the prime 2 has non-solvable image and has Serre's weight $k(\overline{\rho_2}) = 4$, the workaround to get a crystalline lift is to first consider a weight 2 lift (and its compatible family) which is Steinberg at 2. Then apply Lemma 2.3 to this family to obtain a new family $\{\rho_\ell'\}$. If the family is unramified at 2, we are done (we have weight 2 and ramification only at odd primes). Oterwise, the new family must have $k(\overline{\rho_2'}) = 2$. The reason is that if the weight is 4 (the only other possibility), it corresponds to what is called a très ramifiée situation; in particular the residual representation is flat over a field extension of $\mathbb{Q}_2$ with even ramification index. On the other hand, the ramification type at 2 of the representation produced by the lemma was chosen so that it becomes flat over the extension $\mathbb{Q}_4(\sqrt[3]{2})$, which has ramification degree 3, leading to a contradiction.

Recall from Lemma 2.1 that our good-dihedral prime at N ensures non-solvable residual image at the prime 2. The procedure to remove the prime 2 from the compatible system's level consists of applying the following steps:

(1) If $k\left(\overline{\rho_2^{(r+1)}}\right) = 4$, take a minimal weight 2 lift $\rho_2^{(r+2)}$ with Steinberg type at 2 (which exists by Theorem 1.9 (2)) and make it part of an almost strictly compatible system $\{\rho_\ell^{(r+2)}\}$. Modularity of one compatible system is equivalent to modularity of the other by Theorem 1.5. Otherwise, $k\left(\overline{\rho_2^{(r+1)}}\right) = 2$ in which case go to step (3) (to unify notation, denote $r + 3$ the index).

(2) Now the compatible system $\{\rho_\ell^{(r+2)}\}$ has Steinberg Weil-Deligne type at two. Change its local type at 2 via a congruence at the prime 3 as explained in Lemma 2.3. This lemma gives an almost strictly compatible system $\{\rho_\ell^{(r+3)}\}$ whose local type at 2 is either unramified or it comes from an order 3 character. If the local type at 2 is unramified,

move to Paso 5.

Lemma 2.3 proves that modularity of the compatible system $\{\rho_\ell^{(r+2)}\}$ is equivalent to modularity of the compatible system $\{\rho_\ell^{(r+3)}\}$ in both cases.

(3) At this step, $k\left(\overline{\rho_2^{(r+3)}}\right) = 2$ (either because we came from (1) or by Remark 6 if we come from (2)). Let $\{\rho_\ell^{(r+4)}\}$ be an almost strictly compatible system containing a minimal crystalline weight 2 lift of $\overline{\rho_2^{(r+3)}}$ (whose existence is warranted by Theorem 1.9 (1)). Modularity of one family is equivalent to modularity of the other by Theorem 1.5.

Then we end with a system $\{\rho_\ell'\}$ unramified outside $\{N\}$.

Paso 5: killing the good-dihedral prime. The reduction modulo N of the system $\{\rho_\ell'\}$ lies in one of the following cases:

(1) The representation $\overline{\rho_N'}$ is reducible, in which case ρ_N' is modular by Theorem 1.6. End of the proof. Incidentally, as the reader may easily check, this is the only step in the whole proof where the fact that the compatible systems that we are considering are only known to be *almost* strictly compatible is relevant. Since the local parameter at N of this compatible system is ramified and we are assuming that $\overline{\rho_N'}$ is reducible, the conditions in the definition of "almost compatible systems" imply that we do not know the local behavior at N of ρ_N', we only know that it is a de Rham representation. Luckily, this is enough for Theorem 1.6 to hold.

(2) The representation is irreducible and not bad-dihedral. Take an almost strictly compatible system $\{\rho_\ell''\}$ containing a minimal crystalline lift of it (such a lift exists by Theorem 1.9 (3)). Observe that the ramification set S of this system is the empty set. Modularity of the system $\{\rho_\ell''\}$ is equivalent to that of $\{\rho_\ell'\}$ by Theorem 1.4.

(3) The representation is bad-dihedral. Serre's level of the representation $\overline{\rho_N'}$ equals 1 (as the system was unramified outside N), then this case cannot occur by Lemma 1.14 and Remark 5, because $N \equiv 1 \pmod 4$.

Paso 6: reduction of the weight. Move to the prime $p = 5$, where the same three possibilities can occur:

(1) The representation $\overline{\rho_5''}$ is reducible, in which case ρ_5'' is modular by Theorem 1.6. End of the proof.

(2) The representation $\overline{\rho_5''}$ has absolutely irreducible image, but it is bad-dihedral. Since our family has level 1, we know by Remark 5 that this is not possible, since $5 \not\equiv 3 \pmod 4$.

