SERRE'S MODULARITY CONJECTURE

MICHAEL M. SCHEIN

1. Introduction

These notes are based on lectures given by the author at the winter school on Galois theory held at the University of Luxembourg in February 2012. Their aim is to give an overview of Serre's modularity conjecture and of its proof by Khare, Wintenberger, and Kisin [36] [37] [39], as well as of the results of other mathematicians that played an important role in the proof. Along the way we will remark on some recent work concerning generalizations of the conjecture. We have tried as much as possible to concentrate on giving a broad picture of the structure of the arguments and have ignored technical details in places. Some results are given incomplete statements, where we have chosen not to list technical hypotheses; we request the reader's forbearance.

Let F be a totally real number field. We will denote by G_F the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/F)$. It was shown in Prof. Böckle's lectures in this volume how, under some hypotheses, a Hilbert modular eigenform f over F gives rise to a compatible system $\{\rho_{f,v}\}$ of p-adic Galois representations; see [34] for the most general theorem. These representations are extracted from the cohomology of a suitable algebraic variety, and this construction is more or less the only method we have for obtaining p-adic Galois representations. Therefore the following question is of acute interest: given a Galois representation $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$, when is ρ *modular*, i.e. when does there exist a Hilbert modular eigenform f and a place $v|p$ of F such that $\rho \simeq \rho_{f,v}$?

This question is a very difficult one. We will split it into two questions, which are still very difficult, by introducing the notion of the reduction of a Galois representation. The following result is classical; the proof given here is attributed to N. Katz and appears in print, for instance, at the beginning of section 2 of [55].

Proposition 1.1. *Let G be a compact Hausdorff group, and let $\rho : G \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ be a continuous representation. Then ρ is equivalent to a representation ρ' such that $\rho'(G) \subset \mathrm{GL}_n(\mathcal{O}_L)$, where $\mathcal{O}_L$ is the ring of integers of some finite extension $L/\mathbb{Q}_p$.*

Proof. Since G is compact and Hausdorff, it admits a Haar measure μ; without loss of generality, $\mu(G) = 1$. Now,

$$\mathrm{GL}_n(\overline{\mathbb{Q}}_p) = \bigcup_{[L:\mathbb{Q}_p]<\infty} \mathrm{GL}_n(L)$$

The author was supported by grants from the German-Israeli Foundation for Scientific Research and Development and from the Pollack Foundation.

and hence

$$G = \bigcup_{[L:\mathbb{Q}_p]<\infty} \rho^{-1}(\mathrm{GL}_n(L)).$$

Since there are countably many finite extensions $L/\mathbb{Q}_p$, there must be some L such that $\mu(\rho^{-1}(\mathrm{GL}_n(L))) > 0$. Hence, $\rho^{-1}(\mathrm{GL}_n(L)) \subset G$ is a closed subgroup of finite index. Then $\rho^{-1}(\mathrm{GL}_n(\mathcal{O}_L))$ is an open subgroup of the compact group $\rho^{-1}(\mathrm{GL}_n(L))$, so it has finite index in it and thus in G. Let $g_1, \ldots, g_m$ be a collection of coset representatives for $\rho^{-1}(\mathrm{GL}_n(\mathcal{O}_L))$. Let $\Lambda \subset L^n$ be the lattice generated by $\rho(g_1)\mathcal{O}_L^n, \ldots, \rho(g_m)\mathcal{O}_L^n$. This is a lattice of maximal rank, so $\Lambda \simeq \mathcal{O}_L^n$. Furthermore, Λ is stable under the action of G. Let $T \in \mathrm{GL}_n(L)$ be a linear transformation that takes $\mathcal{O}_L^n$ to Λ, and set $\rho'(g) = T^{-1}\rho(g)T$. □

The proposition above applies, in particular, to continuous representations $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ such as we have been considering. If ρ is equivalent to $\rho' : G_F \to \mathrm{GL}_2(\mathcal{O}_L)$, then we define the reduction $\overline{\rho}$ to be the semisimplification of the composition $\tilde{\rho}' : G_F \xrightarrow{\rho'} \mathrm{GL}_2(\mathcal{O}_L) \to \mathrm{GL}_2(k_L) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$, where k_L is the residue field of L and the inclusion $k_L \hookrightarrow \overline{\mathbb{F}}_p$ is induced from $L \hookrightarrow \overline{\mathbb{Q}}_p$. In other words, $\overline{\rho}$ is the direct sum of the Jordan-Hölder constituents of $\tilde{\rho}'$. This definition is independent of all choices, and we call $\overline{\rho}$ the reduction modulo p of ρ.

Remark 1.2. Throughout these notes, except for Section 4, we will usually use $\overline{\rho}$ to denote a mod p Galois representation. The bar simply serves to emphasize that we are dealing with a mod p representation. It does not necessarily mean that we have any p-adic representation ρ in mind, of which $\overline{\rho}$ is to be the reduction.

We say that a mod p Galois representation $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is *modular* if it "arises from geometry" in a way that will be made precise in the next section. If $F = \mathbb{Q}$, then $\overline{\rho}$ is modular if and only if there exists a modular eigenform f such that $\overline{\rho} \simeq \overline{\rho_{f,p}}$, while for larger totally real fields we will require a somewhat more subtle notion of modularity. If a p-adic Galois representation ρ is modular, then its reduction modulo p will be modular as well. We will consider two questions:

(1) Let $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be a p-adic Galois representation. Suppose that $\overline{\rho}$ is modular. Is ρ modular?
(2) Let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a mod p Galois representation. When is $\overline{\rho}$ modular?

It is clear that if we knew complete answers to both of these questions, their union would resolve the question of when a general p-adic representation is modular. Affirmative responses to the first question are known in a variety of different cases; results of this type are called modularity lifting theorems. A conjectural response to the second question is given by Serre's modularity conjecture and its generalizations. Serre's original conjecture, covering the case of $F = \mathbb{Q}$, is now a theorem of Khare, Wintenberger, and Kisin. However, as we shall see, even if we do not know whether a mod p Galois representation is modular, we can say a lot about the Hilbert modular forms f that it could come from if it were modular. The research towards resolving each of these two questions is tightly interconnected with work concerning the other, as shall become evident in these notes.

1.1. **Acknowledgements.** The author is very grateful to Sara Arias de Reyna, Lior Bary-Soroker, and Gabor Wiese, the organizers of the Winter School on Galois Theory held in Luxembourg in February 2012, for inviting him to present these lectures. He is grateful to the audience for their stimulating questions, to Tommaso Centeleghe and Nicolas Billerey for allowing him to make use of the notes they took in his lectures, and to the anonymous referee for a thorough reading. The exposition given here has in places drawn on expositions of similar material given elsewhere, such as the lectures by Toby Gee, Richard Taylor, and Teruyoshi Yoshida at an MSRI workshop on modularity in 2006, of Richard Taylor at the summer school on Serre's conjecture at Luminy in 2007, and those by Fred Diamond at the Galois Trimester at the Institut Henri Poincaré in Paris in 2010. All errors and inaccuracies are, of course, entirely the responsibility of the author.

2. Statement of Serre's modularity conjecture

2.1. **The classical conjecture.** If we fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$, then complex conjugation is a well-defined element $c \in G_{\mathbb{Q}}$. Since it is an involution, any Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ must send c to a matrix with determinant ± 1. We say that $\overline{\rho}$ is *odd* if $\det \overline{\rho}(c) = -1$. Similarly, if F is a totally real field with $[F : \mathbb{Q}] = d$, then the d embeddings $F \hookrightarrow \mathbb{R}$ induce d complex conjugation automorphisms $c_1, \dots, c_d \in G_F$. We say that $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is *totally odd* if $\det \overline{\rho}(c_i) = -1$ for each $i = 1, \dots, d$.

The original statement of Serre's conjecture, which essentially dates back to the 1960's, appeared in some cases in [52], and was properly published only in [54], is the following.

Conjecture 2.1. Let $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a mod p Galois representation. If $\overline{\rho}$ is continuous, irreducible, and odd, then there exists a modular eigenform $f \in S_k(\Gamma_1(N))$, for some weight k and level N, such that $\overline{\rho} \simeq \overline{\rho_{f,p}}$.

This statement has a natural generalization for totally real fields:

Conjecture 2.2 (Weak Serre conjecture)**.** If F is a totally real field and $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is continuous, irreducible, and totally odd, then it is modular.

While this conjecture is already very powerful, one generally wants to know in what weights and levels to look for a modular form giving rise to $\overline{\rho}$. In fact, Serre gave such a strengthened version of his own conjecture, in which he specified a minimal weight $k(\overline{\rho})$ and level $N(\overline{\rho})$ such that there should exist a modular eigenform $f \in S_{k(\overline{\rho})}(\Gamma_1(N(\overline{\rho})))$ with $\overline{\rho} \simeq \overline{\rho_{f,p}}$. We shall not give the explicit formulae for $k(\overline{\rho})$ and $N(\overline{\rho})$ here, but the reader will be able to extract them from our statement of a generalized conjecture later on.

2.2. **Serre weights.** One of the most basic and useful facts about mod p representation theory, and one which is responsible for much of the difference in flavor between it and representation theory in characteristic zero, is the following. A proof may be found in [19].

Proposition 2.3. *Let G be a profinite group, let $P \subset G$ be a normal pro-p-group, and let $\tau : P \to \mathrm{GL}(V)$ be a continuous finite-dimensional representation of P on an $\overline{\mathbb{F}}_p$-vector space V. Let $V^P = \{v \in V : \forall a \in P, \tau(a)v = v\}$. Then $V^P \neq \{0\}$. Moreover, if τ is irreducible, then $V^P = V$.*

Proof. Clearly we may restrict to the case where V is irreducible as a G-module. Since τ is continuous and hence has finite image, we in fact have $\tau : G \to \mathrm{GL}(W)$, where W is a finite-dimensional vector space over a finite extension of $\mathbb{F}_p$. Since P is normal in G, clearly $W^P \subset W$ is a sub-G-module. Since P is a pro-p-group, every non-trivial orbit of the P-action on W must have cardinality divisible by p. However, the cardinality of W is itself divisible by p. Hence, $|W^P| \neq 1$, and thus $W^P = W$. □

Definition 2.4. Let F be a number field, let v be a place of F, and let $k_v = \mathcal{O}_F/v$ be the residue field at v.

(1) A *Serre weight* is an irreducible $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_2(\mathcal{O}_F/p)$.
(2) A *local Serre weight at v* is an irreducible $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_2(k_v)$.

Since $\mathrm{GL}_2(\mathcal{O}_F/p)$ is a finite group, there are only finitely many Serre weights for any number field F. Moreover, suppose that the ideal $p\mathcal{O}_F$ decomposes into prime factors as $p\mathcal{O}_F = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_r^{e_r}$. Then by the Chinese remainder theorem, $\mathrm{GL}_2(\mathcal{O}_F/p) = \mathrm{GL}_2(\mathcal{O}_F/\mathfrak{p}_1^{e_1}) \times \cdots \times \mathrm{GL}_2(\mathcal{O}_F/\mathfrak{p}_r^{e_r})$. For each $1 \leq i \leq r$, let k_i denote the residue field $\mathcal{O}_F/\mathfrak{p}_i$. Since the kernel of the natural projection

$$\mathrm{GL}_2(\mathcal{O}_F/p) \to \mathrm{GL}_2(k_1) \times \cdots \times \mathrm{GL}_2(k_r)$$

is a p-group, all Serre weights factor through this projection by Proposition 2.3. It follows that all Serre weights have the form

$$\sigma = \bigotimes_{v|p} \sigma_v,$$

where σ_v is a local Serre weight at v. The representation theory of GL_2 of a finite field is well known [27], and the distinct local Serre weights at v are precisely the following:

$$\sigma_v = \bigotimes_{\tau : k_v \hookrightarrow \overline{\mathbb{F}}_p} \det{}^{w_\tau} \otimes (\mathrm{Sym}^{r_\tau} k_v^2 \otimes_{k_v,\tau} \overline{\mathbb{F}}_p).$$

Here, k_v^2 is the standard action of $\mathrm{GL}_2(k_v)$ on a two-dimensional vector space over k_v, whereas $0 \leq r_\tau \leq p-1$ for all τ and $0 \leq w_\tau \leq p-1$ for all τ, with the stipulation that the w_τ are not all $p-1$.

