Motives and L-functions

Frank Calegari

ABSTRACT. We survey some conjectures and recent developments in the Langlands program, especially on the conjectures linking motives and Galois representations with automorphic forms and L-functions. We then give an idiosyncratic discussion of various recent modifications of the Taylor–Wiles method due in part to the author in collaboration with David Geraghty, and then give some applications. The final goal is to give some hints about ideas in recent joint work with George Boxer, Toby Gee, and Vincent Pilloni. This is an expanded version of the author's talk in the 2018 Current Developments in Mathematics conference. It will be presented in three parts: exposition, development, and capitulation.

CONTENTS

1. Part I: Introduction

1.1. While you sit here (as I give the talk). Ben Green once explained to me his philosophy about mathematics, which, to paraphrase in a way that the source may possibly disavow, was that he only wanted to prove theorems which he could explain (the statement) to an undergraduate. So let me start with such a problem. Fix a polynomial $f(x)$ of degree d with integer coefficients. Let us assume also that the discriminant Δ of $f(x)$ is non-zero, which is equivalent to requiring that $f(x)$ has no repeated roots. For a prime p, let

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & a \equiv 0 \mod p, \\ 1, & a \equiv n^2 \mod p \text{ for some } n \not\equiv 0 \mod p, \\ -1, & \text{otherwise.} \end{cases}$$

This is the quadratic residue symbol. Now consider the following sum as p varies over primes not dividing Δ:

$$b_p := \sum_{n=0}^{p-1} \left(\frac{f(n)}{p}\right) \in \mathbf{Z}.$$

One might guess that the values $f(n)$ — when they are non-zero — will be squares half the time and non-squares half the time. It follows that one expects b_p to be small, at least in comparison to p. This (and much more) is true, but it is not our concern here. If $f(x)$ is linear, then b_p is exactly zero. If $f(x)$ is quadratic, then b_p is always equal to ± 1, in a way that depends precisely on whether the leading coefficient of $f(x)$ is a square or not. In particular, if

$$a_p = b_p + \begin{cases} \left(\dfrac{x^d \text{ coefficient of } f(x)}{p}\right), & d \text{ even}, \\ 0, & d \text{ odd}, \end{cases}$$

then $a_p = 0$ whenever $d = 1$ or $d = 2$. On the other hand, we have the following:

THEOREM 1.1.1. *Let $f(x)$ be any polynomial of degree d with non-zero discriminant. Suppose that $d = 3$, 4, 5, or 6. Then $a_p > 0$ for infinitely many p, and $a_p < 0$ for infinitely many p.*

We say more about this theorem in §1.13, but while you are still reading, let me make a few remarks. It is relatively easy to show that, for any $f(x)$ of degree at least 3, then *at least one* of the two desired statements always holds. The actual theorems which have been proved in this case are significantly stronger, but already the theorem in the form presented above (for $d = 5$ and 6) was completely open before this year, even (say) for $f(x) = x^5 + x + 1$. One certainly expects this result to hold for all $d \geq 3$, but the general case seems totally hopeless — including the general case for $d = 7$. Even worse, it seems equally hopeless to show that there are *any explicit squarefree polynomials at all* with $a_p > 0$ for infinitely many p and of sufficiently large degree. However, we offer the following exercise for the duration of this talk:

EXERCISE 1.1.2. *Find a* single *polynomial $f(x)$ of degree $d = 3$ for which you can prove Theorem 1.1.1 in an elementary a way as possible. Extra points if you avoid invoking any theorems less than 100 years old.*

Kevin Buzzard points out the following: if $f(x)$ is any polynomial, then the theorem holds for $g(x) = Df(x)$ for all but *at most one* squarefree integer D. The point is that

$$a_p(D) := a_p(g(x)) = \left(\frac{D}{p}\right) a_p(f(x)),$$

and so half the signs for $a_p(D)$ will be different from $a_p(D')$ whenever D/D' is not a square.

1.2. Acknowledgments. I first and foremost thank my coauthors George Boxer, Toby Gee, and Vincent Pilloni; without them, none of this would have been possible. I thank, too, other coauthors who have helped shaped my understanding over the years of the Taylor–Wiles method both through conversations and through collaboration, in particular both Matthew Emerton and David Geraghty. Parts of these notes are lifted from [**CG18**] and [**BCGP18**], including but not limited to parts of the introduction I wrote for [**BCGP18**] but were then overwritten by my coauthors. Finally, I would like to thank Toby Gee, Matthew Emerton, and Eric Stubley for looking through this manuscript and pointing out a number of typos and minor inaccuracies.

1.3. Introduction. Let X be an algebraic variety cut out by a system of polynomial equations with rational coefficients. Let us suppose (mostly for convenience of exposition) that X is smooth and proper. (This assumption will be in effect throughout, except when it is not convenient.) The complex points $X(\mathbf{C})$ of X cut out a smooth compact orientable manifold of dimension $2 \dim(X)$. It seems to be a fact of nature that the geometry of $X(\mathbf{C})$ (thought of in any number of ways: topologically, geometrically, algebraically) is deeply linked to the arithmetic of X. Possibly the ur-example

of this relationship is the Mordell conjecture proved by Faltings [**Fal83**]:
if $X(\mathbf{C})$ is a surface of genus at least two, then $X(F)$ is finite for any num-
ber field $F \subset \mathbf{Q}$. One may conversely ask how the arithmetic of X relates to
the geometry of X. For example, the variety X is cut out by a finite number
of equations with coefficients in $\mathbf{Q}$. One may clear the denominators so that
the coefficients of these equations lie in $\mathbf{Z}$. It is an elementary fact that, for
all but finitely many primes l, the mod-l reduction of these equations de-
fines a smooth variety $X_{\mathbf{F}_l}$ over $\mathbf{F}_l$. For every such prime l, one can count
the number $N_l := \#X(\mathbf{F}_l)$ of points of X which are defined over $\mathbf{F}_l$. (This
passage from $\mathbf{Q}$ to $\mathbf{Z}[1/S]$ goes under the fancy name "spreading out.") For
example, if $X = \mathbf{P}_{\mathbf{Q}}^1$, then

$$N_l = l + 1$$

for every prime l. Then one can ask: how close do the numbers $\#X(\mathbf{F}_l)$ come
to determining X? Furthermore, to the extent that the latter question has a
positive answer, one can further ask: what do the numbers N_l tell us about
the arithmetic and geometry of X? Some of these questions lead directly
to central conjectures in the Langlands program and beyond. However, we
prefer to take not one but two steps back, and begin in a context broader
than the Langlands program.

1.4. Cycle Maps. Suppose we forget (for a moment) the fact that X is
defined over $\mathbf{Q}$, and consider it merely as a smooth algebraic variety over $\mathbf{C}$.
It is a maxim that there are only two things one can do in mathematics:
combinatorics and linear algebra. In particular, geometry is hard. But there
is a universal way to linearize geometry, namely, to consider cohomology. So
it is very natural to study the cohomology groups $H^*(X(\mathbf{C}), \mathbf{Z})$ and ask how
the information in these groups translates into information about X. Our
smoothness and properness assumptions guarantee that $X(\mathbf{C})$ is a compact
orientable manifold. In topology, one usually thinks of homology classes of a
manifold as being represented (possibly virtually) by sub-manifolds (possibly
immersed). For algebraic varieties, there is a corresponding notion of an al-
gebraic cycle which we discuss below. The first observation is that one can't
possibly hope to represent all homology or cohomology classes as coming
from algebraic subvarieties of X, because these subvarieties will be complex
manifolds and so have even dimension. On the other hand, it is certainly
possible for $X(\mathbf{C})$ to have cohomology in odd degrees. (If X is a curve of
genus g, then $H^1(X(\mathbf{C}), \mathbf{Z}) = \mathbf{Z}^{2g}$, where g is the genus of X.) One might still
ask *which* cohomology classes of $X(\mathbf{C})$ arise in this way. Being a little more
precise, one can define an algebraic version of this is as follows. Let $Z_i(X)$ de-
note the free abelian group generated by i-dimensional subvarieties (i-cycles)
on X. One notion of equivalence between cycles (rational equivalence) is to
mod out by the divisors of rational functions on $i + 1$-dimensional subvari-
eties, and the corresponding quotient is the Chow group $\mathrm{CH}_i(X)$. For ex-
ample, if $i = \dim(X) - 1$, then $\mathrm{CH}_{n-1}(X)$ is the divisor class group of X.
If $X \subset \mathbf{P}^N$ is projective, then there is a class in $\mathrm{CH}_{n-1}(X)$ given by the class