(3) The representation $\overline{\rho_5''}$ is irreducible and not bad-dihedral. If Serre's weight $k(\overline{\rho_5''}) = 2$ or 4, take an almost strictly compatible system $\{\tilde\rho_\ell\}$, whose ramification set is empty, containing a minimal crystalline lift of $\overline{\rho_5''}$, which exists by Theorem 1.9 (3) and has Hodge-Tate weights $(0, k(\overline{\rho_5''}) - 1)$. If Serre's weight is 6, take a weight 2 lift, with Steinberg Weil-Deligne type at $p = 5$, unramified outside 5, which exists by Theorem 1.9 (4) and make it part of an almost strictly compatible system $\{\tilde\rho_\ell\}$. In both cases modularity of the system $\{\rho_\ell''\}$ follows from that of the system $\{\tilde\rho_\ell\}$ by Theorem 1.4.

In the cases of weight 2 and 4, we look at the 3-adic representation $\tilde\rho_3$ and we reduce modulo 3. Theorem 1.1 implies that there are no irreducible residual Galois representations unramified outside 3, so the representation $\overline{\tilde\rho_3}$ is reducible. Furthermore, $\tilde\rho_3$ is ordinary by [2, Corollary 4,3], because the weight is either 2 or $4 = 3 + 1$, so the third hypothesis of Theorem 1.7 is satisfied. Since Serre's weight is not 3, the image of $\overline{\tilde\rho_3}|_{D_3}$ is non-trivial, hence Theorem 1.7 implies that $\tilde\rho_3$ is modular.

The last case to consider is when $\tilde{\rho}_5$ has weight 2, is Steinberg at 5 and unramified at any other prime. In this case, by [33, Proposition 9.4.1] the representation $\tilde{\rho}_5$ matches the 5-adic representation of an abelian variety $A/\mathbb{Q}$ of GL_2-type. Since $\tilde{\rho}_5$ is Steinberg at 5 and unramified at any other prime, the variety A is semistable and has good reduction outside 5 contradicting Theorem 1.2.

We have seen that by a combination of suitable modularity lifting theorems the modularity of the given representation was deduced by propagation from the base cases proved by Serre and Schoof and cases where the residual image is solvable. This concludes the proof of Serre's conjecture in the case of odd characteristic.

3 The case $p = 2$

The projective image is either solvable or non-solvable. Recall that all solvable subgroups of $PGL_2(\overline{\mathbb{F}_2})$ in the irreducible case have dihedral projective image (the S_4 case does not occur, while the A_4 case corresponds to a Borel subgroup over $\mathbb{F}_4$ see [17, Lemma 6.1]). The dihedral case was proven in [27]. If the image is non-solvable, take an odd weight two lift (guaranteed by Theorem 1.9(1) and (2)), make it part of an almost strictly compatible system and reduce modulo a prime greater than 3 and which is not in the ramification set of the compatible system (since the residual Serre weight is 2 under these conditions, applying Lemma 1.14 we see that $p > 2 \cdot 2 - 1 = 3$ is enough to avoid the bad-dihedral situation) to end in a situation covered before, namely: the representation is either reducible (so modularity follows from Theorem 1.6) or irreducible in odd characteristic hence it is covered by the cases of Serre's conjecture we have already solved, in which case modularity of the system follows from Theorem 1.4 since we know that the residual representation is not bad-dihedral.

References

1. Boston, N., Lenstra, H.W., Jr., Ribet, K.A.: Quotients of group rings arising from two-dimensional representations. C. R. Acad. Sci. **312**(4), 323–328 (1991)
2. Berger, L., Li, H., Zhu, H.J.: Construction of some families of 2-dimensional crystalline representations. Math. Ann. **329**(2), 365–377 (2004)
3. Deligne, P.: Formes modulaires et représentations l-adiques. In: Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363, volume 175 of Lecture Notes in Math., pages Exp. No. 355, 139–172. Springer, Berlin, (1971)
4. Dieulefait, L.V.: Existence of families of Galois representations and new cases of the Fontaine-Mazur conjecture. J. Reine Angew. Math. **577**, 147–151 (2004)