What is the connection between these Serre weights and the modularity of Galois representations? Observe, first of all, that if $F = \mathbb{Q}$, then the Serre weights are just $\det^w \otimes \mathrm{Sym}^r \overline{\mathbb{F}}_p^2$, where $0 \leq r \leq p-1$ and $0 \leq w \leq p-2$. Let $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a mod p Galois representation, and let $\mathbf{T}$ be the Hecke algebra generated over $\mathbb{Z}$ by the Hecke operators T_l for $l \nmid pN(\overline{\rho})$. Define the maximal ideal $\mathfrak{m}_{\overline{\rho}} \subset \mathbf{T}$ to be the kernel of the map

$$\begin{aligned} \mathbf{T} &\to \overline{\mathbb{F}}_p \\ T_l &\mapsto \mathrm{tr}\overline{\rho}(\mathrm{Frob}_l). \end{aligned}$$

The following was observed by Ash and Stevens [2].

Proposition 2.5. *Let $k \geq 2$. A Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is modular of level N and weight k if and only if $H^1(\Gamma_1(N), \mathrm{Sym}^{k-2}\overline{\mathbb{F}}_p^2)_{\mathfrak{m}_{\overline{\rho}}} \neq 0$.*

We remark that by Theorem 3.4 of [18], any collection of eigenvalues of $\mathbf{T}$ for which there exists an eigenform of some weight has an eigenform of weight at most $p+1$. Thus we do not lose generality by concentrating on forms associated to Serre weights. The twist by a power of the determinant that can occur in the Serre weight corresponds to Nebentypus.

2.3. **Modularity.** Inspired by the observations above, we will formulate a new definition of modularity. We use quaternionic Shimura curves in place of modular curves.

Recall that F is our fixed totally real field and that $d = [F : \mathbb{Q}]$. Let B/F be a quaternion algebra that splits at exactly one real place and at all places above p (note that we make no conditions at places away from p, so there are no parity issues). In other words, we are able to fix isomorphisms

$$\begin{aligned} B \otimes \mathbb{R} &\simeq M_2(\mathbb{R}) \times \mathbb{H}^{d-1} \\ B \otimes \mathbb{Q}_p &\simeq M_2(F \otimes \mathbb{Q}_p). \end{aligned}$$

Denote by $\mathbb{A}$ the adeles of $\mathbb{Q}$, and if S is a finite set of places of $\mathbb{Q}$ then we set $\mathbb{A}^S$ to be the adeles away from S. Define the group $G = \mathrm{Res}_{F/\mathbb{Q}} B^*$, and let $U \subset G(\mathbb{A}^\infty)$ be an open compact subgroup of the form $U = U_p \times U^p$, where $U^p \subset G(\mathbb{A}^{\infty,p})$ and

$$U_p = \ker \left(\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v}) \to \prod_{v|p} \mathrm{GL}_2(k_v) \right).$$

Then the Shimura curve

$$M_U(\mathbb{C}) = G(\mathbb{Q}) \backslash (G(\mathbb{A}^\infty) \times (\mathbb{C} - \mathbb{R}))/U$$

has a model over F. Set $V \subset G(\mathbb{A}^\infty)$ to be the following open compact subgroup:

$$V = \left(\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v}) \right) \times U^p.$$

It follows from a simple modification of an argument of Carayol [16] that if U^p is sufficiently small (see [46] for the precise definition of "sufficiently small") then the natural map $M_U \to M_V$ is a Galois cover of Shimura curves with Galois group $V/U \simeq \prod_{v|p} \mathrm{GL}_2(k_v)$.

Definition 2.6. Let F be a totally real field, let σ be a Serre weight, and let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a mod p Galois representation. We say that $\overline{\rho}$ is *modular of weight* σ if there exist a quaternion algebra B/F and an open compact subgroup $U \subset G(\mathbb{A}^\infty)$ as above such that

$$\overline{\rho} \subset (\mathrm{Pic}^0(M_U)[p] \otimes \sigma)^{\prod_{v|p} \mathrm{GL}_2(k_v)}. \tag{1}$$

The reader is referred to the introduction of [14] and to Propositions 2.5 and 2.10 of the same paper for a discussion of why the naive notion of modularity (a Galois representation is modular if it arises from a Hilbert modular form of weight $(k_1, \ldots, k_d)$) is insufficient and of the connection between this naive modularity and the notion of being modular of a certain Serre weight.

Remark 2.7. It is not hard to show, using the Eichler-Shimura relation, that the condition (1) is equivalent to the following:

$$\overline{\rho}^{\vee} = \mathrm{Hom}(\overline{\rho}, \overline{\mathbb{F}}_p) \subset H^1_{\acute{e}t}(M_V \otimes \overline{\mathbb{Q}}, \mathcal{L}_\sigma),$$

where $\mathcal{L}_\sigma$ is the mod p local system associated to σ. While this fact is well-known, the author does not know of a clear proof in the literature. See Proposition 2.6 of [46].

Remark 2.8. The reader may wonder why we assume throughout that F is a totally real field. Why do we not work with arbitrary number fields? This question is strengthened by the fact that the proofs of many of the results about Galois representations $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ that are described below do not use global information about $\overline{\rho}$, but only local information about the restrictions of $\overline{\rho}$ to the decomposition subgroups at various places. It is very difficult to study modular Galois representations over arbitrary number fields F because of the lack of nice algebraic varieties over F that store automorphic data and in whose cohomology we could look for our Galois representations. The theory of Shimura varieties over totally real fields has no good analogue over general number fields. A number of mathematicians have studied the modularity for quadratic imaginary fields: see, for example, [28], [57], and [6]. In this case one can translate the problem to Siegel modular surfaces. See [7] and the end of [15] for modularity lifting theorems in the quadratic imaginary case.

2.4. **Serre's weight conjecture.** For every place v of F dividing p, denote by k_v the corresponding residue field and let its cardinality be $q_v = p^{f_v}$. Recall that G_{F_v} can be embedded non-canonically into G_F as follows. For every finite extension L/F, we choose a place v_L of L such that $v_L|v$ and such that if $L \subset L'$, then $v_{L'}|v_L$. Now let $G_v = \{\alpha \in G_F : \forall L/F, \alpha(v_L) = v_L\}$. This G_v is called a decomposition subgroup at v, and it is easy to show that $G_v \simeq G_{F_v}$. If we replace the system $\{v_L\}_L$ by a different compatible system of places, we will get a subgroup conjugate to G_v. Inside G_v, we have the inertia subgroup I_v consisting of all $\alpha \in G_v$ such that for each L/F, the induced automorphism of $\mathcal{O}_L/v_L$ is trivial. The wild inertia P_v is the pro-p-Sylow subgroup of I_v. Note that the isomorphism $G_v \simeq G_{F_v}$ induces isomorphisms $I_v \simeq \mathrm{Gal}(\overline{\mathbb{Q}}_p/F_v^{nr})$ and $P_v \simeq \mathrm{Gal}(\overline{\mathbb{Q}}_p/F_v^{tr})$, where F_v^{nr} and F_v^{tr} are the maximal unramified and maximal tamely ramified extensions of F_v, respectively. In particular, since F_v^{tr} is a Galois extension of F_v, we see that P_v is a normal subgroup of G_v. We have an exact sequence

$$1 \to I_v \to G_v \to \mathrm{Gal}(\overline{\mathbb{F}}_p/k_v) \to 1. \tag{2}$$

Now, let $r \geq 1$ and consider an embedding of fields $\tau : \mathbb{F}_{p^r} \hookrightarrow \overline{\mathbb{F}}_p$. Fix a uniformizer $\pi_v \in \mathcal{O}_{F_v}$ and define $\psi_\tau : I_v \to \overline{\mathbb{F}}_p^*$ to be the following composition of homomorphisms:

$$I_v \simeq \mathrm{Gal}(\overline{\mathbb{Q}}_p/F_v^{nr}) \to \mathrm{Gal}(F_v^{nr}(\sqrt[p^r-1]{\pi_v})/F_v^{nr}) \simeq \mathbb{F}_{p^r}^* \xrightarrow{\tau} \overline{\mathbb{F}}_p^*.$$

Let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a mod p Galois representation. Then P_v acts trivially on $(\overline{\rho}_{|G_v})^{ss}$ by Proposition 2.3. It follows that $((\overline{\rho}_{|G_v})^{ss})_{|I_v}$ factors through a representation of the abelian group I_v/P_v and thus is a sum of characters $\varphi \oplus \varphi'$. If Frob_v is a lift to G_v via (2) of the map $x \mapsto x^{q_v}$, which is a topological generator of $\mathrm{Gal}(\overline{\mathbb{F}}_p/k_v)$, then Frob_v acts on $((\overline{\rho}_{|G_v})^{ss})_{|I_v}$ by conjugation. Therefore, $\{\varphi, \varphi'\} = \{(\varphi)^{q_v}, (\varphi')^{q_v}\}$. Let $[k_v' : k_v] = 2$. We have two possibilities:

(1) If $\overline{\rho}_{|G_v}$ is reducible, then $\overline{\rho}_{|I_v} \sim \begin{pmatrix} \varphi & * \\ 0 & \varphi' \end{pmatrix}$, where φ and φ' each factor through k_v.

(2) If $\overline{\rho}_{|G_v}$ is irreducible, then $\overline{\rho}_{|I_v} \sim \begin{pmatrix} \varphi & 0 \\ 0 & \varphi' \end{pmatrix}$, where φ and φ' factor through $(k_v')^*$ and we have $\varphi' = \varphi^{q_v}$ and $\varphi = (\varphi')^{q_v}$.

We are finally ready to state Serre's conjecture. Given a mod p Galois representation $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$, our aim is to define a set $W(\overline{\rho})$ of Serre weights and then conjecture:

Conjecture 2.9 (Strong Serre conjecture)**.** If $\overline{\rho}$ is continuous, irreducible, and totally odd, then it is modular. Moreover, if σ is a Serre weight, then $\overline{\rho}$ is modular of weight σ if and only if $\sigma \in W(\overline{\rho})$.

A crucial property of the set $W(\overline{\rho})$ is that it is defined locally. This means that for each place $v|p$ we will specify a set $W_v(\overline{\rho})$ of local Serre weights at v (recall Definition 2.4) and then define

$$W(\overline{\rho}) = \left\{ \sigma = \bigotimes_{v|p} \sigma_v : \forall v|p, \sigma_v \in W_v(\overline{\rho}) \right\}. \tag{3}$$

Given a place $v|p$, let e_v be the ramification index of $v|p$, so that $[F_v : \mathbb{Q}_p] = e_v f_v$. Let S_v be the collection of all field embeddings $k_v \hookrightarrow \overline{\mathbb{F}}_p$. The definition of $W_v(\overline{\rho})$ involves several cases, corresponding to the cases above.

(1) If $\overline{\rho}_{|G_v}$ is reducible and semisimple (i.e. the direct sum of two one-dimensional representations) then a local Serre weight at v

$$\sigma_v = \bigotimes_{\tau \in S_v} \left(\det{}^{w_\tau} \otimes \mathrm{Sym}^{r_\tau} k_v^2 \otimes_{k_v,\tau} \overline{\mathbb{F}}_p \right) \tag{4}$$

is contained in $W_v(\overline{\rho})$ if and only if there exists a subset $A \subset S_v$ and an integer $0 \le \delta_\tau \le e_v - 1$ for each $\tau \in S_v$, such that

$$\overline{\rho}_{|I_v} \sim \begin{pmatrix} \prod_{\tau \in A} \psi_\tau^{w_\tau + r_\tau + 1 + \delta_\tau} \prod_{\tau \notin A} \psi_\tau^{w_\tau + e_v - 1 - \delta_\tau} & 0 \\ 0 & \prod_{\tau \in A} \psi_\tau^{w_\tau + e_v - 1 - \delta_\tau} \prod_{\tau \notin A} \psi_\tau^{w_\tau + r_\tau + 1 + \delta_\tau} \end{pmatrix}. \tag{5}$$

(2) If $\overline{\rho}_{|G_v}$ is irreducible, then σ_v as in (4) is contained in $W_v(\overline{\rho})$ if and only if for each $\tau \in S_v$ there exists an integer $0 \le \delta_\tau \le e_v - 1$ and a lift $\tilde{\tau} : k_v' \hookrightarrow \overline{\mathbb{F}}_p$ of τ (recall that k_v' is a quadratic extension of k_v) such that

$$\overline{\rho}_{|I_v} \sim \begin{pmatrix} \varphi & 0 \\ 0 & \varphi^{q_v} \end{pmatrix},$$

where

$$\varphi = \prod_{\tau \in S_v} \psi_{\tilde{\tau}}^{(q_v+1)w_\tau + r_\tau + 1 + \delta_\tau + q_v(e_v - 1 - \delta_\tau)}. \tag{6}$$

(3) If $\overline{\rho}_{|G_v}$ is indecomposable, i.e. reducible but not semisimple, then

$$\overline{\rho}_{|G_v} \sim \begin{pmatrix} \varphi & * \\ 0 & \varphi' \end{pmatrix},$$

where $*$ corresponds to an element $c_{\overline{\rho}} \in \mathrm{Ext}^1(\varphi', \varphi) \simeq H^1(G_v, \varphi(\varphi')^{-1})$. For each local Serre weight σ_v and each subset $A \subset S_v$ one defines a subspace $L_{A,\sigma_v} \subset H^1(G_v, \varphi(\varphi')^{-1})$. The definition of L_{A,σ_v} is intricate, and we omit it here. Then a local Serre weight as in (4) is contained in $W_v(\overline{\rho})$ if and only if there exist a subset $A \subset S_v$ and an integer $0 \leq \delta_\tau \leq e_v - 1$ for each $\tau \in S_v$ such that $((\overline{\rho}_{|G_v})^{ss})_{|I_v}$ has the form of (5) and in addition $c_{\overline{\rho}} \in L_{A,\sigma_v}$.