of a hyperplane (note that all hyperplanes are equivalent, by considering the restriction to X of the rational function given by the ratio of the linear forms vanishing on each hyperplane.) Assuming that X is smooth of dimension d, one may define $\mathrm{CH}^i(X)$ to be the Chow group on codimension i cycles, which is just $\mathrm{CH}_{d-i}(X)$. In particular, for a smooth projective variety X, the hyperplane class gives an element of $\mathrm{CH}^1(X)$. The groups $\mathrm{CH}^*(X)$ turn out to form a commutative (graded) ring, with the product of two classes (roughly) corresponding to taking the class of their intersection (this is literally correct if the classes are transverse). There now turns out to be a desired cycle map

$$\mathrm{CH}^i(X) \to H^{2i}(X(\mathbf{C}), \mathbf{Z}) \to H^{2i}(X(\mathbf{C}), \mathbf{Q}).$$

As mentioned above, there is no hope that cycles will be able to capture all the cohomology of X. But there are further obstructions on a cohomology class to turn up in the image of the cycle class map. Recall first of all that de Rham cohomology gives an identification between $H^{2i}(X(\mathbf{C}), \mathbf{C})$ and $H^{2i}_{\mathrm{dR}}(X(\mathbf{C}), \mathbf{C})$. Because $X(\mathbf{C})$ is a compact Riemannian manifold, all cohomology classes have harmonic representatives. However, the complex structure gives a further refinement of these forms. Let us examine this more closely in the case of a genus one curve. If X has genus one, then the complex points $X(\mathbf{C})$ may be identified (as a complex manifold) with a torus $\mathbf{C}/\Lambda$. The one forms in $H^1_{\mathrm{dR}}(X(\mathbf{C}), \mathbf{C})$ have a particularly nice description: they are spanned by the differentials dx and dy. On the other hand, the complex structure of $X(\mathbf{C})$ singles out a canonical line inside this space spanned by the *holomorphic* differential

$$dz = dx + idy.$$

Moreover, there is also a canonical complement to this line, given by the span of the *anti-holomorphic* differential $d\bar{z} = dx - idy$. This decomposition is something which applies very generally to all de Rham cohomology groups of complex manifolds. Namely, there is a Hodge decomposition

$$H^n_{\mathrm{dR}}(X(\mathbf{C}), \mathbf{C}) = \bigoplus_{p+q=n} H^{p,q}(X(\mathbf{C}), \mathbf{C}),$$

where $H^{p,q}$ is the subspace of n forms which, locally at any point, consists of harmonic n forms which may be written as a harmonic function times a differential of the form

$$dz_1 \wedge \ldots \wedge dz_p \wedge d\bar{z}_{p+1} \wedge \ldots \wedge d\bar{z}_{p+q}.$$

It is elementary to show that the image of the Chern class map from $\mathrm{CH}^i(X)$ always lands inside $H^{i,i}$. The **Hodge Conjecture** is, as a first approximation, a converse to this statement.

CONJECTURE 1.4.1 (Hodge Conjecture). *Every class in $H^{2i}(X(\mathbf{C}), \mathbf{Q}) \cap H^{i,i}(X(\mathbf{C}), \mathbf{C})$ is a (rational) linear combination of elements in the image of the cycle class map.*

Note that we consider $\mathbf{Q}$ rather than $\mathbf{Z}$ coefficients and allow rational linear combinations of cycles. What is essential in the formulation of this conjecture is that we restrict to coefficients in $\mathbf{Q}$. Namely, the intersection $H^{2i}(X(\mathbf{C}), \mathbf{C}) \cap H^{i,i}(X(\mathbf{C}), \mathbf{C})$ is obviously $H^{i,i}(X(\mathbf{C}), \mathbf{C})$, but it is often the case that this group is non-trivial but $H^{2i}(X(\mathbf{C}), \mathbf{Q}) \cap H^{i,i}(X(\mathbf{C}), \mathbf{C})$ *is* trivial, or at least of dimension smaller than $\dim H^{i,i}$.

One extreme example of the Hodge conjecture is when *all* the cohomology is of type $H^{i,i}$. In this case, the Hodge conjecture predicts that all the cohomology of X is explained by algebraic cycles. One example of such a variety is projective space, but there are plenty more:

EXAMPLE 1.4.2 (Spherical Varieties). Let X be a projective variety with an action of a connected reductive group G, and suppose that X contains a dense orbit for some Borel $B \subset G$. Then $H^{p,q} = 0$ unless $p = q$.

This includes a wide variety of interesting varieties, for example, Grassmannians, general flag varieties, and projective toric varieties. In these cases, there are often very natural sources of algebraic cycles, indeed, usually enough to show the Hodge Conjecture is true. The study of these varieties is often very well understood in terms of geometric combinatorics by exploiting the explicit nature of these geometric cycles. These varieties are less obviously interesting from an "arithmetic" perspective; for example, as we shall see below, the point counts N_p of these varieties X essentially contain little information beyond the dimension of the cohomology groups $H^*(X(\mathbf{C}), \mathbf{C})$.

1.5. Varieties over number fields. Let us now remember that our smooth projective variety X is defined over a number field (which we assume is $\mathbf{Q}$). A natural question is how to exploit this fact. The first remark is that the absolute Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of $\mathbf{Q}$ acts on X thought of as a variety over $\overline{\mathbf{Q}}$. Since we are still interested in considering linear data associated to X, it is natural to consider how this action interacts with the cohomology groups $H^*(X(\mathbf{C}), \mathbf{Z})$. One problem is that the $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ action does not extend into any natural way from $X/\overline{\mathbf{Q}}$ to $X(\mathbf{C})$, so there is no obvious action on cohomology via transport of structure. The group $H^0(X(\mathbf{C}), \mathbf{Z})$ is a free abelian group on the connected components of $X(\mathbf{C})$. It is an exercise to show that the components of $X(\mathbf{C})$ actually coincide with the irreducible algebraic components of $X/\overline{\mathbf{Q}}$. But these components are certainly all defined over some finite extension of $\mathbf{Q}$, and are permuted by the Galois action. So there is always an honest Galois action on $H^0(X(\mathbf{C}), \mathbf{Z})$. This construction does not work for higher cohomology. There is, however, a suitable replacement, given by the étale cohomology groups $H^*_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$. Let us give a very rough hint of how to construct such groups in the case of H^1. The first step is to give a definition of cohomology which has a suitably algebraic flavour (since, as we have previously seen, representing classes by algebraic subvarieties cannot possibly work). One nice description of H^1 is in terms of abelian covering spaces $Y \to X$. Here the notion of a covering space has a

clean analogue, at least for maps of finite degree (étale morphisms). One can then hope to prove that any finite topological covering map can be realized geometrically over a finite extension of $\mathbf{Q}$, and hence give rise to a well defined group $H^1_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Z}/n\mathbf{Z})$. The corresponding cover Y (together with the group of deck transformations) then has a natural action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, which translates into an action of this group on $H^1_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Z}/n\mathbf{Z})$. One now defines

$$H^1_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) := \mathrm{proj}\lim H^1_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Z}/l^m\mathbf{Z}) \otimes \mathbf{Q}_p.$$

The groups $H^*_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ have the following felicitous properties, which includes the consequences of the Weil Conjectures as proved by Deligne [**Del74**]

(1) There are natural continuous actions of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $H^*_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$.