5. Dieulefait, L.: The level 1 weight 2 case of Serre's conjecture. Rev. Mat. Iberoam. **23**(3), 1115–1124 (2007)
6. Dieulefait, L.: Remarks on Serre's modularity conjecture. Manuscr. Math. **139**(1–2), 71–89 (2012)
7. Deligne, P., Serre, J.-P.: Formes modulaires de poids 1. Ann. Sci. École Norm. Sup. (4) **7**(507–530), 1974 (1975)
8. Edixhoven, B.: The weight in Serre's conjectures on modular forms. Invent. Math. **109**(3), 563–594 (1992)
9. Emerton, M.: Local-global compatibility in the p-adic Langlands programme for $GL_{2/\mathbb{Q}}$, preprint (2011)
10. Gee, T.: Automorphic lifts of prescribed types. Math. Ann. **350**(1), 107–144 (2011)
11. Yongquan, H., Tan, F.: The Breuil-Mézard conjecture for non-scalar split residual representations. Ann. Sci. Éc. Norm. Supér. (4) **48**(6), 1383–1421 (2015)
12. Khare, C.: Serre's modularity conjecture: the level one case. Duke Math. J. **134**(3), 557–589 (2006)
13. Kisin, M.: The Fontaine-Mazur conjecture for GL_2. J. Am. Math. Soc. **22**(3), 641–690 (2009)
14. Kisin, M.: Modularity of 2-adic Barsotti-Tate representations. Invent. Math. **178**(3), 587–634 (2009)
15. Kisin, M.: Moduli of finite flat group schemes, and modularity. Ann. Math. (2) **170**(3), 1085–1180 (2009)
16. Khare, C., Wintenberger, J.-P.: On Serre's conjecture for 2-dimensional mod p representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$. Ann. Math. (2) **169**(1), 229–253 (2009)
17. Khare, C., Wintenberger, J.-P.: Serre's modularity conjecture. I. Invent. Math. **178**(3), 485–504 (2009)
18. Khare, C., Wintenberger, J.-P.: Serre's modularity conjecture. II. Invent. Math. **178**(3), 505–586 (2009)
19. Langlands, R.P.: Base change for GL(2). Annals of Mathematics Studies, No. 96. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, (1980)
20. Pan, L.: The fontaine-mazur conjecture in the residually reducible case. To appear in JAMS, (2019)
21. Paškūnas, V.: On the Breuil-Mézard conjecture. Duke Math. J. **164**(2), 297–359 (2015)
22. Paškūnas, V.: On 2-dimensional 2-adic Galois representations of local and global fields. Algebra Number Theory **10**(6), 1301–1358 (2016)
23. Ribet, K.A.: On modular representations of $Gal(\overline{Q}/Q)$ arising from modular forms. Invent. Math. **100**(2), 431–476 (1990)
24. Ribet, K.A.: Report on mod l representations of $Gal(\overline{\mathbf{Q}}/\mathbf{Q})$. In Motives (Seattle, WA, 1991), volume 55 of Proc. Sympos. Pure Math., pages 639–676. Amer. Math. Soc., Providence, RI, (1994)
25. Ribet, K.A.: Images of semistable Galois representations. Pacific J. Math., (Special Issue):277–297, (1997). Olga Taussky-Todd: in memoriam
26. Ribet, K.A., Stein, W.A.: Lectures on Serre's conjectures. In Arithmetic algebraic geometry (Park City, UT, 1999), volume 9 of IAS/Park City Math. Ser., pages 143–232. Amer. Math. Soc., Providence, RI, (2001)
27. Rohrlich, D.E., Tunnell, J.B.: An elementary case of Serre's conjecture. Number Special Issue, pages 299–309. Olga Taussky-Todd: in memoriam (1997)
28. Schoof, R.: Abelian varieties over $\mathbb{Q}$ with bad reduction in one prime only. Compos. Math. **141**(4), 847–868 (2005)
29. Serre, J.-P.: Valeurs propres des opérateurs de Hecke modulo l. In: Journées Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, 1974), pages 109–117. Astérisque, Nos. 24–25. (1975)
30. Serre, J.-P.: Sur les représentations modulaires de degré 2 de $Gal(\overline{Q}/Q)$. Duke Math. J. **54**(1), 179–230 (1987)
31. Serre, J.-P.: Abelian l-adic representations and elliptic curves, volume 7 of Research Notes in Mathematics. A K Peters Ltd., Wellesley, MA, (1998). With the collaboration of Willem Kuyk and John Labute, Revised reprint of the 1968 original
32. Serre, J.-P.: Oeuvres/Collected papers. III. 1972–1984. Springer Collected Works in Mathematics. Springer, Heidelberg, (2013). Reprint of the 2003 edition [of the 1986 original MR0926691]
33. Snowden, A.: On two dimensional weight two odd representations of totally real fields. arXiv:0905.4266v1 [math.NT], (2009)
34. Skinner, C.M., Wiles, A.J.: Residually reducible representations and modular forms. Inst. Hautes Études Sci. Publ. Math. **89**, 5–126 (2000). (**1999**)
35. Tate, J.: Number theoretic background. In Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, pages 3–26. Amer. Math. Soc., Providence, R.I., (1979)
36. Tate, J.: The non-existence of certain Galois extensions of $\mathbf{Q}$ unramified outside 2. In Arithmetic geometry (Tempe, AZ, 1993), volume 174 of Contemp. Math., pages 153–156. Amer. Math. Soc., Providence, RI, (1994)
37. Tunnell, J.: Artin's conjecture for representations of octahedral type. Bull. Amer. Math. Soc. (N.S.) **5**(2), 173–175 (1981)
38. Tung, S.-N.: On the automorphy of 2-dimensional potentially semistable deformation rings of $G_{\mathbb{Q}_p}$. Algebra Number Theory **15**(9), 2173–2194 (2021)

39. Tung, S.-N.: On the modularity of 2-adic potentially semi-stable deformation rings. Math. Z. **298**(1–2), 107–159 (2021)
40. Taylor, R., Wiles, A.: Ring-theoretic properties of certain Hecke algebras. Ann. Math. (2) **141**(3), 553–572 (1995)
41. Wiles, A.: Modular elliptic curves and Fermat's last theorem. Ann. Math. (2) **141**(3), 443–551 (1995)