As we mentioned above, this conjecture was first stated in the case of $F = \mathbb{Q}$ by Serre several decades ago, in a slightly different language. We will now indicate how to extract the minimal weight $k(\overline{\rho})$ from the conjecture formulated above. It follows from Proposition 2.5 and the definition of modularity that $\overline{\rho}$ arises from a modular form of weight $2 \leq k \leq p+1$ (and trivial Nebentypus) precisely when it is modular of the Serre weight $\mathrm{Sym}^{k-2}\overline{\mathbb{F}}_p^2$. Furthermore, it is easily shown by an argument with Eisenstein ideals that, for arbitrary $k \geq 2$, we have $H^1(\Gamma_1(N), \mathrm{Sym}^{k-2}\overline{\mathbb{F}}_p^2)_{\mathfrak{m}_{\overline{\rho}}} \neq 0$ if and only if $H^1(\Gamma_1(N), W)_{\mathfrak{m}_{\overline{\rho}}} \neq 0$ for some Jordan-Hölder constituent W of $\mathrm{Sym}^{k-2}\overline{\mathbb{F}}_p$. Thus, the minimal weight $k(\overline{\rho})$ conjectured by Serre is just the minimal $k \geq 2$ such that $\overline{\rho}$ is modular (in the sense of Conjecture 2.9) of some Jordan-Hölder constituent of $\mathrm{Sym}^{k-2}\overline{\mathbb{F}}_p^2$. This $k(\overline{\rho})$ is easily computed using the following decomposition.

Lemma 2.10. *Let $m > p-1$. Then the following holds, where "ss" denotes semisimplification:*

$$(\mathrm{Sym}^m \overline{\mathbb{F}}_p^2)^{ss} = (\det \otimes \mathrm{Sym}^{m-p-1}\overline{\mathbb{F}}_p^2)^{ss} \oplus \mathrm{Sym}^r \overline{\mathbb{F}}_p^2 \oplus \det{}^r \otimes \mathrm{Sym}^{p-1-r}\overline{\mathbb{F}}_p^2,$$

where r is the unique integer in the range $0 \leq r \leq p-2$ such that $r \equiv m$ modulo $p-1$.

The first generalization of Serre's conjecture beyond the original case of $F = \mathbb{Q}$, to the case of totally real fields F in which p is unramified, was by Buzzard, Diamond, and Jarvis [14] and circulated for nearly a decade before their paper appeared in print. A conjecture for arbitrary totally real fields F but semisimple $\overline{\rho}_{|G_v}$ (i.e. the first two of the three cases above) was made by the author; see [45], Theorems 2.4 and 2.5. At around the same time, Herzig [31] made a conjecture for n-dimensional representations $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ for arbitrary n, under the assumption that $\overline{\rho}_{|G_p}$ is semisimple. These conjectures were later restated by other authors [3] [21] to cover arbitrary $\overline{\rho}$ at the cost of becoming less explicit: if σ is a local Serre weight at $v|p$, then $\sigma \in W_v(\overline{\rho})$ is conjectured to be equivalent to the existence of a p-adic lift of $\overline{\rho}$ with some specified local properties. Such opacity appears already in the indecomposable case of the conjecture of [14], in the form of the spaces L_{A,σ_v}.

2.5. **The level in Serre's conjecture.** Before proceeding, we will say a few words about the level in Serre's conjecture. For this, the notion of the Artin conductor of a Galois representation is crucial. Roughly speaking, the Artin conductor $\mathfrak{n}(\overline{\rho})$ of $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is an ideal $\mathfrak{n}(\overline{\rho}) = \prod_v v^{a_v}$ of $\mathcal{O}_F$, where v runs over the finite places of F and a_v measures the ramification of $\overline{\rho}$ at v. If $\overline{\rho}$ is unramified at v, i.e. if $I_v \subset \ker \overline{\rho}$, then $a_v = 0$. Otherwise, the exponent a_v reflects how far down the upper ramification filtration of I_v one has to go to find a subgroup contained in $\ker \overline{\rho}$. More precisely, let L/F be a finite Galois extension such that $\overline{\rho}$ factors through $\mathrm{Gal}(L/F)$; this exists because $\overline{\rho}$ is continuous and G_F is compact, hence $\overline{\rho}$ has finite image. if $v \nmid p$, then let v_L be a place of L lying above v, and for each $i \geq 0$ let

$G_i \subset G_{v_L} \simeq \mathrm{Gal}(\overline{L_{v_L}}/L_{v_L})$ be the i-th ramification subgroup; see, for instance, Chapter IV of [51] for the definition. If $V_{\overline{\rho}}$ is a two-dimensional $\overline{\mathbb{F}}_p$-vector space on which $\overline{\rho}$ acts, then we define

$$a_v = \sum_{i=0}^{\infty} \frac{1}{[I_{v_L} : G_i]} \dim_{\overline{\mathbb{F}}_p}(V_{\overline{\rho}}/V_{\overline{\rho}}^{G_i}),$$

where $V_{\overline{\rho}}^{G_i}$ is the subspace of G_i-invariants. More details about the Artin conductor may be found in Chapter VI of [51].

When $F = \mathbb{Q}$, any modular Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ arises from a modular form of weight N that is prime to p, and Serre specified a minimal such weight $N(\overline{\rho})$: this is just the prime-to-p part of the Artin conductor of $\overline{\rho}$. If F is an arbitrary totally real field, we are no longer so lucky. We can no longer expect $\overline{\rho}$ to always arise in level prime to p (this would correspond to $U_p = \prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})$ in Section 2.3); see the introduction of [14] for a discussion of why not. However, the prime-to-p part of the level may conjecturally always be taken to be the prime-to-p part of $\mathfrak{n}(\overline{\rho})$, which means that we may take $U^p = \prod_{v \nmid p} U_1(v^{a_v})$, where the group $U_1(v^{a_v}) \subset \mathrm{GL}_2(\mathcal{O}_{F_v})$ consists of all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_{F_v})$ such that $a - 1 \in v^{a_v}$ and $c \in v^{a_v}$.

3. Weak Serre implies strong Serre

The "weak" version of Serre's modularity conjecture (Conjecture 2.2) is actually a very strong statement. It has been proved only in the case $F = \mathbb{Q}$. This was achieved by Khare, Wintenberger, and Kisin [36], [37], [39], and we will sketch some of their methods below. As we will see at the end of these notes, some of these methods fail crucially whenever F is a totally real field with $[F : \mathbb{Q}] > 1$, so that a major new idea is needed for any substantial further progress on the conjecture.

In the meantime, much research in the area has focused on proving, in various settings, that the weak version of Serre's conjecture (Conjecture 2.2) implies the strong version (Conjecture 2.9), in other words that if $\overline{\rho}$ is modular, then it is modular of precisely the predicted Serre weights. In the case $F = \mathbb{Q}$, this fact has essentially been known since the late 1970's except for a few cases where $p = 2$ and was an important input in the proof of Serre's conjecture (the stubborn cases with $p = 2$ were also settled by Khare, Wintenberger, and Kisin's work). It was proved by Deligne for $\overline{\rho}_{|G_p}$ reducible and by Fontaine for $\overline{\rho}_{|G_p}$ irreducible. Fontaine's work was never published, and a (somewhat different) proof of the theorem first appeared in print in [18].

We will now mention some of the "weak Serre implies strong Serre" theorems that have been proved in recent years. The conjecture of [14], for F in which p is unramified, was proved by Gee [23] for most Serre weights by deformation-theoretic methods; see also [46] for a more geometric proof of most cases of one direction of this conjecture for locally irreducible $\overline{\rho}$. We will say more about the methods of these papers in the remainder of this section. The remaining cases of the Buzzard-Diamond-Jarvis conjecture were attacked in a series of papers by Gee and coauthors, until it was finally proved completely in [24]. The results of [46] were

extended in [45] to cases where p ramifies in F. Moreover, the conjecture of [45] was proved, for most cases where p is totally ramified in F, by Gee and Savitt [26]. More cases in the related, but not equivalent, unitary setting were resolved in [25]. Some non-totally ramified cases with $e = f = 2$ were addressed by R. Smith in his Ph.D. thesis at the University of Arizona, but it seems that new ideas are needed to make substantial further progress.

3.1. **A sketch of Gee's argument.** The claim that weak Serre implies strong Serre consists, of course, of two claims in opposite directions:

(1) If $\overline{\rho}$ is modular of weight σ, then $\sigma \in W(\overline{\rho})$.
(2) If $\overline{\rho}$ is modular of some weight and $\sigma \in W(\overline{\rho})$, then $\overline{\rho}$ is modular of weight σ.

The most successful method for proving "weak Serre implies strong Serre" has been that of relating the modularity of $\overline{\rho}$ to the existence of lifts of $\overline{\rho}$ with some specific local properties and then using p-adic Hodge theory to investigate the existence of such lifts.

An important breakthrough was Gee's paper [23], which proved the following result. Its statement involves the following definition: a local Serre weight at v is said to be regular if it is of the form (4) with $1 \leq r_\tau \leq p-3$ for all $\tau \in S_v$. A Serre weight $\sigma = \bigotimes_{v|p} \sigma_v$ is called regular if all the σ_v are regular.

Theorem 3.1. *Suppose that $p \geq 5$ is unramified in the totally real field F, that $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is modular of some Serre weight, and that $\overline{\rho}_{|G_{F(\zeta_p)}}$ is irreducible. Suppose that σ is a regular Serre weight. If $\overline{\rho}$ is modular of weight σ, then $\sigma \in W(\overline{\rho})$. Conversely, if $\sigma \in W(\overline{\rho})$ and some further technical conditions are satisfied at places $v|p$ where $\overline{\rho}_{|G_v}$ is reducible and σ_v arises from $A = S_v$ or $A = \varnothing$ in the recipe of Section 2.4, then $\overline{\rho}$ is modular of weight σ.*

We will give a very brief sketch of part of the argument of [23] to illustrate the method; the reader is referred to that paper (and the papers cited in it!) for further details. It should be noted that Gee works with a different notion of modularity than the one given above; he uses definite quaternion algebras, rather than indefinite ones. It is not possible to translate theorems directly from one setting to the other, but his local arguments can be translated. In this section, we will assume that p is unramified in the totally real field F. Let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous, irreducible, and totally odd mod p Galois representation. Recall that in Section 2.4 we defined a set $W_v(\overline{\rho})$ of local Serre weights at v for each $v|p$.

Let $v|p$; for the purposes of this section, we will say that a p-adic representation η_v of $\mathrm{GL}_2(k_v)$ is *good* if it is either an irreducible principal series or supercuspidal; in other words, η_v is any irreducible p-adic representation of $\mathrm{GL}_2(k_v)$ that is not one-dimensional or special. We regard η_v as a representation of the group $\mathrm{GL}_2(\mathcal{O}_{F_v})$ via the obvious inflation. Then η_v has an associated inertial type τ_v, namely a p-adic representation of I_v with the property that for any irreducible p-adic representation π of $\mathrm{GL}_2(F_v)$, we have $\eta_v \subset \pi_{|GL_2(\mathcal{O}_{F_v})}$ if and only if $\mathrm{LLC}(\pi)_{|I_v} \simeq \tau_v$. Here $\mathrm{LLC}(\pi)$ is the Weil-Deligne representation associated to π by the local Langlands correspondence. See Henniart's appendix to [11] for an exposition of the theory of types for GL_2.

Proposition 3.2 ([23], Lemma 2.1.4)**.** *Let $\overline{\rho}$ be as above, and for each $v|p$ let η_v be a good representation as above. Then $\overline{\rho}$ is modular of some Serre weight $\sigma \in \mathrm{JH}(\bigotimes_{v|p}(\eta_v \otimes \overline{\mathbb{F}}_p))$ if and*

only if $\overline{\rho}$ has a modular p-adic lift $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ such that for each $v|p$, the restriction $\rho_{|G_v}$ is potentially Barsotti-Tate (i.e. potentially crystalline with Hodge-Tate weights $(0,1)$) and $\mathrm{WD}(\rho_{|G_v})_{|I_v} = \tau_v$.