(2) There are isomorphisms of vector spaces

$$H^*_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \simeq H^*(X(\mathbf{C}), \mathbf{Q}_p).$$

(3) If $l \neq p$ is a prime of good reduction for X, then, choosing an embedding of $\overline{\mathbf{Q}}$ into $\overline{\mathbf{Q}}_l$, there is a natural action of $\mathrm{Gal}(\overline{\mathbf{Q}}_l/\mathbf{Q}_l)$ on $H^*_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$, and this action factors through the unramified quotient group of the finite field $\mathrm{Gal}(\overline{\mathbf{F}}_l/\mathbf{F}_l) = \widehat{\mathbf{Z}}$ which is canonically topologically generated by the Frobenius element Frob_l.

(4) The fixed points of Frob_l^m on $X/\overline{\mathbf{F}}_l$ are precisely the points $X(\mathbf{F}_q)$ where $q = l^m$. The number of fixed points of Frob_q is given by the formula

$$\sum_{i=0}^{2\dim(X)} (-1)^i \mathrm{Tr}\left(\mathrm{Frob}_q \,\big|\, H^i_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)\right).$$

(5) There is a Poincaré duality perfect pairing

$$H^i_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \times H^{2d-i}_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \to H^{2d}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \simeq \mathbf{Q}_p(-d),$$

where $d = \dim(X)$, and $\mathbf{Q}_p(-d)$ is the one dimensional vector space where Frob_l acts by l^{-d}.

(6) The eigenvalues of Frob_l on $H^i_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ have absolute value $l^{-i/2}$.

There are a number of immediate consequences of this result for the sequence of numbers $N_l = \#X(\mathbf{F}_l)$.

LEMMA 1.5.1. *The data of the numbers $\{N_l\}$ is precisely equivalent to the data of the Galois representations $H^i_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ for any prime p up to semi-simplification.*

This is essentially a consequence of the Brauer–Nesbitt theorem. It is certainly expected (see below) that the Galois representations associated to projective varieties should all be semi-simple.

Let us now return to cycle maps. For any $X/\mathbf{Q}$, the Chow Groups have natural definitions over any finite extension $F/\mathbf{Q}$. Moreover, there exist p-adic cycle maps

$$\mathrm{CH}^i(X/F) \to H^{2i}_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p(i)).$$

The twist in the coefficients can be pulled out of the cohomology groups, and only serves to twist the Galois action on the right hand side so that the resulting map is actually compatible with the action of $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. In particular, the image of the cycle map is invariant under $\text{Gal}(\overline{\mathbf{Q}}/F)$. In this setting, the analog of the Hodge Conjecture is the Tate Conjecture:

CONJECTURE 1.5.2 (Tate Conjecture). *The image of* $\text{CH}^i(X/F) \otimes \mathbf{Q}_p$ *under the cycle map is precisely the* $\text{Gal}(\overline{\mathbf{Q}}/F)$*-invariant classes in* $H^{2i}_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p(i))$.

If the action of $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on a class in $H^{2i}_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ has Frob_p for all p acting by p^{-i}, then we say that the class is Tate. If the same holds for all Frob_p in $\text{Gal}(\overline{\mathbf{Q}}/F)$ for some finite extension $F/\mathbf{Q}$, we call the class potentially Tate. The Tate conjecture then says that all potentially Tate classes are explained by algebraic cycles.

One interpretation of the Tate conjecture is that the classification of varieties over $\mathbf{Q}$ can be broken up into two pieces; one, which deals with the linear algebra, consists of understanding Galois representations, and the other deals with the combinatorics of understanding cycles. Moreover, the data one can recover from X is precisely the information corresponding to the latter data together with the dimensions of the corresponding cohomology groups.

1.6. p-adic Hodge Theory. The Tate Conjecture and the Hodge Conjecture are obviously of a similar flavour, but they are not directly related. The problem is that (as stated) the Hodge decomposition relies heavily on the complex description of X, whereas the étale cohomology relies on a p-adic description on X. The subject of p-adic Hodge Theory was created (in large) part to provide a p-adic version of this story. The first step is to recall the algebraic construction of de Rham cohomology. The construction here is not too surprising. The key point of the usual definition of de Rham cohomology is the Poincaré lemma that closed forms on an open ball are exact. Sheafifying this construction, one may consider the corresponding de Rham complex (of sheaves):

$$\mathcal{O}_X \to \Omega^1_X \to \Omega^2_X \to \dots$$

In the category of smooth manifolds, this is a sequence of acyclic sheaves which is exact. In the algebraic world, neither of these things are true — the sheaves Ω^i_X can have higher cohomology, and the sequence fails to be exact because the Zariski opens are "too big" for the Poincaré lemma to hold. None the less, it turns out that the correct way to define $H^*_{\text{dR}}(X/\mathbf{Q}_p)$ is as the hypercohomology of this complex. There is a corresponding spectral sequence

$$H^p(X, \Omega^q) \Rightarrow H^{p+q}_{\text{dR}}(X).$$

The degeneration of this spectral sequence gives a filtration on de Rham cohomology with filtered pieces $H^{p,q} = H^p(X, \Omega^q)$. Over $\mathbf{C}$, these pieces may be identified with the Hodge decomposition described before.

One of the most influential developments in arithmetic geometry over the past 50 years has been the realization that there is a *p*-adic version of Hodge Theory, and that this has direct links to arithmetic. This subject had it genesis in the paper of Tate [**Tat67**], but was otherwise fashioned almost single handedly by Fontaine [**Fon82, Fon90, Fon94a, Fon94b**]. To give a very rough exposition, first note that one can define the de Rham cohomology groups $H_{\mathrm{dR}}^*(X)$ in a more algebraic way, and in particular, thinking of X as a variety over $\mathbf{Q}_p$, give well defined groups $H_{\mathrm{dR}}^*(X, \mathbf{Q}_p)$. These groups admit a Hodge filtration. If X is smooth over $\mathbf{Z}_p$, then these groups even turn out only to depend on the special fibre of X and so admit the action of an extra operator coming from Frobenius. As we shall see later, any information one can recover from the de Rham cohomology groups should also be encoded in the étale cohomology groups $H_{\mathrm{et}}^*(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$. It is thus natural to ask how one can do this directly. We have no time to do this subject any justice, so let us confine ourselves to some cursory remarks. The first is as follows. If $\mathbf{C}_p$ is the completion of $\overline{\mathbf{Q}}_p$, then it turns out that $\mathbf{C}_p$ is complete and algebraically closed, and admits an action of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. This induces an action of this group on $H_{\mathrm{et}}^*(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \otimes_{\mathbf{Q}_p} \mathbf{C}_p$, where $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ acts on both factors. The first highly non-trivial fact is that there is an isomorphism

$$H_{\mathrm{et}}^m(X/\overline{\mathbf{Q}}, \mathbf{Q}_p) \otimes_{\mathbf{Q}_p} \mathbf{C}_p \simeq \bigoplus_{p+q=m} H^p(X, \Omega^q) \otimes \mathbf{C}_p(q).$$

In particular, one *can* recover the Hodge numbers from étale cohomology. Much more is true — one can recover the entire de Rham cohomology groups along with their filtration. To do this, one still needs to enlarge coefficients, but instead of tensoring with $B_{\mathrm{HT}} := \bigoplus \mathbf{C}_p(k)$, one tensors with the ring B_{dR}, a filtered ring with a $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$-action whose associated graded is B_{HT}. The corresponding functor

$$V \mapsto D_{\mathrm{dR}}(V) = (V \otimes B_{\mathrm{dR}})^{\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)}$$

maps vector spaces over $\mathbf{Q}_p$ with a $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$-action to *p*-adic over $\mathbf{Q}_p$ of dimension $\leq \dim_{\mathbf{Q}_p}(V)$ which acquire a filtration after tensoring with some finite extension $K/\mathbf{Q}_p$.

DEFINITION 1.6.1. A *p*-adic representation V of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is *de Rham* if $\dim D_{\mathrm{dR}}(V) = \dim(V)$.

The implied theorem above says that Galois representations in étale cohomology are de Rham.