Here, and subsequently, we write $JH(V)$ for the set of Jordan-Hölder constituents of a representation V, whereas $\mathrm{WD}(\rho_{|G_v})$ denotes the Weil-Deligne representation corresponding to the local Galois representation $\rho_{|G_v}$. The reader is referred to the classic article [56] for the correspondence between Galois and Weil-Deligne representations.

Suppose that we know how to prove the first of the two claims at the beginning of this section, namely that if $\overline{\rho}$ is modular of weight σ, then $\sigma \in W(\overline{\rho})$. Assuming that, here is a strategy for proving the second claim. Let $\sigma \in W(\overline{\rho})$ be a Serre weight. If it is regular, then for each $v|p$ there exists a good η_v as above such that $\mathrm{JH}(\otimes_{v|p}(\eta_v \otimes \overline{\mathbb{F}}_p)) \cap W(\overline{\rho}) = \{\sigma\}$. Then by Proposition 3.2 it suffices to find a modular lift ρ with the properties specified in the statement of that proposition.

The most daunting aspect of coming up with a lift ρ of $\overline{\rho}$ that satisfies the conditions of Proposition 3.2 is clearly that of showing that the ρ we have constructed is modular. Fortunately, Gee's adaption of a modularity lifting theorem of Kisin comes to the rescue. This is the first of many close connections that we will see in these lectures between Serre's modularity conjecture and modularity lifting theorems.

Proposition 3.3. *Suppose that the hypotheses of Theorem 3.1 hold and that $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ is a lift of $\overline{\rho}$ such that $\rho_{|G_v}$ is potentially Barsotti-Tate and $\mathrm{WD}(\rho_{|G_v})_{|I_v} = \tau_v$ for each $v|p$. Suppose that there exists a cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_F)$ such that for every $v|p$, the local Galois representation $\rho_{\pi,v}$ is potentially ordinary if and only if $\rho_{|G_v}$ is potentially ordinary. Then ρ is modular.*

Note that the hypothesis on $\rho_{|G_{F(\zeta_p)}}$ in the statement of Theorem 3.1 is common in modularity lifting theorems à la Kisin, and this is the point in the proof where it is necessary.

Now we need to construct a lift ρ satisfying the conditions of Proposition 3.3. The theory of Breuil modules allows us to translate local conditions on Galois representations into linear-algebraic data.

Let k be a finite field of characteristic $p > 2$, let $W(k)$ be the associated ring of Witt vectors, and let $K_0 = W(k)[1/p]$ be its fraction field. Let K/K_0 be a totally tamely ramified Galois extension of degree e. Let $B \subset K_0$ be a subfield such that there exists a uniformizer $\pi \in \mathcal{O}_K$ satisfying $\pi^e \in B$. Choose such a π. Let $2 \leq k \leq p-1$ be an integer; this conflict of notation is standard and will produce no confusion. Let $E/\mathbb{F}_p$ be a finite extension. The category $\mathrm{BrMod}^{k-1}_{dd,B}$ of Breuil modules with descent data has as objects quintuples $(M, M_{k-1}, \varphi_{k-1}, N, \hat{g})$ such that:

(1) M is a finitely generated $(k \otimes_{\mathbb{F}_p} E)[u]/u^{ep}$-module that is free over $k[u]/u^{ep}$.
(2) M_{k-1} is a submodule such that $u^{e(k-1)}M \subset M_{k-1}$.
(3) $\varphi_{k-1} : M_{k-1} \to M$ is an E-linear and Frobenius-semilinear homomorphism whose image generates M as a $(k \otimes_{\mathbb{F}_p} E)[u]/u^{ep}$-module. Frobenius-semilinear in this case means that if $a \in k[u]/u^{ep}$ and $m \in M_{k-1}$, then $\varphi_{k-1}(am) = a^p \varphi_{k-1}(m)$.
(4) $N : M \to uM$ is a $(k \otimes_{\mathbb{F}_p} E)$-linear map satisfying

(a) $N(um) = uN(m) - um$ for all $m \in M$.
(b) $u^e N(M_{k-1}) \subset M_{k-1}$.
(c) $\varphi_{k-1}(u^e N(m)) = -\frac{\pi^e}{p} N(\varphi_{k-1}(m))$ for all $m \in M_{k-1}$.

(5) For each $g \in \mathrm{Gal}(K/B)$, there is an additive bijection $\hat{g} : M \to M$ such that
(a) Each $\hat{g}$ commutes with, φ_{k-1}, M, and the E-action.
(b) $\hat{1}$ is the identity map, where $1 \in \mathrm{Gal}(K/B)$ is the identity automorphism.
(c) $\hat{g} \circ \hat{h} = \widehat{g \circ h}$ for all $g, h \in \mathrm{Gal}(K/B)$.
(d) $\hat{g}(au^i m) = g(a)((g(\pi)/\pi)^i \otimes 1)u^i \hat{g}(m)$ for all $a \in k \otimes_{\mathbb{F}_p} E$, $m \in M$, and $i \geq 0$. To make sense of $g(a)$, note that k is the residue field of K_0, hence of K, and so is acted on by $\mathrm{Gal}(K/B)$. We let $\mathrm{Gal}(K/B)$ act trivially on the second component of $k \otimes_{\mathbb{F}_p} E$.

The connection between Breuil modules and potentially Barsotti-Tate Galois representations is evidenced, for instance, by the fact that the category $\mathrm{BrMod}^1_{dd,B}$ is equivalent to the category of finite flat group schemes over $\mathcal{O}_K$ with an action of E and descent data to B. Gee proves that the existence of a lift ρ which is potentially Barsotti-Tate at v of inertial type τ_v is equivalent to the existence of a Breuil module satisfying certain conditions. As we see from the definition above, Breuil modules with descent data are complicated objects but are very explicit, and one constructs the needed Breuil module by hand.

The proof for arbitrary (i.e. not necessarily regular) Serre weights follows the same lines, but the theory of Breuil modules, which itself is an extension of Fontaine-Laffaille theory, is not powerful enough. Here one uses Liu's theory of Kisin modules.

3.2. **Modular weights are predicted ones: some algebraic geometry.** In this section we will sketch how to prove that if $\overline{\rho}$ is modular of a Serre weight σ, then $\sigma \in W(\overline{\rho})$. In order to illustrate the variety of methods applicable to this problem, we will give an algebraic-geometry argument following [46] and [45]. This method was used to obtain the earliest results in this direction, but it has turned out to be less effective than the deformation-theoretic and p-adic Hodge-theoretic techniques of which a flavor was given in the previous section. The reader may, of course, find further details in [46].

In this section we will not impose such severe limitations on the ramification of p in the totally real field F, but we will suppose that $\overline{\rho}_{|G_v}$ is irreducible for all $v|p$. For each $v|p$, let e_v be the ramification index of $F_v/\mathbb{Q}_p$. Let $\sigma = \bigotimes_{v|p} \sigma_v$ be a Serre weight such that for each v and each $\tau \in S_v$ we have $0 \leq r_\tau \leq p - e_v - 1$ (so in particular we are assuming here that $e_v \leq p-1$).

Now we will recall some notions from Sections 2.2 and 2.3. Assume that $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is modular of weight σ. By definition, this implies the existence of a quaternion algebra B/F, giving rise to an algebraic group G, and an open compact subgroup $V = \left(\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})\right) \times U^p \subset G(\mathbb{A}^\infty)$ such that $H^1_{\acute{e}t}(M_V \otimes \overline{\mathbb{Q}}, \mathcal{L}_\sigma)_{\mathfrak{m}_{\overline{\rho}}} \neq 0$. For each $v|p$, let $U_1^{bal}(v) \subset \mathrm{GL}_2(\mathcal{O}_{F_v})$ be the subgroup of matrices whose reductions modulo v are unipotent upper triangular, i.e. of

matrices that are congruent to $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$ modulo v. Consider the open compact subgroup

$$U_1^{bal}(p) = \left(\prod_{v|p} U_1^{bal}(v) \right) \times U^p \subset G(\mathbb{A}^\infty).$$

Fix a place $v|p$. Let $D = W(k_v)$ be a ring of Witt vectors, let $K = F_v^{nr}$ be the fraction field of D, let $K' = K(\sqrt[q_v-1]{\pi_v})$ be a totally tamely ramified extension with $\mathrm{Gal}(K'/K) \simeq k_v^*$, and let $D' = \mathcal{O}_{K'}$. Then $M_{U_1^{bal}(p)}$ has an integral model over D, which we shall denote $\mathbf{M}_{U_1^{bal}(p)}$. Moreover, $\mathbf{M}_{U_1^{bal}(p)} \times_D D'$ has a well-behaved special fiber consisting of two smooth curves intersecting transversally at finitely many points.

Let $j : \mathrm{Gal}(K'/K) \to \mathcal{O}_{F_v}^*/(1+v)$ be the isomorphism induced by the Artin reciprocity map of local class field theory (normalized so as to send arithmetic Frobenius to uniformizers). We have natural actions of $\mathrm{GL}_2(\mathcal{O}_{F_v})$ (coming from the p-component of $G(\mathbb{A}^\infty)$) and of $\mathrm{Gal}(K'/K)$ on the special fiber of $\mathbf{M}_{U_1^{bal}(p)} \times_D D'$, and Carayol ([16], 10.3) shows that the action of $\gamma \in \mathrm{Gal}(K'/K)$ is equal to that of $\begin{pmatrix} j(\gamma)^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & j(\gamma)^{-1} \end{pmatrix}$, respectively, on the two components of the special fiber.

Recall that we are assuming that $\overline{\rho}$ is modular of a given Serre weight σ. Let $B(k_v) \subset \mathrm{GL}_2(k_v)$ be the Borel subgroup of upper triangular matrices, and let $\theta : B(k_v) \to \overline{\mathbb{F}}_p^*$ be a character such that $\sigma_v \in \mathrm{JH}(\mathrm{Ind}_{B(k_v)}^{\mathrm{GL}_2(k_v)}\theta)$. Let C be the Néron model over D' of the curve $\mathrm{Pic}^0(\mathbf{M}_{U_1^{bal}(p)}) \times K'$. Then $C[p^\infty]$ is a p-divisible group, and the reduction $C[p^\infty] \otimes \mathbf{T}/\mathfrak{m}_{\overline{\rho}}$ contains a finite piece G_θ on which the diagonal matrices in $\mathrm{GL}_2(\mathcal{O}_{F_v})$ act via the character θ. By the main result of [9], $G_\theta[\mathfrak{m}_{\overline{\rho}}]_K$ is a direct sum of a finite number of copies of $\overline{\rho}$. As in Section 2.4 above, $\overline{\rho}_{|I_v}$ is a direct sum of two characters, φ and φ', that satisfy $\varphi^{q_v} = \varphi'$ and $(\varphi')^{q_v} = \varphi$. We can pick out a subspace $H \subset G_\theta[\mathfrak{m}_{\overline{\rho}}]_K$ of rank q_v^2 on which $\mathrm{Gal}(\overline{K}/K) \simeq I_v$ acts by the character φ.

Let $\mathbb{F}$ be a finite field, sufficiently large so that $\mathrm{im}(\overline{\rho}_{|G_v}) \subset \mathrm{GL}_2(\mathbb{F})$ and $\mathbb{F}_{q_v^2} \subset \mathbb{F}$. We will apply Raynaud's theory of vector space schemes [44]. An $\mathbb{F}$-vector space scheme over D is a commutative group scheme W/D carrying an action of $\mathbb{F}$. Let $\mathcal{I} \subset \mathcal{O}_W$ be the augmentation ideal, so that $\mathcal{O}_W = \mathcal{I} \oplus \mathcal{O}_D$. Here $\mathcal{O}_W$ is the structure sheaf of W. It is easy to see that $\mathcal{I}$ decomposes as follows:

$$\mathcal{I} = \bigoplus_{\chi : \mathbb{F}^* \to D^*} \mathcal{I}_\chi,$$

where $\mathcal{I}_\chi$ is the piece of $\mathcal{I}$ on which $\mathbb{F}$ acts via the character χ. We see that H is an $\mathbb{F}_{q_v^2}$-vector space scheme, and it satisfies the additional crucial property that each $\mathcal{I}_\chi$ is a non-zero invertible sheaf. The vector space scheme H is endowed with two Galois actions:

(1) As we noted before, $\mathrm{Gal}(\overline{K}/K) \simeq I_v$ acts on $H(\overline{K})$ by the character φ, which we are trying to determine.
(2) $\mathrm{Gal}(K'/K) \simeq k_v^*$ acts on the cotangent space $\mathrm{cot}(H_{D'} \times_{D'} \overline{\mathbb{F}}_p)$. Thanks to Carayol's congruences mentioned above, we can express this action explicitly in terms of the character θ.

From Raynaud's work one deduces an explicit relation between these two different Galois actions. We will not perform the calculations here, but the reader can find them in Section 3 of [45]. At the end we obtain a collection $\Phi(\theta)$ of characters φ that are compatible with the known action of $\mathrm{Gal}(K'/K)$. It turns out that these are precisely the characters φ arising from mod p Galois representations $\overline{\rho}$ that are modular of some Serre weight $\sigma' \otimes \sigma^v$, where $\sigma' \in \mathrm{JH}(\mathrm{Ind}_{B(k_v)}^{\mathrm{GL}_2(k_v)}\theta)$ and $\sigma^v = \bigotimes_{w|p, w\neq v} \sigma_w$, where σ_w is an arbitrary local Serre weight at w. Observe that this is the best result that we can hope to obtain at this stage of the proof, since so far we have only used θ in our calculations and not σ_v itself.

To get a more precise result, we consider all the characters $\theta : B(k_v) \to \overline{\mathbb{F}}_p^*$ such that $\sigma_v \in \mathrm{JH}(\mathrm{Ind}_{B(k_v)}^{\mathrm{GL}_2(k_v)}\theta)$. Clearly all the φ associated to $\overline{\rho}$ that are modular of weight $\sigma_v \otimes \sigma^v$ lie in the intersection $\bigcap_{\sigma_v \in \mathrm{JH}(\mathrm{Ind}\theta)} \Phi_\theta$. We hope that this intersection will turn out to be exactly the collection of representations $\overline{\rho}$ such that $\sigma_v \in W_v(\overline{\rho})$. The hope comes true when σ_v is of the form (4) with $0 \leq r_\tau \leq p-1-e_v$ for all $\tau \in S_v$, which is the reason for the hypothesis to this effect that we made above. The combinatorial issues that prevent this method from giving us as good a theorem as we would like when r_τ does not satisfy the constraint $0 \leq r_\tau \leq p-1-e_v$ are essentially also what prevents the method of [23] from handling the non-regular Serre weights.

4. The mod p local Langlands correspondence

The Langlands philosophy postulates a deep connection between algebra and analysis and is one of the main motivations behind modern research on Serre's modularity conjecture and its generalizations. In this section we will show a very brief glimmer of the connection between them. Let $n \geq 1$, let $F/\mathbb{Q}_p$ be a p-adic field, and let E be a field. In very rough terms, we would like to have a correspondence between certain Galois representations $\rho : G_F \to \mathrm{GL}_n(E)$ and certain representations of $\mathrm{GL}_n(F)$ on vector spaces over E; one of the most difficult parts of this problem is finding the correct definition of "certain." Often one can attach L-functions to each of these types of objects, and the L-functions of the objects paired by the correspondence should match.

In the case of $E = \mathbb{C}$, the correspondence was proved by Harris and Taylor [29] and Henniart [30], working with Weil-Deligne representations instead of the closely related Galois representations. If $E = \overline{\mathbb{F}}_l$, with $l \neq p$, then considerable progress was made by Vignéras [60]. However, if $E = \overline{\mathbb{F}}_p$, then very little is known. In many respects the study of the mod p local Langlands correspondence is at the stage in its development where the complex local Langlands correspondence was in the 1970's: one tries to classify objects on both sides and pair them up explicitly in a natural way, but no deep underlying theory is yet available. Moreover, understanding the mod p representation theory of $\mathrm{GL}_n(F)$ has turned out to be remarkably difficult.

In this section, we will use the following notation. We let $G = \mathrm{GL}_n(F)$ and consider the maximal open compact subgroup $K = \mathrm{GL}_n(\mathcal{O}_F)$ and the center $Z = Z(G) \simeq F^*$. Let $\pi \in \mathcal{O}_F$ be a uniformizer, and let $k_F = \mathcal{O}_F/(\pi)$ be the residue field as usual. Let $q = p^f$ be the cardinality of k. Let $I \subset K$ be the Iwahori subgroup consisting of matrices that are upper triangular modulo π, and let $I(1)$ be the pro-p-Sylow subgroup of I. For instance, if

$n = 2$ then

$$\begin{aligned} I &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_F) : c \in \pi\mathcal{O}_F \right\} \\ I(1) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_F) : c \in \pi\mathcal{O}_F; a, d \in 1 + \pi\mathcal{O}_F \right\}. \end{aligned}$$

Let σ be an irreducible $\overline{\mathbb{F}}_p$-representation of K. By Proposition 2.3, σ factors through the natural reduction map $K \to \mathrm{GL}_n(k)$, since the kernel of this map is a pro-p group. Therefore, σ arises from an irreducible $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_n(k)$ by inflation; these are exactly the objects that we called local Serre weights above in the case $n = 2$. Moreover, we can view σ as a representation of the larger group KZ by decreeing that $\begin{pmatrix} \pi & 0 \\ 0 & \pi \end{pmatrix}$ acts trivially.

If $H \subset G$ is any open subgroup, and τ is an $\overline{\mathbb{F}}_p$-representation of H, we can consider the compact induction $\mathrm{ind}_H^G \tau$. A model for this representation is given by the space of functions $f : G \to V_\tau$ that are locally constant, compactly supported modulo Z, and satisfy the condition $f(hg) = \tau(h) \cdot f(g)$ for every $h \in H$ and $g \in G$. Here V_τ is the underlying $\overline{\mathbb{F}}_p$-vector space of τ. The action of G is given by $(gf)(x) = f(xg)$ for all $g, x \in G$. Note that if H is a subgroup of finite index, then local constancy and compact support are automatic and this is just the usual induction. The endomorphisms of this compact induction were computed by Barthel and Livné [4] for $n = 2$. For $n \geq 2$, see [48] for an explicit computation and [33] for a more conceptual argument on the level of algebraic groups.

Proposition 4.1. *Let σ be an irreducible $\overline{\mathbb{F}}_p$-representation of K. The endomorphism algebra $\mathrm{End}_G(\mathrm{ind}_{KZ}^G \sigma)$ is equal to a polynomial ring $\overline{\mathbb{F}}_p[T_1, \ldots, T_{n-1}]$, where the T_i are explicitly defined endomorphisms.*

Let W be an irreducible $\overline{\mathbb{F}}_p$-representation of G with central character, i.e. such that the elements of Z act by scalars. Twisting by an unramified character, we may assume that $\begin{pmatrix} \pi & 0 \\ 0 & \pi \end{pmatrix}$ acts trivially. If $\sigma \subset W_{|K}$ is a K-submodule of G, then by Frobenius reciprocity we obtain a non-zero homomorphism $\mathrm{ind}_{KZ}^G \sigma \to W$ of G-modules, which must be a surjection by the irreducibility of W. We say that W is *admissible* if the space of invariants $W^U = \{w \in W : \forall u \in U, uw = w\}$ is finite-dimensional for any open subgroup $U \subset G$; since W is an $\overline{\mathbb{F}}_p$-representation this is in fact equivalent to $W^{I(1)}$ being finite-dimensional.

The endomorphism algebra $\mathrm{End}_G(\mathrm{ind}_{KZ}^G \sigma)$ is commutative by Proposition 4.1, and it acts on $\mathrm{Hom}_G(\mathrm{ind}_{KZ}^G \sigma, W)$ in the obvious way. If W is assumed to be admissible, then $\mathrm{Hom}_G(\mathrm{ind}_{KZ}^G \sigma, W) \simeq \mathrm{Hom}_{KZ}(\sigma, W_{|KZ})$ is finite-dimensional (because σ must contain a non-zero $I(1)$-invariant, which must map to an element of $W^{I(1)}$) and necessarily contains an eigenvector for the $\mathrm{End}_G(\mathrm{ind}_{KZ}^G \sigma)$-action. We obtain the following result.

Proposition 4.2. *Let W be a smooth irreducible $\overline{\mathbb{F}}_p[G]$-module with central character. Assume that W is admissible if $n \geq 3$. Let σ be an irreducible $\overline{\mathbb{F}}_p[K]$-module σ such that $\sigma \subset W_{|K}$. Then there exist an unramified character $\chi : F^* \to \overline{\mathbb{F}}_p^*$ and scalars $\lambda_1, \ldots, \lambda_{n-1} \in \overline{\mathbb{F}}_p$*

such that there exists a surjection of G-modules

$$(\chi \circ \det) \otimes \mathrm{ind}_{KZ}^{G}\sigma/(T_1 - \lambda_1, \ldots, T_{n-1} - \lambda_{n-1})\mathrm{ind}_{KZ}^{G}\sigma \twoheadrightarrow W. \tag{7}$$

Proof. If W is admissible, then we have sketched out the proof. If $n = 2$, then Barthel and Livné (see Theorems 32 and 33 of [4]) obtain this result without assuming admissibility of W by using the fact that $\mathrm{End}_G(\mathrm{ind}_{KZ}^{G}\sigma)$ has Krull dimension 1. □

For the rest of this section, suppose that $n = 2$. In this case, the endomorphism algebra $\mathrm{End}_G(\mathrm{ind}_{KZ}^{G}\sigma)$ has a single generator T_1, which we will call T. Up to unramified twist, we know that every irreducible $\overline{\mathbb{F}}_p[G]$-module with central character is a quotient of $\mathrm{ind}_{KZ}^{G}\sigma/(T - \lambda)(\mathrm{ind}_{KZ}^{G}\sigma)$ for some σ and some $\lambda \in \overline{\mathbb{F}}_p$. We say that an irreducible W as above is *supersingular* if it is a quotient of some $\mathrm{ind}_{KZ}^{G}\sigma/(T - \lambda)(\mathrm{ind}_{KZ}^{G}\sigma)$. Barthel and Livné proved a partial classification of the irreducible $\overline{\mathbb{F}}_p[G]$-modules with central character as follows. Note that if $G = \mathrm{GL}_2(\mathbb{Q}_p)$, then Berger [5] recently showed that all irreducible $\overline{\mathbb{F}}_p[G]$-modules have central character, but this is not known even for $G = \mathrm{GL}_2(F)$ whenever $F \neq \mathbb{Q}_p$.

Theorem 4.3 (Barthel-Livné)**.** *Let σ be an irreducible $\overline{\mathbb{F}}_p[K]$-module.*

1. *If σ has dimension other than 1 or p^f (the minimal and maximal dimensions possible) or if $\lambda \neq \pm 1$, then $\mathrm{ind}_{KZ}^{G}\sigma/(T - \lambda)(\mathrm{ind}_{KZ}^{G}\sigma)$ is irreducible and is isomorphic to the parabolic induction of a character from the upper triangular Borel subgroup $B \subset G$.*
2. *The induction $\mathrm{ind}_{B}^{G}\mathbf{1}$, where $\mathbf{1}$ is the trivial character of B, has length two. Its subquotients are a one-dimensional representation* det *and an infinite-dimensional analogue of the Steinberg representation, denoted* St*.*
3. *Up to unramified twist, every smooth irreducible $\overline{\mathbb{F}}_p[G]$-module W with central character satisfies exactly one of the following statements:*
   (a) *$W \simeq \mathrm{ind}_{KZ}^{G}\sigma/(T - \lambda)(\mathrm{ind}_{KZ}^{G}\sigma)$, where σ has dimension other than 1 or p^f, or $\lambda \neq \pm 1$.*
   (b) *$W \simeq \chi \circ \det$ for some smooth character $\chi : F^* \to \overline{\mathbb{F}}_p^*$.*
   (c) *$W \simeq (\chi \circ \det) \otimes \mathrm{St}$ for some smooth character $\chi : F^* \to \overline{\mathbb{F}}_p^*$.*
   (d) *W is supersingular.*

Remark 4.4. The previous theorem classifies all non-supersingular (smooth, with central character) $\overline{\mathbb{F}}_p$-representations of $\mathrm{GL}_2(F)$ for arbitrary finite extensions $F/\mathbb{Q}_p$. Herzig [32] proved a generalization of this theorem of $\mathrm{GL}_n(F)$ for $n > 2$, in which all smooth admissible representations of $\mathrm{GL}_n(F)$ with central character are classified in terms of the supersingular representations of $\mathrm{GL}_m(F)$ for $m \leq n$. A representation of $\mathrm{GL}_n(F)$ is called supersingular if it is a quotient of $\mathrm{ind}_{KZ}^{G}\sigma/(T_1, \ldots, T_{n-1})$. Abe [1] further generalized this result to a wider class of reductive groups.