In fact, there is an entire menagerie of functors constructed by Fontaine. If X is smooth over $\mathbf{Z}_p$, then, as mentioned above, the de Rham cohomology can be identified with the crystalline cohomology of the special fibre, and there is a corresponding functor $D_{\mathrm{cris}}(V)$ which admits an action by Frobenius. There is also a functor $D_{\mathrm{pst}}(V)$ which is modeled on the cohomology of varieties which admit a semistable model over a finite extension. Quite possibly all smooth varieties $X/\mathbf{Q}_p$ admit a semistable model over a finite

extension $K/\mathbf{Q}_p$, but this is probably tricky — fortunately any Galois representation V coming from the étale cohomology of such an X behaves as if it does ([**Tsu99**]), in the sense that $\dim D_{\mathrm{pst}}(V) = \dim V$ for such V.

1.7. Zero Dimensional Varieties. One interesting basic case to address are smooth algebraic varieties of dimension zero over $\mathbf{Q}$. Certainly, as algebraic varieties over $\mathbf{C}$, such varieties are simply given by d points for some integer d. The structure of $X/\mathbf{Q}$ is then purely determined by the Galois descent data, or equivalently, by the action of a permutation representation

$$\varrho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to S_d$$

which is well defined up to conjugation (on the source and the target). For example, one could consider the solutions to a separable polynomial $f(x) = 0$, and then consider the action of the Galois group on the roots. Associated to ϱ is then a corresponding *linear* representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_d(\mathbf{Q}).$$

In a different optic, one can think of ρ as the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the vector space $H^0(X(\mathbf{C}), \mathbf{Q})$ which is generated by the components of X. In particular, by extending coefficients to $\mathbf{Q}_p$, the representation $\rho_p = \rho \otimes \mathbf{Q}_p$ is the representation:

$$\rho_p : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(H^0_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)).$$

The representations ϱ and ρ factor through a minimal finite quotient $\mathrm{Gal}(L/\mathbf{Q})$ for some finite extension $L/\mathbf{Q}$ with Galois group G. The number of points of X modulo p (at least for a prime of good reduction) is precisely the number of fixed points of the Frobenius element $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. By standard Galois theory, there is a corresponding conjugacy class of elements also denoted by Frob_p in $\mathrm{Gal}(L/\mathbf{Q})$. The number of fixed points of $g \in G$ under a permutation representation ϱ is precisely the trace of the corresponding linearized representation ρ, and so

$$N_p = \mathrm{Tr}(\rho_p(\mathrm{Frob}_p)).$$

The Cebotarev density theorem implies that any conjugacy class in G is represented by Frob_p (infinitely often), and so the data of N_p is enough to recover the values $\mathrm{Tr}(\rho_p(g))$ for all $g \in G$, and hence to recover the representation ρ_p. The Tate conjecture (and the Hodge conjecture) are both obvious in this setting.

It is worthwhile to note, however, that the data of ρ_p for any (or all l) still falls short of recovering X up to isomorphism. Certainly ρ determines the finite quotient G and the Galois extension $L/\mathbf{Q}$. However, the data of ρ is not enough to recover the permutation representation ϱ. For example, suppose that ϱ is a transitive representation. Let H be the one point stabilizer of ρ (well defined up to conjugation). Then one has

$$\rho = \mathrm{Ind}_H^G \mathbf{C}.$$

It turns out that the representation $\operatorname{Ind}_H^G \mathbf{C}$ does *not* always determine the conjugacy class of H. Since the fixed field $K = L^H$ is determined up to isomorphism by the conjugacy class of H, this means that there exist pairs of non-isomorphic fields $K = \mathbf{Q}(\theta)$ and $K' = \mathbf{Q}(\theta')$ which give rise to varieties X and X' with the same Galois representations ρ_p acting on $H_{\text{et}}^0(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$, and so have the same number of points modulo l for all primes l. But this is a "combinatorial" problem in finite group theory rather than anything to do with arithmetic.

1.8. Elliptic Curves. Let us move on to the special case of elliptic curves over a number field K. The Tate conjecture doesn't say anything particularly interesting about the cohomology of $H_{\text{et}}^*(A/\overline{\mathbf{Q}}, \mathbf{Q}_p)$, since Tate classes can only exist in H^0 and H^2, and they moreover do exist in these degrees and are trivially given by the image of cycles. What *is* interesting, however, is to consider $X = A \times B$ for elliptic curves A and B. By the Künneth formula (and using the fact that $H_{\text{et}}^1(B, \mathbf{Q}_p)$ is self-dual up to a twist by $\mathbf{Q}_p(1)$) there is an identification

$$H_{\text{et}}^2(A \times B, \mathbf{Q}_p(1)) = \mathbf{Q}_p^2 \oplus \operatorname{Hom}(H_{\text{et}}^1(A, \mathbf{Q}_p), H_{\text{et}}^1(B, \mathbf{Q}_p)).$$

There are two obvious classes in $\operatorname{CH}^1(A \times B/K)$ coming from A times a point and a point times B, and these generate the $\mathbf{Q}_p^2$ factor under the class map. If A is isogenous to B, then the image of the graph of the isogeny gives an isomorphism between the corresponding cohomology groups. But now suppose that the cohomology groups are isomorphic. Then the Tate conjecture predicts the existence of a corresponding class $[c]$ in $\operatorname{CH}^1(A \times B)$, which is geometrically represented after scaling by (a linear combination of) cycles C. Such a C comes with projections $\pi_A : C \to A$ and $\pi_B : C \to B$, and thus induced maps (Albanese and Picard functoriality) $\pi_A^* : A \to \operatorname{Jac}(C)$ and $\pi_{B,*} : \operatorname{Jac}(C) \to B$, and hence an induced isogeny $\phi = \pi_{B,*} \circ \pi_A^* : A \to B$. The corresponding map is non-trivial precisely when $[c]$ is not in the span of the two tautological classes. Note, furthermore, that if $H_{\text{et}}^1(A, \mathbf{Q}_p)$ is reducible as a Galois representation and $\operatorname{Hom}(H_{\text{et}}^1(A, \mathbf{Q}_p), H_{\text{et}}^1(B, \mathbf{Q}_p))$ has two dimensions worth of $\operatorname{Gal}(\overline{\mathbf{Q}}/K)$-invariant classes, we obtain a second isogeny ψ from A to B which is not a linear multiple of the first. Assuming the Tate conjecture, this implies that A and B have complex multiplication over K.

It is instructive also to compare this with a similar calculation with the Hodge conjecture. Let $A = \mathbf{C}/\Lambda_\tau$ and $B = \mathbf{C}/\Lambda_\sigma$. We have

$$\dim(H^{1,1}(A \times B, \mathbf{C})) = 4,$$

and there are always at least two dimensions worth of classes in the image of the cycle map from $\operatorname{CH}^1(A \times B)$ coming from A times a point and a point times B. It follows that $\dim(H^2(\mathbf{Q}) \cap H^{1,1}(\mathbf{C})) = 2$, 3, or 4. If $\dim(H^2(\mathbf{Q}) \cap H^{1,1}(\mathbf{C})) = 2$, then the Hodge conjecture is trivially true, so we assume that it is equal to 3 or 4. If dz_A and $d\overline{z}_A$ are the obvious holomorphic and

anti-holomorphic one forms on A, and dz_B and $d\overline{z}_B$ the corresponding forms on B, then bases for $H^1(A, \mathbf{Z}) \subset H^1_{\mathrm{dR}}(A, \mathbf{C})$ and $H^1(B, \mathbf{Z}) \subset H^1_{\mathrm{dR}}(B, \mathbf{C})$ over $\mathbf{Z}$ are given by

$$\omega_1 = \frac{\tau d\overline{z}_A - \overline{\tau} dz_A}{\tau - \overline{\tau}}, \quad \omega_2 = \frac{dz_A - d\overline{z}_A}{\tau - \overline{\tau}},$$

$$\eta_1 = \frac{\sigma d\overline{z}_B - \overline{\sigma} dz_B}{\sigma - \overline{\sigma}}, \quad \eta_2 = \frac{dz_B - d\overline{z}_B}{\sigma - \overline{\sigma}}.$$

For example, we find that

$$\int_0^1 \omega_1 = 1, \int_0^\tau \omega_1 = 0, \int_0^1 \omega_2 = 0, \int_0^\tau \omega_2 = 1,$$

and similarly with η_1 and η_2. Also, we find that

$$\omega_1 \wedge \omega_2 = -\frac{dz_A d\overline{z}_A}{\tau - \overline{\tau}} = \frac{2i dx_A \wedge dy_A}{\tau - \overline{\tau}} = \frac{dx_A \wedge dy_A}{\mathrm{im}(\tau)},$$

and the latter has volume one over any fundamental domain for A. Clearly the classes $\omega_1 \wedge \omega_2$ and $\eta_1 \wedge \eta_2$ lie in $H^{1,1}(A \times B, \mathbf{C})$. These are the image under the cycle map of the two obvious classes mentioned previously. Thus to compute $H^2(A \times B, \mathbf{Q}) \cap H^{1,1}(A \otimes B, \mathbf{C})$, it suffices to see what (if any) $\mathbf{Q}$-linear combinations of $\omega_i \wedge \eta_j$ for $i, j \in \{1, 2\}$ lie in $H^{1,1}$. That is, it suffices to compute the kernel of the resulting map from $\mathbf{Q}^4$ generated by these four forms to $\mathbf{C}^2$ given by looking at the coefficients of $dz_A \wedge dz_B$ and $d\overline{z}_A \wedge d\overline{z}_B$. The coefficients of these forms for $\omega_1 \wedge \eta_1$, $\omega_1 \wedge \eta_2$, $\omega_2 \wedge \eta_1$, and $\omega_2 \wedge \eta_2$ respectively are the column vectors in the following matrix:

$$\begin{pmatrix} \dfrac{\tau\sigma}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{\tau}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{\sigma}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{1}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} \\[2mm] \dfrac{\overline{\tau}\overline{\sigma}}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{\overline{\tau}}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{\overline{\sigma}}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} & \dfrac{1}{-4\,\mathrm{im}(\tau)\,\mathrm{im}(\sigma)} \end{pmatrix}.$$

We are assuming that $H^2(A \times B, \mathbf{Q}) \cap H^{1,1}(A \otimes B, \mathbf{C}) > 2$, so there exists some $\mathbf{Q}$-linear combination of column vectors which vanishes. Note that the second line is the complex conjugate of the first. Hence we are simply looking for $\mathbf{Q}$ linear combinations of 1, σ, τ, and $\sigma\tau$ which vanish, e.g.

$$a + b\sigma + c\tau + d\sigma\tau = 0.$$

But this implies immediately that

$$\tau = -\frac{a + b\sigma}{c + d\sigma}.$$

This condition is exactly equivalent to the condition that A is isogenous to B. The existence of an isogeny ϕ implies the existence of a geometric cycle. In particular, this proves the Hodge Conjecture in this case if there is an equality $\dim(H^2(\mathbf{Q}) \cap H^{1,1}(\mathbf{C})) = 3$. If $\dim(H^2(\mathbf{Q}) \cap H^{1,1}(\mathbf{C})) = 4$, we

have to produce a different cycle. In this case, however, we deduce that there is a second equality

$$\frac{a + b\sigma}{c + d\sigma} = \frac{p + q\sigma}{r + s\sigma}$$

for a corresponding matrix which is not a scalar multiple of the first. But this implies that $\sigma \in \mathbf{C} \backslash \mathbf{R}$ satisfies a quadratic equation with integer coefficients, which implies that A and B are isogenous CM elliptic curves, and hence both have CM by some order $\mathcal{O}$ in some imaginary quadratic field. But now if $\phi : A \to B$ is any isogeny, then the image in cohomology of the cycle classes corresponding to $[\phi]$ and $[x \circ \phi]$ for any $x \in \mathrm{End}(B)$ with $x \in \mathcal{O} \setminus \mathbf{Z}$ are linearly independent, and we have proven the Hodge Conjecture in this case as well. (It probably would be remiss not to mention at this point that the Hodge Conjecture for H^2 of projective varieties — of which the above is a very special case — follows from the Lefschetz theorem on $(1, 1)$-classes.)

Amazingly enough, the Hodge conjecture is not known even for general CM abelian varieties (although it is known for H^2). To give an example of the subtleties, one could ask whether the Hodge conjecture holds after weakening the condition that we have an abelian variety to simply a complex torus. But then it turns out ([**Zuc77**]) to be false! (see also [**Voi02**]).

Returning to the Tate conjecture, If we now consider curves of higher genus, the situation becomes decidedly more murky. Any non-trivial $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-equivariant homomorphism from $H^1_{\mathrm{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ to $H^1_{\mathrm{et}}(Y/\overline{\mathbf{Q}}, \mathbf{Q}_p)$ for smooth projective curves X and Y certainly implies the existence of an abelian variety $A/\mathbf{Q}$ together with isogenies $A \to \mathrm{Jac}(X)$ and $A \to \mathrm{Jac}(Y)$. But it is not the case (for example) that a surjection on cohomology need be induced by a covering map $X \to Y$. These "combinatorial" aspects of cycles beyond the Tate conjecture, while very interesting, are certainly beyond the scope of this idiosyncratic survey. Even if $\mathrm{Jac}(X) \simeq \mathrm{Jac}(Y)$ as unpolarized abelian varieties, then X need not even be geometrically isomorphic to Y. One beautiful example is given [**How96**, Theorem 1] by the pair of curves

$$C : y^2 = x^5 + x^3 + x^2 - x - 1$$
$$D : y^2 = x^5 - x^3 + x^2 - x - 1,$$

over $\mathbf{F}_3$.

1.9. Motives and the Fontaine–Mazur Conjecture. We have the following cartoon visualization for understanding algebraic varieties given by taking étale cohomology at some prime p.
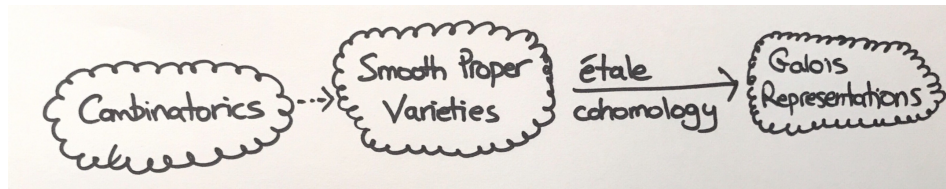
The category of (proper) algebraic varieties is not abelian, but one way to think of this diagram is an exact sequence. The category on the right hand side *is* abelian. So the "linear algebra" part of the classification problem becomes one of determining the image. The Tate conjecture suggests a way to define this category. Namely, it suggests that all decompositions of étale cohomology arise due to the existence of cycles. In particular, given a smooth projective variety X, one first singles out an integer m (the cohomological degree). Then one additionally considers triples (X, π, m), where π is a geometric correspondence. Roughly, suppose one has a Galois invariant subspace $V \subset H^m_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_l)$. By Poincaré duality, there is a perfect pairing

$$H^n_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_l) \times H^{2d-n}_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_l) \to \mathbf{Q}_l(-d).$$

It follows that $V^\vee(-d)$ is a quotient of $H^{2d-n}_{\text{et}}(X/\overline{\mathbf{Q}}, \mathbf{Q}_l)$. Assuming that the cohomology is semi-simple as a Galois module, this implies, by Künneth, that there is an inclusion

$$V \otimes V^\vee \subset H^{2d}(X \times X/\overline{\mathbf{Q}}, \mathbf{Q}_l(d)).$$

However, there is a Galois invariant subspace of $V \otimes V^\vee$, and thus, assuming the Tate conjecture, there exists a corresponding cycle class $\pi \in \text{CH}^d(X \times X, \mathbf{Q})$. Continuing further, the Tate conjecture also suggests the correct way to define morphisms and equivalences in this category. Note that cohomology only exists for $n \geq 0$, so everything included so far is pure of non-negative weight. If one also wants to include duals (so allowing $\mathbf{Z}(n)$ for each n), then one may formally invert the Tate object. In light of Poincaré duality, this suffices to allow duals of all objects. The problem with these definitions is precisely that they require the Tate conjecture to be true to even make sense. There is somehow no way to avoid this problem, however — without a way of proving the existence of appropriate cycles, any conjecture free definition of motives is a mere simulacrum of the truth.