Let L be a number field and v a place of L such that $L_v \simeq F$. If $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is an irreducible local Galois representation, let $\tilde{\rho} : G_L \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a global representation such that $\tilde{\rho}_{|G_v} \simeq \rho$. Recall that in Conjecture 2.9 we defined a set $W_v(\tilde{\rho})$ of local Serre weights at v, which in fact depends only on ρ. Thus we can speak of a set $W(\rho)$ of modular local Serre weights.

Now suppose that $G = \mathrm{GL}_2(\mathbb{Q}_p)$. In this case, the irreducible $\overline{\mathbb{F}}_p[K]$-modules have the form $\sigma = \det^w \otimes \mathrm{Sym}^r \overline{\mathbb{F}}_p^2$ with $0 \leq w \leq p-2$ and $0 \leq r \leq p-1$. We define an involution on the set of these local Serre weights as follows. For σ as above, define $\sigma' = \det^{w+r} \otimes \mathrm{Sym}^{p-1-r} \overline{\mathbb{F}}_p^2$. Note that $(\sigma')' = \sigma$. It is easy to compute from the statement of Conjecture 2.9 that if $\rho : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is irreducible, then $W(\rho)$ is necessarily of the form $W(\rho) = \{\sigma, \sigma'\}$.

Breuil [10] completed the classification of the irreducible $\overline{\mathbb{F}}_p$-representations of $\mathrm{GL}_2(\mathbb{Q}_p)$ with the following result.

Theorem 4.5 (Breuil)**.** *Let $G = \mathrm{GL}_2(\mathbb{Q}_p)$. Then for every local Serre weight σ, the G-module $\mathrm{ind}_{KZ}^G \sigma / T(\mathrm{ind}_{KZ}^G \sigma)$ is irreducible. Moreover, for every σ we have*

$$\mathrm{ind}_{KZ}^G \sigma / T(\mathrm{ind}_{KZ}^G \sigma) \simeq \mathrm{ind}_{KZ}^G \sigma' / T(\mathrm{ind}_{KZ}^G \sigma') \tag{8}$$

and these are the only isomorphisms among supersingular $\overline{\mathbb{F}}_p[G]$-modules.

Remark 4.6. Note that the two operators T appearing in (8) are different objects. The T on the left-hand side is the generator of the endomorphism algebra of $\mathrm{ind}_{KZ}^G \sigma$, while the one on the right-hand side generates the endomorphism algebra of $\mathrm{ind}_{KZ}^G \sigma'$.

Proof. Let $W = \mathrm{ind}_{KZ}^G \sigma / T(\mathrm{ind}_{KZ}^G \sigma)$ and let $U \subset W$ be an irreducible G-submodule. By explicit computation, one shows that $W^{I(1)}$ is two-dimensional and that every non-zero element of $W^{I(1)}$ generates W as a G-module. But $U^{I(1)} \neq 0$ by Proposition 2.3 and hence $U = W$. The isomorphisms of (8) are constructed explicitly, and one shows that

$$\mathrm{soc}_K(\mathrm{ind}_{KZ}^G \sigma / T(\mathrm{ind}_{KZ}^G \sigma)) \simeq \sigma \oplus \sigma', \tag{9}$$

implying that there are no other isomorphisms. Recall that for a G-module M, the socle $\mathrm{soc}_K(M)$ is the direct sum of all irreducible K-submodules of M. □

If $\rho : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is an irreducible local Galois representation, define an $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ by $\pi(\rho) = \mathrm{ind}_{KZ}^G \sigma / T(\mathrm{ind}_{KZ}^G \sigma)$, where $\sigma \in W(\rho)$. It is immediate from the results just presented that $\pi(\rho)$ is well-defined and that this construction provides a bijection between irreducible Galois representations $\rho : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ and supersingular representations of $\mathrm{GL}_2(\mathbb{Q}_p)$. Note that the following relation is satisfied:

$$\mathrm{soc}_K(\pi(\rho)) = \bigoplus_{\sigma \in W(\rho)} \sigma. \tag{10}$$

In the same paper [10], Breuil constructed $\pi(\rho)$ for semisimple reducible ρ, and eventually Colmez defined $\pi(\rho)$ for indecomposable ρ, thereby completing the mod p local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$. These constructions are more complicated than the one presented above, and we will not give them here, nor shall we argue why these definitions of $\pi(\rho)$ are the "correct" ones. However, it is important to note that the property (10) remains true for all ρ.

If $F \neq \mathbb{Q}_p$, then almost nothing is known about the mod p local Langlands correspondence for $\mathrm{GL}_2(F)$, and the statements that are known are almost all negative. For instance, we know that there cannot be a bijection between irreducible Galois representations $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ and supersingular representations of $\mathrm{GL}_2(F)$ because there are far too many of the latter. The

condition (10) does not isolate a $\pi(\rho)$ because, for unramified extensions $F/\mathbb{Q}_p$, Breuil and Paskunas [12] have proved the existence of infinite families of supersingular representations W satisfying $\mathrm{soc}_K(W) \simeq \bigoplus_{\sigma \in W(\rho)} \sigma$. Moreover, although it is immediate from Zorn's Lemma that supersingular representations of $\mathrm{GL}_2(F)$ exist, we do not have a single explicit construction of one; the proof of Breuil and Paskunas uses the theory of diagrams and involves the taking of injective envelopes, which makes their work very non-explicit. In fact, Schraen [49] has shown that if $F/\mathbb{Q}_p$ is quadratic, then no supersingular representation of $\mathrm{GL}_2(F)$ is finitely presented. This makes it difficult to contemplate generalizations of Colmez's construction.

Let e be the ramification index of $F/\mathbb{Q}_p$, and recall that the residue field k of F satisfies $[k : \mathbb{F}_p] = f$. Let F_0 be the maximal unramified subextension of $F/\mathbb{Q}_p$, and observe that, since F and F_0 have the same residue field, the Serre weights for F_0 are the same as those for F. If $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is irreducible, then (see [47]) one defines e^f irreducible representations $\rho_1, \ldots, \rho_{e^f} : G_{F_0} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that one expects

$$\mathrm{soc}_K(\pi(\rho)) \simeq \bigoplus_{i=1}^{e^f} \bigoplus_{\sigma \in W(\rho_i)} \sigma.$$

The formula above explains the multiplicities of the different constituents of the K-socle of $\pi(\rho)$. Ongoing work of Breuil and Diamond aims to specify the K-socles of $\pi(\rho)$ for reducible ρ. This section has only scratched the surface of the mod p local Langlands correspondence and has said almost nothing about current research, but we hope that it has sufficiently piqued the reader's interest to consult the literature for more details about the field.

5. Potential modularity and compatible systems

After the digression about mod p local Langlands in the previous section, we return to our discussion of Serre's modularity conjecture. In particular, we return to the notation of Section 2, so that F is now again a totally real number field.

5.1. **A wish list.** Suppose that we have two mod p Galois representations $\overline{\rho}_1 : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_{p_1})$ and $\overline{\rho}_2 : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_{p_2})$, where p_1 and p_2 are two primes, possibly distinct. It clearly would be useful to be able to prove statements of the form "if $\overline{\rho}_1$ is modular and certain conditions are satisfied, then $\overline{\rho}_2$ is modular as well." Such theorems would allow us to leverage knowledge of Serre's conjecture in some special cases to prove it for larger classes of Galois representations.

How can we relate the modularity of two different Galois representations? A crucial idea is to think about p-adic Galois representations, and to recall that when we first encountered them, in Prof. Böckle's lectures, they were constructed in families. Indeed, for a modular form f, we obtained a representation $\rho_{f,l} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$ for each prime l. The $\rho_{f,l}$ for different l were very intimately related.

An important starting point for work on modularity is an axiomatization of this phenomenon: the notion of weakly and strongly compatible systems that we saw in Prof. Böckle's lectures. A strongly compatible system $\{\rho_l : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_l)\}$ of Galois representations behaves like a family of representations arising from a modular form. In particular, if one

member of the system is modular, then they all are, and the same is true of their reductions. This gives us a general strategy for proving the "if $\overline{\rho}_1$ is modular, then $\overline{\rho}_2$ is too" theorems that we wished for at the beginning of this section. Suppose we could find a compatible system $\{\rho_l\}$ such that $\overline{\rho}_1 \simeq \overline{\rho_{p_1}}$ and $\overline{\rho}_2 \simeq \overline{\rho_{p_2}}$. We are assuming that $\overline{\rho}_1$ is modular. If we could somehow prove that ρ_{p_1} is modular, the compatible system would allow us to conclude that ρ_{p_2} is modular as well, and hence that $\overline{\rho}_2$ is modular.

Three major ingredients are involved in implementing this strategy. Starting with a representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$, we have the following **wish list**:

(1) Find nice lifts $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ of $\overline{\rho}$. "Nice" will mean that ρ satisfies hypotheses that make the other parts of the wish list available.
(2) Given such a lift ρ, embed it in a compatible system $\{\rho_l\}$ such that $\rho \simeq \rho_p$.
(3) Modularity lifting theorems.

5.2. **Potential modularity of mod p Galois representations.** In this section we will sketch a proof of the following theorem of Taylor [58].

Proposition 5.1. *Let $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(k)$ be a continuous, irreducible, odd Galois representation, where k is a finite field of characteristic p. Then there exists a Galois totally real extension $F/\mathbb{Q}$ that is unramified at p and such that $\overline{\rho}_{|G_F}$ is modular.*

If $\overline{\rho}$ has solvable image, then this problem may be handled by the methods of Langlands and Tunnell, so we will assume that this is not the case. Passing to a suitable totally real extension $F/\mathbb{Q}$ that is unramified at p, we may assume that the determinant of $\overline{\rho}_{|G_v}$ is the mod p cyclotomic character for all places $v|p$ of F, and that $\overline{\rho}$ has the following form at all $v|p$:

$$\overline{\rho}_{|G_v} \sim \begin{pmatrix} \varepsilon\chi_v^{-1} & * \\ 0 & \chi_v \end{pmatrix},$$

where ε is the mod p cyclotomic character and $\chi_v : G_v \to k^*$ is a character. Now $\overline{\rho}_{G_F}$ looks like it could be the restriction of an ordinary p-adic representation of Hodge-Tate weights $\{0, 1\}$ coming from an abelian variety, and our task is to show that this is indeed the case.

Recall that if A/F is an abelian variety, then for every finite place v of F, the Galois group G_F acts on the torsion $A[v]$, and the reductions of these representations give us a strictly compatible system $\{\overline{\rho}_{A,v}\}$. We are looking for an abelian variety A/F such that $\overline{\rho}_{|G_F} \simeq \overline{\rho}_{A,v}$ for some v. Let $v' \neq v$ be another place of F lying over p. We will cleverly set up a moduli problem of abelian varieties in such a way that a handy theorem of Morel-Bailly [42], quoted below, will give us exactly the existence of the A that we need.

Let $M/\mathbb{Q}$ be an imaginary quadratic extension and $\psi : G_M \to \overline{\mathbb{Q}}_p^*$ a character. Consider the moduli problem of triples $(A, \varphi_v, \varphi_{v'})$ such that A is a Hilbert-Blumenthal abelian variety (this is an abelian variety carrying an action of the ring of integers $\mathcal{O}_E$ of a specified totally real field E and some additional structure such as a Rosati involution and a polarization; we will not give a precise definition here but only mention that there is a well-developed theory of moduli problems for these objects, founded by Rapoport in [43]), and the φ_v and $\varphi_{v'}$ are isomorphisms $\varphi_v : \overline{\rho}_{|G_F} \xrightarrow{\sim} \overline{\rho}_{A,v}$ and $\varphi_{v'} : \mathrm{Ind}_{G_M}^{G_{\mathbb{Q}}} \overline{\psi} \xrightarrow{\sim} \overline{\rho}_{A,v'}$. The general theory of Hilbert-Blumenthal abelian varieties tells us that this moduli problem is representable by some moduli

space $X/\mathbb{Q}$, and if we knew that this space had a rational point, it would correspond to the abelian variety A that we are looking for.

Proposition 5.2 (Moret-Bailly)**.** *Let K be a number field and S a finite set of places of K. If X/K is a geometrically irreducible smooth quasi-projective scheme and $X(K_v) \neq \varnothing$ for all $v \in S$, then $X(K_S)$ is Zariski dense in X. Here K_S/K is the maximal extension of K in which all $v \in S$ split completely.*

By choosing E, v, v', ψ wisely, it can be arranged that the hypotheses of Moret-Bailly's theorem are satisfied for $K = \mathbb{Q}$ and $X/\mathbb{Q}$ the moduli space considered above. In fact, Moret-Bailly's result appears to be far stronger than what we need to prove the existence of a rational point. This gives us the freedom to strengthen Proposition 5.1 by imposing a number of additional properties on the totally real field F, such as requiring it to be linearly disjoint from any specified number field. These strengthenings turn out to be essential, as they allow $\overline{\rho}_{|G_F}$ to satisfy the hypotheses of the modularity lifting theorems that we will call upon later.