Once one has a theory of Motives (conjectural or otherwise), the Tate conjecture implies that the map from M to Galois representations given by étale cohomology is a fully faithful functor. Assuming this, it then becomes very interesting to determine the image.

CONJECTURE 1.9.1 (Fontaine–Mazur). *The image of pure irreducible motives under étale cohomology (with $\mathbf{Q}_p$-coefficients) consists of continuous irreducible representations V of $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ with the following properties:*

  (1) *The representation V is unramified for all but finitely many primes.*
  (2) *The representation V restricted to $\text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is de Rham.*

One can make this conjecture without referring to motives; one simply predicts that V arises up to Tate twist as a sub-quotient of the étale cohomology of some smooth proper variety. (By resolution of singularities, it is equivalent to arise inside the étale cohomology of any variety.) The first constraint is obvious to impose. It is possible (although complicated) to write down representations not satisfying (1), but it takes some effort. What is

amazing about this conjecture is that the local condition at p has such a strong global implication. It is probably worth noting that the local-at-p version of this conjecture is false. Namely, suppose that V is a p-adic representation of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ which is de Rham. Then it need not arise from the cohomology of an algebraic variety over $\mathbf{Q}_p$. Consider, for example, a one dimensional unramified representation on which Frobenius acts by $\alpha \in \mathbf{Q}_p^\times$. This can only arise from a motive if α is a root of unity. More generally, if $V \subset H^n(X/\overline{\mathbf{Q}}_p, \mathbf{Q}_p)$, then the crystalline eigenvalues on $D_{\mathrm{pst}}(V)$ will be Weil numbers.

1.9.2. *Caveats.* It is tempting to extend the Fontaine–Mazur conjecture to the context of mixed motives. For example, let $E/\mathbf{Q}$ be an elliptic curve, let ∞ denote the origin of E, let $Z = \infty \cup P$ where $P - \infty$ has infinite order on $E(\mathbf{Q})$, and let $U = E \backslash Z$. Then the étale cohomology group $H^1_{\mathrm{et}}(U, \mathbf{Q}_p(1))$ is a non-split extension of $\mathbf{Q}_p$ by the Tate module $H^1_{\mathrm{et}}(E, \mathbf{Q}_p(1))$ of E. This extension is de Rham. Hence it gives rise to a class in the $(\mathbf{Q}_p)$ Selmer group of E. But now (conversely) suppose that one has a non-trivial class in the $\mathbf{Q}_p$-Selmer group, Then one hopes (wishes!) to be able to recover a point P of infinite order. This seems exactly like something that a generalized Fontaine–Mazur conjecture (with the Tate conjecture) might be able to do. But there are at least a few pitfalls to the most naïve generalization, coming from the fact that extension groups of $\mathbf{Q}_p$ representations give vector spaces over $\mathbf{Q}_p$ (which are uncountable whenever they are non-zero) and yet algebraic varieties are countable. So, for example, whenever a Selmer group has $\mathbf{Q}_p$-rank at least two, there are uncountably many lines in $\mathbf{Q}_p^2$, and thus not every extension can be a $\mathbf{Q}_p$-multiple of some geometric class.

1.10. Compatible Systems. One initially surprising implication of the Fontaine–Mazur conjecture is that, given one global p-adic Galois representation which is de Rham, it should come from a motive, and hence come from a compatible system, and thus give rise to l-adic representations for $l \neq p$ for which the trace of general Frobenius elements is the same, even though the Galois representations are quite different. (For example, the extension $\mathbf{Z}_p(1)$ coming from the p-power roots of unity comes from the motive $\mathbf{Z}(1)$ and then generates the extensions coming from l-power roots of unity for all l.) Without having to think about motives, we should at least, however, think about the notion of a compatible system of Galois representations which should arise from a motive.

We now recall some definitions from [**BLGGT14**, §5] and [**PT15**, §1]. Let F denote a number field. By a *rank n weakly compatible system of l-adic representations $\mathcal{R}$ of G_F defined over M* we mean a 5-tuple

$$(M, S, \{Q_v(X)\}, \{\rho_\lambda\}, \{H_\tau\})$$

where

    (1) M is a number field;
    (2) S is a finite set of primes of F;

(3) for each prime $v \notin S$ of F, $Q_v(X)$ is a monic degree n polynomial in $M[X]$;

(4) for each prime λ of M (with residue characteristic l, say)

$$\rho_\lambda : G_F \to \mathrm{GL}_n(\overline{M}_\lambda)$$

is a continuous, semi-simple, representation such that
- if $v \notin S$ and $v \nmid l$ is a prime of F, then ρ_λ is unramified at v and $\rho_\lambda(\mathrm{Frob}_v)$ has characteristic polynomial $Q_v(X)$,
- while if $v|l$, then $\rho_\lambda|_{G_{F_v}}$ is de Rham and in the case $v \notin S$ crystalline;

(5) for $\tau : F \hookrightarrow \overline{M}$, H_τ is a multiset of n integers such that for any $\overline{M} \hookrightarrow \overline{M}_\lambda$ over M we have $\mathrm{HT}_\tau(\rho_\lambda) = H_\tau$.

Let is consider some examples of compatible systems with $F = \mathbf{Q}$. The first example of a compatible system arises from a Galois representation with finite image:

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\mathbf{C}).$$

Up to conjugation, the image lands in $\mathrm{GL}_n(M)$ for some finite cyclotomic field M. The corresponding Galois representations ρ_λ are then the composite of the map to $\mathrm{GL}_n(M)$ with the localization $M \to M_\lambda$. The second example of a compatible system comes from (the duals of the) Tate module of an elliptic curve $E/\mathbf{Q}$. The third example comes from $H_{\mathrm{et}}^n(X, \mathbf{Q}_p)$ and varying p for some smooth proper variety $X/\mathbf{Q}$. The last two examples have coefficients in $M = \mathbf{Q}$. For an example with coefficients, let $F = K$ be an imaginary quadratic field of class number one (say $\mathbf{Q}(\sqrt{-1})$), and let E/K be a CM elliptic curve. The p-adic Galois representations:

$$\rho_{E,p} : \mathrm{Gal}(\overline{\mathbf{Q}}/K) \to \mathrm{GL}_2(\mathbf{Q}_p)$$

have image in the subgroup $(K \otimes \mathbf{Q}_p)^\times$. When p splits in K, the image is conjugate to a subgroup of a split torus. When p is inert, however, the image is conjugate to a subgroup of a non-split torus. In other words, the representations $\rho_{E,p}$ in this case are irreducible but not absolutely irreducible. On the other hand, if we take $M = K$, then the representations:

$$\rho_{E,K,\lambda} : \mathrm{Gal}(\overline{\mathbf{Q}}/K) \to \mathrm{GL}_2(K_\lambda)$$

for each prime λ in $\mathcal{O}_K$ are reducible. We may even decompose this compatible system into the direct sum of two compatible systems over K with coefficients K, corresponding to Grossencharacters χ of $\mathbf{A}_K^\times$. (The fact that the coefficients of the compatible system are related to the ground field of the Galois group is not an accident and happens more generally, and is widely acknowledged to be annoying.)

1.11. The Brauer Group. One might naturally hope that any n-dimensional compatible system with coefficients in M arises from some rank n motive with coefficients in M. In fact, there is a slight subtlety that already arises for $M = \mathbf{Q}$ and representations with finite image. Let $F/\mathbf{Q}$ be a Galois extension with $\mathrm{Gal}(F/\mathbf{Q}) = Q$, the quaternion group of order 8. There

exists a 2-dimensional representation of Q over $\mathbf{C}$ whose traces lie in $\mathbf{Q}$. Hence this gives rise to a 2-dimensional compatible system with coefficients over $\mathbf{Q}$. But Q itself does not have a 2-dimensional irreducible representation over $\mathbf{Q}$; there is a Brauer obstruction. Instead, there exists a 4-dimensional representation $V_{\mathbf{Q}}$ of Q which is irreducible, and such that $\operatorname{End}(V_{\mathbf{Q}}) = D$, the algebra of Hamilton's quaternions.