5.3. **Deformation theory and modularity lifting results.** In the previous section we laid out the ingredients of the proof of a potential modularity theorem for mod p representations. Now we want to build on that result to get a potential modularity theorem for p-adic representations, which will be used in Section 5.4. Suppose that we are given a continuous, odd, irreducible mod p Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(k)$, as usual. First of all, we want to find a nice p-adic lift of $\overline{\rho}$ as in the first item of the wish list of Section 5.1.

Consider the following deformation problem. We want to study deformations $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(A)$, where A is a complete local noetherian algebra with residue field k, such that ρ lifts $\overline{\rho}$. In addition, for each prime l we fix an equivalence class τ_l of representations of the inertia group $I_l \simeq G_{\mathbb{Q}_l^{nr}}$, such that all but finitely many of the τ_l are trivial. Let $\chi_p : G_{\mathbb{Q}} \to \overline{\mathbb{Q}}_p^*$ be a character; we have $\chi_p = \omega_p^{k-1}$, where ω_p is the p-adic cyclotomic character. We require that $\det \rho = \chi_p$, that τ_l be the restriction to inertia of the Weil-Deligne representation associated to $\rho_{|G_l}$ for each l, and that $\rho_{|G_p}$ be crystalline of Hodge-Tate weights $\{0, k-1\}$.

By general deformation theory, this deformation problem is represented by a complete noetherian local ring $R^X_{\overline{\rho},\mathbb{Q}}$ with residue field k. It can be proved with very considerable effort, using Galois cohomology and the Euler characteristic formula (see [8]), that $\dim R^X_{\overline{\rho},\mathbb{Q}} \geq 1$ (by the dimension of a ring we mean the Krull dimension). To get our nice lift of $\overline{\rho}$, we need to show that $R^X_{\overline{\rho},\mathbb{Q}}$ has a point over an algebra of characteristic zero.

Consider the Hecke algebra $\mathbf{T} = \mathbb{Z}_p[T_l : l \neq p]$, which acts on the space of modular forms $S_k(\Gamma_1(N(\overline{\rho}))$. We get a natural surjection

$$R^X_{\overline{\rho},\mathbb{Q}} \twoheadrightarrow \mathbf{T}_{\mathfrak{m}_{\overline{\rho}}} \tag{11}$$

by the universal property of $R^X_{\overline{\rho},\mathbb{Q}}$. On the other hand, if every deformation classified by $R^X_{\overline{\rho},\mathbb{Q}}$ is modular, then the universal deformation must factor through $\mathbf{T}_{\mathfrak{m}_{\overline{\rho}}}$ and therefore $R^X_{\overline{\rho},\mathbb{Q}} \simeq \mathbf{T}_{\mathfrak{m}_{\overline{\rho}}}$. Proving a modularity lifting theorem, therefore, comes down to proving such an isomorphism, i.e. an "$R = T$" theorem. One of the breakthroughs of the Taylor-Wiles method [61] [59] was the understanding that modularity lifting results could often be reduced to statements about

ring-theoretic properties of Hecke algebras. For example, if $R^X_{\overline{\rho},\mathbb{Q}}$ were an integral domain, then proving $\dim R^X_{\overline{\rho},\mathbb{Q}} = \dim \mathbf{T}_{\mathfrak{m}_{\overline{\rho}}}$ would suffice to establish that (11) is an isomorphism, since the quotient of an integral domain by a non-trivial ideal has strictly lower Krull dimension than the original ring. In general, Spec $R^X_{\overline{\rho},\mathbb{Q}}$ will have more than one irreducible component, and proving that (11) is an isomorphism often amounts to showing that each component overlaps the image of Spec $\mathbf{T}_{\mathfrak{m}_{\overline{\rho}}}$ as well as comparing Krull dimensions.

In fact, we do not know that (11) is an isomorphism. However, we know from Proposition 5.1 that $\overline{\rho}_{|G_F}$ is modular for some totally real fields F. We may consider an analogous deformation problem to the one studied above, but over F; it is represented by a deformation ring $R^X_{\overline{\rho},F}$. Moreover, if we choose F correctly, then a modularity lifting theorem is known by work of Diamond [17] and Fujiwara [22]; in that case one can prove that $R^X_{\overline{\rho},F}$ is isomorphic to a suitable localized Hecke algebra $\mathbf{T}_F$.

The map Spec $R^X_{\overline{\rho},\mathbb{Q}} \to$ Spec $R^X_{\overline{\rho},F}$ corresponding to restriction to the subgroup G_F of representations of $G_{\mathbb{Q}}$ is clearly quasi-finite, i.e. has finite fibers. Indeed, G_F has finite index in $G_{\mathbb{Q}}$ and it is not hard to see that there are only finitely many ways to extend a representation of G_F to the larger group $G_{\mathbb{Q}}$. Moreover, the Hecke algebra $\mathbf{T}_F$ is finitely generated as a $\mathbb{Z}_p$-module, since it embeds in the endomorphism algebra of a suitable abelian variety. This implies that $\mathbf{T}_F/(p)$ is a finite set, therefore that $R^X_{\overline{\rho},F}/(p)$ is finite, and therefore that $R^X_{\overline{\rho},\mathbb{Q}}/(p)$ is finite, hence has dimension zero. Hence, any prime ideal of $R^X_{\overline{\rho},\mathbb{Q}}$ containing (p) is necessarily maximal.

On the other hand, recall that $\dim R^X_{\overline{\rho},\mathbb{Q}} \geq 1$. This means that there exists a non-maximal prime ideal $P \subset R^X_{\overline{\rho},\mathbb{Q}}$. By the above, we know that P does not contain (p). Since $R^X_{\overline{\rho},\mathbb{Q}}$ is finitely generated over $\mathbb{Z}_p$, it follows that the quotient $R^X_{\overline{\rho},\mathbb{Q}}/P$ embeds into the ring of integers $\mathcal{O}_L$ of a suitable finite extension $L/\mathbb{Q}_p$. Now by the universal property of $R^X_{\overline{\rho},\mathbb{Q}}$, the embedding $R^X_{\overline{\rho},\mathbb{Q}}/P \hookrightarrow \mathcal{O}_L$ corresponds to a p-adic Galois representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O}_L)$ lifting $\overline{\rho}$.

We have now achieved the first part of the wish list in Section 5.1. In fact, by doing all of this more carefully we could ensure that the obtained lift ρ has a variety of good properties.

5.4. **Constructing compatible systems.** In the previous section, we started with a mod p representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ and found a finite extension $L/\mathbb{Q}_p$ and a p-adic representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O}_L)$ lifting $\overline{\rho}$, thereby fulfilling the first part of the wish list of section 5.1. In this section we will build a compatible system around ρ.

By Taylor's potential modularity theorem (Proposition 5.1), we know that there is a Galois totally real field $F/\mathbb{Q}$ such that $\overline{\rho}_{|G_F}$ is modular. By the modularity lifting theorems of Diamond and Fujiwara that were mentioned in the previous section, we know that $\rho_{|G_F}$ is modular as well. Let $G = \mathrm{Gal}(F/\mathbb{Q})$. By Brauer's theorem (see, for instance, chapter 10 of [53]) there exist solvable subgroups $H_i \subset G$, integers $n_i \in \mathbb{Z}$, and one-dimensional representations χ_i of H_i such that

$$\mathbf{1} = \sum_{i=1}^{t} n_i \mathrm{Ind}_{H_i}^{G} \chi_i, \tag{12}$$

in the Grothendieck group of G, where $\mathbf{1}$ is the trivial representation of G. Note that even though $\mathbf{1}$ is a true representation, some of the n_i might be negative. This will cause us problems later. Set F_i to be the fixed field of H_i. Tensoring with ρ, we obtain that

$$\rho = \sum_{i=1}^{t} n_i \mathrm{Ind}_{G_{F_i}}^{G_{\mathbb{Q}}}(\rho_{|G_{F_i}} \otimes \chi_i). \tag{13}$$

Since $\mathrm{Gal}(F/F_i) = H_i$ is solvable, we conclude by Langlands-Tunnell solvable base change that each $\rho_{|G_{F_i}}$ arises from an automorphic form π_i on F_i. Hence we can trivially rewrite (13) as

$$\rho = \rho_p = \sum_{i=1}^{t} n_i \mathrm{Ind}_{G_{F_i}}^{G_{\mathbb{Q}}}(\rho_{\pi_i,p} \otimes \chi_i). \tag{14}$$

Since each $\rho_{\pi_i,p}$ comes from an automorphic form and therefore sits in a compatible system of representations of G_{F_i}, it is very tempting to define

$$\rho_l = \sum_{i=1}^{t} n_i \mathrm{Ind}_{G_{F_i}}^{G_{\mathbb{Q}}}(\rho_{\pi_i,l} \otimes \chi_i) \tag{15}$$

for arbitrary primes l. In fact, this idea works. If we knew that the ρ_l were true representations and not just virtual ones, then the compatible system properties of the $\{\rho_{\pi_i,l}\}$ would easily imply that $\{\rho_l\}$ is a compatible system as well. In fact, it can indeed be checked that the $\{\rho_l\}$ are true representations. This fulfills the second part of the wish list.

6. Proof of Serre's conjecture

We are finally in a position to give an exceedingly impressionistic sketch of the strategy behind the proof of Serre's conjecture. For more detail, the reader is referred to Wintenberger's excellent expository article [62] and to Khare's exposition [35], which has somewhat fewer details but paints the big picture in bold strokes. For simplicity, we will only consider the level one case of Serre's conjecture. This means that we start with a Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ that is continuous, irreducible, odd, and unramified outside p. Recall from Section 2.5 that the lack of ramification outside p means that the prime-to-p part of the Artin conductor $\mathfrak{n}(\overline{\rho})$ is trivial, and hence $N(\overline{\rho}) = 1$. We aim to prove that $\overline{\rho}$ is modular.

It is important to note that some special cases of Serre's conjecture were known well before Khare's idea of applying Taylor's potential modularity results and Kisin's modularity lifting techniques to this problem. In the 1970's Tate used discriminant bounds to prove that there are no continuous irreducible odd representations $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_2)$, and therefore that the level one case of Serre's modularity conjecture is vacuously true for $p = 2$. Serre extended his argument to $p = 3$ shortly afterwards, and these two results are essential to the work of Khare and Wintenberger, since they constitute the base cases of their induction argument. We note that the level one case of Serre's conjecture for $p = 5$ was proved by Brueggeman [13] contingent on the generalized Riemann hypothesis, and that, with some local hypotheses at small primes but no restriction on the level, the conjecture was proved for $\overline{\rho}$ with image lying in $\mathrm{GL}_2(\mathbb{F}_7)$ by Manoharmayum [40] and for $\overline{\rho}$ with image lying in $\mathrm{GL}_2(\mathbb{F}_9)$ by Ellenberg [20].

To give a taste of the inductive argument that proves Serre's conjecture, and to illustrate its crucial reliance on modularity lifting theorems, we will first flagrantly disregard the current reality and describe what the proof would look like if modularity lifting technology were more advanced than it actually is. Assume the following, for the moment:

Dream 6.1. *Let $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be continuous, irreducible, unramified outside p, and crystalline at p with Hodge-Tate weights $\{w, 0\}$ for some $w \leq 2p$. Suppose that its reduction $\overline{\rho}$ is modular. Then ρ is modular.*

This dream follows, of course, from the Fontaine-Mazur conjecture. It was considered totally out of reach when Khare and Wintenberger did their work, but such a modularity lifting result has since been proved in most cases by Kisin [38]. In fact, since Serre's conjecture is now known, his work implies most cases of Fontaine-Mazur for two-dimensional representations of $G_{\mathbb{Q}}$. In any case, let us assume the dream and then prove Serre's conjecture in level one.