In particular, given a pure irreducible weakly compatible system $\mathcal{R}$ with coefficients in F, one predicts the existence of a simple pure irreducible motive M with coefficients in F but now possibly of larger rank. If one extends the coefficients to a suitably large extension, then M may split into some number μ of copies of some motive (in the same way that the irreducible 4-dimensional representation $V_{\mathbf{Q}}$ of Q satisfies $V_{\mathbf{Q}} \otimes \mathbf{C} = W_{\mathbf{C}}^2$, for the irreducible complex representation $W_{\mathbf{C}}$ of Q of dimension 2.)

EXAMPLE 1.11.1 (Elliptic Curves and Fake Elliptic Curves). Let $\mathcal{R}$ be a weakly compatible system of Galois representations of G_F which is irreducible of dimension 2 with coefficients in $\mathbf{Q}$, and suppose that $H_\tau = [0, 1]$ for all τ. (One source of such $\mathcal{R}$ is elliptic curves E/F.) The Fontaine–Mazur conjecture predicts that they should come from a pure irreducible motive $M/\mathbf{Q}$. Let us assume the standard conjectures [**Kle94**]. Let $\operatorname{End}_{\mathbf{Q}}(M) \otimes \mathbf{Q} = D$. From the absolute irreducibility of the Galois representations, we deduce that M is also simple, and hence D is a division algebra. The center of D is a number field E. We claim that $E = \mathbf{Q}$. It suffices to show that for all p, the center of $D \otimes \mathbf{Q}_p$ is $\mathbf{Q}_p$. By the Tate conjecture, however, we know that $\operatorname{End}_{\mathbf{Q}}(M) \otimes \mathbf{Q}_p = D \otimes \mathbf{Q}_p$ can be determined from the endomorphisms of the étale realization of M, which is isomorphic to a direct sum of some number of copies of a representation $V/\mathbf{Q}_p$, where $V \otimes \overline{\mathbf{Q}}_p$ is isomorphic by the Brauer–Nesbitt theorem to a finite number of copies of the representation r_p. But it follows that $\operatorname{End}_{\mathbf{Q}_p}(V)$ is a division algebra with center $\mathbf{Q}_p$, and the endomorphisms of the p-adic étale realization of M is a matrix algebra over this division algebra, and thus also has center $\mathbf{Q}_p$. Hence $E = \mathbf{Q}$.

Let us fix an embedding $F \to \mathbf{C}$. The Hodge realization of M is a polarized Hodge structure of weight one, which gives a polarized torus, and thus (by Riemann) an abelian variety B over $\mathbf{C}$. If we assume the Hodge Conjecture, then (by a standard argument, see the proof of Lemma 10.3.2 of [**BCGP18**] for more details) we can prove that M comes from an abelian variety over $\mathbf{Q}$. Now taking into account that the center of D is $\mathbf{Q}$, we deduce from the Albert classification (see [**Mum08**, Thm. 2, p. 201]) that A is one of the following three types:

(1) Type I: A/F is an elliptic curve with $\operatorname{End}_{\mathbf{Q}}(A) = \mathbf{Z}$.
(2) Type II: A/F is an abelian surface with $\operatorname{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$, an indefinite quaternion algebra over $\mathbf{Q}$.
(3) Type III: A/F is an abelian surface with $\operatorname{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$, a definite quaternion algebra over $\mathbf{Q}$.

(Note that Type IV of the Albert classification cannot occur, because the center $F = \mathbf{Q}$ of D is not a totally imaginary CM field.)

Hence we deduce that either A is an elliptic curve, or A is a an abelian surface with an action of some order in a division algebra. The latter are known as *fake elliptic curves*. There is an obstruction to the existence of such fake elliptic curves.

LEMMA 1.11.2. *Suppose that $F/\mathbf{Q}$ has at least one real place. Then there do not exist fake elliptic curves A/F.*

PROOF. Let $\sigma : F \hookrightarrow \mathbf{R}$ be a real place. By transport of structure, there is a non-trivial action of D on $H^1(A(\mathbf{R}), \mathbf{Q}) = \mathbf{Q}^2$. However, D does not admit any non-trivial representations of dimension less than 4 over $\mathbf{Q}$. □

It turns out that if $\mathrm{End}^0(A_{\mathbf{C}}) := \mathrm{End}(A_{\mathbf{C}}) \otimes \mathbf{Q} = D$, then D must be indefinite. (The corresponding Shimura curves X_D parametrizing such objects lie in the exceptional class of Shimura varieties with the property that there is a strict containment $D \subsetneq \mathrm{End}_{\mathbf{C}}(A)$ for all complex points A of X_D (see [**BL04**, §9.9]).) However, there can exist an inclusion $D \hookrightarrow \mathrm{End}^0(A_{\mathbf{C}})$ for definite D. For example, let $A = E \oplus E$ where E has CM by an order in K. Then $\mathrm{End}^0(A_{\mathbf{C}}) = M_2(K)$, and $M_2(K)$ certainly contains many definite $D/\mathbf{Q}$ as subalgebras.

The existence of fake elliptic curves means that one has to be careful making conjectures about the existence of Motives over $\mathbf{Q}$ simply from the coefficients of the compatible system. It is true that a classical cuspidal modular eigenform of weight 2 with coefficients in $\mathbf{Z}$ is associated to an elliptic curve rather than a fake elliptic curve, since Lemma 1.11.2 implies there are no fake elliptic curves over $\mathbf{Q}$. However, there is no analogue of Lemma 1.11.2 for abelian surfaces — we explain in [**BCGP18**] why there *do* exist cuspidal Siegel modular eigenforms of parallel weight $(2, 2)$ whose transfer to $\mathrm{GL}(4)/\mathbf{Q}$ is cuspidal and yet correspond to fake abelian surfaces over $\mathbf{Q}$ rather than abelian surfaces of $\mathbf{Q}$. This necessitated a minor modification of the main conjecture of [**BK14**].

1.12. Arithmetic L-functions. The Riemann zeta function is given for $\Re s > 1$ by the formula

$$\zeta(s) = \sum \frac{1}{n^s}.$$

However, an alternative formulation due to Euler is that

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

where the product is over primes p. This second definition can be adapted to a very general arithmetic context. Let X be a smooth, projective variety of dimension m over a number field F with good reduction outside a finite

set of primes S. Associated to X, one may write down a global Hasse–Weil zeta function:

$$\zeta_S(s) = \prod \frac{1}{1 - N(x)^{-s}},$$

where the product runs over all the closed points x of some (any) smooth integral model $\mathcal{X}/\mathcal{O}_F[1/S]$ for X. One can extend this to a product over all primes by looking at the étale cohomology for any prime l, although it is not even known that the corresponding function is independent of that choice of l (although it is true for curves and abelian varieties and finite sets.)

EXAMPLE 1.12.1. Let $f(x)$ be an irreducible polynomial with $K = \mathbf{Q}(x)/f(x)$. Let $X = \operatorname{Spec}(\mathcal{O}_K)$. Let S be the set of primes dividing Δ_K. Then, up to Euler factors dividing S, one has

$$\zeta_S(s) = \zeta_K(s).$$

One can piece together the factors $(1 - N(x)^{-s})$ for all points $x \in X(\overline{\mathbf{F}}_p)$ to write $\zeta_S(s)$ in a slightly different form as a product over primes $\mathfrak{p}$ in $\mathcal{O}_F$, namely

$$\zeta_S(s) = \prod_{\mathfrak{p} \nmid S} \zeta(X/\mathbf{F}_\mathfrak{p}, s),$$

where $\zeta(X/\mathbf{F}_\mathfrak{p}, s)$ is now the local zeta function of a smooth projective variety over a finite field, namely, if $\mathbf{F}_\mathfrak{p} = \mathbf{F}_q$, and $X/\mathbf{F}_\mathfrak{p} = Z/\mathbf{F}_q$, then

$$\zeta(Z/\mathbf{F}_q, s) = \exp\left(\sum_{m=1}^\infty \frac{\#Z(\mathbf{F}_{q^m})}{m} q^{-ms}\right).$$

A consequence of the Weil conjectures is that $\zeta_S(s)$ is absolutely convergent for $\operatorname{Re}(s) > 1 + m$. However, much more is expected:

CONJECTURE 1.12.2 (Hasse–Weil Conjecture, cf. [**Ser70**], in particular Conj. C9.). *The function $\zeta_S(s)$ extends to a meromorphic function of $\mathbf{C}$. There exists a rational number $A \in \mathbf{Q}$, rational functions $P_v(T)$ for $v|S$, and infinite Gamma factors $\Gamma_v(s)$ for $v|\infty$ such that:*

$$\xi(s) = \zeta_S(s) \cdot A^{s/2} \cdot \prod_{v|\infty} \Gamma_v(s) \cdot \prod_{v|S} \frac{1}{P_v(N(v)^{-s})}$$

is non-zero and satisfies the functional equation $\xi(s) = w \cdot \xi(m + 1 - s)$ with $w = \pm 1$.