Let p_n be the n-th prime. We will prove Serre's conjecture by induction on n. It is known for $p_1 = 2$ and $p_2 = 3$ by the theorems of Tate and Serre that were mentioned above. Suppose that it is also known for p_{n-1}. Let $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_{p_n})$ be continuous, irreducible, odd, and unramified outside p_n. By the methods of Sections 5.3 and 5.4, we can find a lift ρ of $\overline{\rho}$ that sits in a compatible system $\{\rho_l\}$, so that $\rho_{p_n} \simeq \rho$. Consider $\overline{\rho}_{p_{n-1}}$; it is modular by induction. Moreover, by the properties of compatible systems, $\rho_{p_{n-1}}$ is unramified outside p_{n-1} and is crystalline of Hodge-Tate weight $(0, k(\overline{\rho}) - 1)$. As we saw at the beginning of these notes, up to a twist we can assume that $k(\overline{\rho}) \leq p_n \leq 2p_{n-1}$, where the second inequality is Bertrand's Postulate. By the Dream, $\rho_{p_{n-1}}$ is modular. Hence ρ_{p_n} is modular by the compatible system, and hence $\overline{\rho}$ is modular and we are done.

The powerful modularity lifting theorem of the Dream can be seen as a fulfillment of the third part of the wish list of Section 5.1. Even though the Dream is not yet known, the modularity lifting theorems available to Khare and Wintenberger in 2005 were enough to prove Serre's conjecture, albeit with lots of technical work. The modularity lifting theorems available now, and still more those available then, come with long lists of technical hypotheses, and one must be very careful to ensure that the liftings of $\overline{\rho}$ and the compatible systems obtained from the methods of Sections 5.3 and 5.4 satisfy these. In these notes we have entirely ignored these technical points, which complicate the work tremendously. However, at its core the basic idea is the simple one presented here.

We conclude with the unfortunate observation that it does not appear to be possible, at least not without a major new idea, to generalize the beautiful argument of Khare and Wintenberger to obtain a proof of the generalizations of Serre's conjecture to totally real fields that were incorporated into Conjecture 2.9 above. While all the ingredients of their proof – potential modularity, construction of lifts, compatible systems, modularity lifting theorems – are less developed for arbitrary totally real fields than for $\mathbb{Q}$, a more fundamental problem is that any inductive argument on the places of F would require that enough base cases be proved first, and we have no idea how to prove them. Recall that Tate and Serre proved (the level one case of) Serre's conjecture for $p = 2$ and $p = 3$ by showing that it was vacuously true, i.e. that there were no $\overline{\rho}$ that were continuous, irreducible, odd, and unramified outside

p. While analogous non-existence theorems have been proved for some small primes and a few specific quadratic real fields (see, for instance, [41] and [50]), we know that for general totally real fields, even quadratic ones, Serre's conjecture is never vacuously true. Indeed, for general totally real fields, for all p there exist continuous, irreducible, odd mod p Galois representations that are unramified outside p; see the introduction to [14] for an example in the case of $F = \mathbb{Q}(\sqrt{29})$. Thus, to get the base case for an induction argument, one would need to establish a sufficiently large number of non-vacuous cases of Serre's conjecture, and it is not clear at all at the present time how to attack this problem. Serre's modularity conjecture will likely continue to be an important motivation and source of research problems for some time to come.

References

[1] Noriyuki Abe. On a classification of irreducible admissible modulo p representations of a p-adic split reductive group. Preprint, available at `http://arxiv.org/pdf/1103.2525v3`.

[2] Avner Ash and Glenn Stevens. Cohomology of arithmetic groups and congruences between systems of Hecke eigenvalues. *J. Reine Angew. Math.*, 365:192–220, 1986.

[3] Thomas Barnet-Lamb, Toby Gee, and David Geraghty. Serre weights for rank two unitary groups. Preprint, 2012.

[4] L. Barthel and R. Livné. Irreducible modular representations of GL_2 of a local field. *Duke Math. J.*, 75(2):261–292, 1994.

[5] Laurent Berger. Central characters for smooth irreducible modular representations of $\mathrm{GL}_2(\mathbf{Q}_p)$. *Rendiconti del Seminario Matematico della Università di Padova*, 127, 2012.

[6] Tobias Berger and Gergely Harcos. l-adic representations associated to modular forms over imaginary quadratic fields. *Int. Math. Res. Not. IMRN*, (23):Art. ID rnm113, 16, 2007.

[7] Tobias Berger and Krzysztof Klosin. An $R = T$ theorem for imaginary quadratic fields. *Math. Ann.*, 349(3):675–703, 2011.

[8] Gebhard Böckle. A local-to-global principle for deformations of Galois representations. *J. Reine Angew. Math.*, 509:199–236, 1999.

[9] Nigel Boston, Hendrik W. Lenstra, Jr., and Kenneth A. Ribet. Quotients of group rings arising from two-dimensional representations. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(4):323–328, 1991.

[10] Christophe Breuil. Sur quelques représentations modulaires et p-adiques de $\mathrm{GL}_2(\mathbf{Q}_p)$. I. *Compositio Math.*, 138(2):165–188, 2003.

[11] Christophe Breuil and Ariane Mézard. Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbf{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ en $l = p$. *Duke Math. J.*, 115(2):205–310, 2002. With an appendix by Guy Henniart.

[12] Christophe Breuil and Vytautas Paskunas. Towards a modulo p Langlands correspondence for GL_2. *Memoirs Amer. Math. Soc.*, 216, 2012.

[13] Sharon Brueggeman. The nonexistence of certain Galois extensions unramified outside 5. *J. Number Theory*, 75(1):47–52, 1999.

[14] Kevin Buzzard, Fred Diamond, and Frazer Jarvis. On Serre's conjecture for mod l Galois representations over totally real fields. *Duke Math. J.*, 55:105–161, 2010.

[15] Frank Calegari and David Geraghty. Modularity lifting beyond the Taylor-Wiles method. Preprint.

[16] Henri Carayol. Sur la mauvaise réduction des courbes de Shimura. *Compositio Math.*, 59(2):151–230, 1986.

[17] Fred Diamond. The Taylor-Wiles construction and multiplicity one. *Invent. Math.*, 128(2):379–391, 1997.

[18] Bas Edixhoven. The weight in Serre's conjectures on modular forms. *Invent. Math.*, 109(3):563–594, 1992.

[19] Bas Edixhoven. Serre's conjecture. In *Modular forms and Fermat's last theorem (Boston, MA, 1995)*, pages 209–242. Springer, New York, 1997.

[20] Jordan S. Ellenberg. Serre's conjecture over $\mathbb{F}_9$. *Ann. of Math. (2)*, 161(3):1111–1142, 2005.

[21] Matthew Emerton, Toby Gee, and Florian Herzig. Explicit Serre weight conjectures. In preparation.

[22] Kazuhiro Fujiwara. Galois deformations and arithmetic geometry of Shimura varieties. In *International Congress of Mathematicians. Vol. II*, pages 347–371. Eur. Math. Soc., Zürich, 2006.
[23] Toby Gee. On the weights of mod p Hilbert modular forms. *Invent. Math.*, 184:1–46, 2011.
[24] Toby Gee and Mark Kisin. The Breuil-Mézard conjecture for potentially Barsotti-Tate representations. Preprint, 2012.
[25] Toby Gee, Tong Liu, and David Savitt. Crystalline extensions and the weight part of Serre's conjecture. *Algebra and Number Theory*, To appear.
[26] Toby Gee and David Savitt. Serre weights for mod p Hilbert modular forms: the totally ramified case. *J. Reine Angew. Math.*, 660:1–26, 2011.
[27] J. A. Green. The characters of the finite general linear groups. *Trans. Amer. Math. Soc.*, 80:402–447, 1955.
[28] Michael Harris, David Soudry, and Richard Taylor. l-adic representations associated to modular forms over imaginary quadratic fields. I. Lifting to $\mathrm{GSp}_4(\mathbf{Q})$. *Invent. Math.*, 112(2):377–411, 1993.
[29] Michael Harris and Richard Taylor. *The geometry and cohomology of some simple Shimura varieties*, volume 151 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2001. With an appendix by Vladimir G. Berkovich.
[30] Guy Henniart. Une preuve simple des conjectures de Langlands pour $\mathrm{GL}(n)$ sur un corps p-adique. *Invent. Math.*, 139(2):439–455, 2000.
[31] Florian Herzig. The weight in a Serre-type conjecture for tame n-dimensional Galois representations. *Duke Math. J.*, 149:37–116, 2009.
[32] Florian Herzig. The classification of irreducible admissible mod p representations of a p-adic GL_n. *Invent. Math.*, 186:373–434, 2011.
[33] Florian Herzig. A Satake isomorphism in characteristic p. *Compos. Math.*, 147(1):263–283, 2011.
[34] Frazer Jarvis. On Galois representations associated to Hilbert modular forms of low weight. *J. Reine Angew. Math.*, 491:199–216, 1997.
[35] Chandrashekhar Khare. Serre's modularity conjecture: a survey of the level one case. In *L-functions and Galois representations*, volume 320 of *London Math. Soc. Lecture Note Ser.*, pages 270–299. Cambridge Univ. Press, Cambridge, 2007.
[36] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. I. *Invent. Math.*, 178(3):485–504, 2009.
[37] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. II. *Invent. Math.*, 178(3):505–586, 2009.
[38] Mark Kisin. The Fontaine-Mazur conjecture for GL_2. *J. Amer. Math. Soc.*, 22(3):641–690, 2009.
[39] Mark Kisin. Modularity of 2-adic Barsotti-Tate representations. *Invent. Math.*, 178(3):587–634, 2009.
[40] Jayanta Manoharmayum. Serre's conjecture for mod 7 Galois representations. In *Modular curves and abelian varieties*, volume 224 of *Progr. Math.*, pages 141–149. Birkhäuser, Basel, 2004.
[41] Hyunsuk Moon and Yuichiro Taguchi. The non-existence of certain mod 2 Galois representations of some small quadratic fields. *Proc. Japan Acad. Ser. A Math. Sci.*, 84(5):63–67, 2008.
[42] Laurent Moret-Bailly. Groupes de Picard et problèmes de Skolem. I, II. *Ann. Sci. École Norm. Sup. (4)*, 22(2):161–179, 181–194, 1989.
[43] Michael Rapoport. Compactifications de l'espace de modules de Hilbert-Blumenthal. *Compositio Math.*, 36(3):255–335, 1978.
[44] Michel Raynaud. Schémas en groupes de type $(p, \dots, p)$. *Bull. Soc. Math. France*, 102:241–280, 1974.
[45] Michael M. Schein. Weights in Serre's conjecture for Hilbert modular forms: the ramified case. *Israel J. Math.*, 166:369–391, 2008.
[46] Michael M. Schein. Weights of Galois representations associated to Hilbert modular forms. *J. Reine Angew. Math*, 622:57–94, 2008.
[47] Michael M. Schein. Reduction modulo p of cuspidal representations and weights in Serre's conjecture. *Bull. London Math. Soc.*, 41:147–154, 2009.

[48] Michael M. Schein. Weights in generalizations of Serre's conjecture and the mod p local Langlands correspondence. In *Symmetries in algebra and number theory*, pages 71–93. Universitätsverlag Göttingen, Göttingen, 2009.

[49] Benjamin Schraen. Sur la présentation des répresentations supersingulières de $\mathrm{GL}_2(F)$. Preprint, 2012.

[50] Mehmet Haluk Şengün. The nonexistence of certain representations of the absolute Galois group of quadratic fields. *Proc. Amer. Math. Soc.*, 137(1):27–35, 2009.

[51] Jean-Pierre Serre. *Corps locaux*. Hermann, Paris, 1968. Deuxième édition, Publications de l'Université de Nancago, No. VIII.

[52] Jean-Pierre Serre. Valeurs propres des opérateurs de Hecke modulo l. In *Journées Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, 1974)*, pages 109–117. Astérisque, Nos. 24–25. Soc. Math. France, Paris, 1975.

[53] Jean-Pierre Serre. *Linear representations of finite groups*. Springer-Verlag, New York, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.

[54] Jean-Pierre Serre. Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987.

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

[56] J. Tate. 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.

[57] Richard Taylor. l-adic representations associated to modular forms over imaginary quadratic fields. II. *Invent. Math.*, 116(1-3):619–643, 1994.

[58] Richard Taylor. Remarks on a conjecture of Fontaine and Mazur. *J. Inst. Math. Jussieu*, 1(1):125–143, 2002.

[59] Richard Taylor and Andrew Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995.

[60] Marie-France Vignéras. La conjecture de Langlands locale pour $\mathrm{GL}(n, F)$ modulo l quand $l \neq p$, $l > n$. *Ann. Sci. École Norm. Sup. (4)*, 34(6):789–816, 2001.

[61] Andrew Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995.

[62] Jean-Pierre Wintenberger. La conjecture de modularité de Serre: le cas de conducteur (d'après C. Khare). *Astérisque*, (311):Exp. No. 956, viii, 99–121, 2007. Séminaire Bourbaki. Vol. 2005/2006.

Department of Mathematics, Bar-Ilan University, Ramat Gan 52900, Israel
E-mail address: mschein@math.biu.ac.il