If $F = \mathbf{Q}$ and X is a point, then $\zeta_S(s)$ is the Riemann zeta function, and Conjecture 1.12.2 follows from Riemann's functional equation [**Rie59**]. If F is a general number field but X is still a point, then $\zeta_S(s)$ is the Dedekind zeta function $\zeta_F(s)$, and Conjecture 1.12.2 is a theorem of Hecke [**Hec20**].

EXAMPLE 1.12.3. Suppose that all the cohomology of X is Tate over $\mathbf{Q}$. Then $\zeta_S(s)$ is a product of shifts of the Riemann zeta function. Suppose that all the cohomology of X is potentially Tate. Then $\zeta_S(s)$ is a product of shifts of Artin *L*-functions.

This applies, for example, whenever the cohomology of X is generated by algebraic cycles over $\overline{\mathbf{Q}}$ (as mentioned previously, this is an equivalence assuming the Tate conjecture). All Artin L-functions are meromorphic by [**Bra47**], and so Conjecture 1.12.2 is true for all such X.

1.12.4. *Hasse–Weil Conjecture for curves.* Let us now consider the Hasse–Weil conjecture for curves $X/\mathbf{Q}$ of genus g. We have the three following theorems, which I shall somewhat cheekily arrange as follows (somewhat anachronistically in the first case):

THEOREM 1.12.5 (Riemann [**Rie59**]). *Let $X/\mathbf{Q}$ be a curve of genus zero. Then the Hasse–Weil conjecture is true for X.*

THEOREM 1.12.6 (Wiles, Taylor–Wiles, Breuil–Conrad–Diamond–Taylor [**Wil95, TW95, CDT99, BCDT01**]). *Let $X/\mathbf{Q}$ be a curve of genus one. Then the Hasse–Weil conjecture is true for X.*

THEOREM 1.12.7 (Boxer–C–Gee-Pilloni [**BCGP18**]). *Let $X/\mathbf{Q}$ be a curve of genus two. Then the Hasse–Weil conjecture is true for X.*

For a curve of genus g, we have $\zeta_S(s) = \zeta_{\mathbf{Q}}(s)\zeta_{\mathbf{Q}}(s-1)/L(V,s)$, where V is the Galois representation associated to $H^1(X, \mathbf{Q}_p)$. Hence the Hasse–Weil conjecture is equivalent to the meromorphy and functional equation of $L(V,s)$. Of course, the main result of [**BCDT01**] proves the stronger result that $L(V,s)$ is *holomorphic* when X has genus one, but this is irrelevant for the Hasse–Weil conjecture. Only potentially modularity is needed to prove 1.12.2. Indeed, the potential modularity results of [**Tay03**] imply Conjecture 1.12.2 for curves X/F of genus one over any totally real field, and Theorem 1.12.7 is proved as a corollary of the following more general result which falls more naturally within the purview of the Langlands Program:

THEOREM 1.12.8. *Let X be either a genus two curve or an abelian surface over a totally real field F. Then X is potentially modular.*

1.13. Sato–Tate, and a question I learnt from Nick Katz. Let X be a smooth curve over $\mathbf{Q}$ of genus g, and let $\widetilde{a}_p = \dfrac{a_p}{\sqrt{p}}$, so $\widetilde{a}_p \in [-2g, 2g]$. The Sato–Tate predicts that the $\widetilde{a}_p$ are uniformly distributed on $[-2g, 2g]$ with respect to the Sato–Tate measure associated to X, which is the pushforward under the trace map of the Haar measure on a certain compact subgroup K of $\mathrm{USp}_{2g}(\mathbf{C})$. If X has genus one, then one knows that K is either $\mathrm{SU}(2)$, $\mathrm{O}(2)$, or $\mathrm{SO}(2)$ (the case $\mathrm{SO}(2)$ not occurring over $\mathbf{Q}$). If X has genus two, then in [**FKRS12**] it is conjecturally shown that there are 52 possible K which occur over a number field, precisely 32 of which occur over $\mathbf{Q}$. It is shown in [**Tay18**] (crucially relying on the results of [**ACC$^+$18**] in the most difficult case) that the Sato–Tate conjecture holds in all but at most one case:

THEOREM 1.13.1 ([**Tay18**]). *Let $X/\mathbf{Q}$ be a genus two curve, and suppose that $\mathrm{End}^0_{\mathbf{C}}(\mathrm{Jac}(X)) \neq \mathbf{Z}$. Then the Sato–Tate conjecture holds.*

(Of course, the missing case is the generic case.) On the other hand, one has the following question, which is a version of a question raised in these notes of Nick Katz [**Kat12**] (entitled *Simple things we don't know*). Suppose, for a curve X, I can compute information concerning the asymptotics of $\widetilde{a}_p$ as p varies. In order to define the rules of the game, say one can at least compute the quantities:

$$\widetilde{a}_{\max}(X) := \limsup \frac{\widetilde{a}_p}{\sqrt{p}} \in [-2g, 2g], \qquad \widetilde{a}_{\min}(X) = \liminf \frac{\widetilde{a}_p}{\sqrt{p}} \in [-2g, 2g].$$

QUESTION 1.13.2. Do $\widetilde{a}_{\max}(X)$ and $\widetilde{a}_{\min}(X)$ determine the genus of X? Do they determine the genus of X even if one restricts to curves of genus $g \leq 2$?

The Sato–Tate conjecture (a theorem [**Tay08, CHT08**]) implies that for a genus 1 curve $X/\mathbf{Q}$ one has $\widetilde{a}_{\max}(X) = 2$ and $\widetilde{a}_{\min}(X) = -2$. For curves of genus $g = 2$ one has, at least, the following easy consequence of Theorem 1.12.8:

THEOREM 1.13.3. *Let X be a genus two curve over $\mathbf{Q}$. Then $\widetilde{a}_{\max}(X) > 0$ and $\widetilde{a}_{\min}(X) < 0$.*

This implies Theorem 1.1.1 for $d = 5$ and $d = 6$. (The Sato–Tate conjecture for elliptic curves implies Theorem 1.1.1 and more for curves of genus $g = 1$). This theorem also gives a partial answer to question 1.13.2 — one can distinguish from $\widetilde{a}_{\max}(X)$ a curve of genus $g = 2$ from a curve of genus $g = 0$. (Again, this problem was open in this formulation before this result.) We give the idea of the argument below. But before then, we say a little more about our previous exercise:

EXERCISE 1.13.4. *Find a single polynomial $f(x)$ of degree $d = 3$ for which you can prove Theorem 1.1.1 in an elementary a way as possible.*

Here is one approach. Let $f(x) = x^3 - x$. If $p \equiv -1 \mod 4$, then $A_p = 0$. Otherwise, writing $p = a^2 + b^2$ for $p \equiv 1 \mod 4$, one has $a_p = 2a$, where the ambiguity in a and b is resolved by insisting that

$$(a + ib) \equiv 1 \mod \pi^3$$

with $\pi = (1 + i)$. One could now appeal to a theorem of Hecke [**Hec20**], which gives equidistribution of $a + bi$ in S^1 in the uniform measure. This *seems* like massive overkill, and one can almost imagine that one can address this problem using class field theory. For example, given a real quadratic field $F = \mathbf{Q}(\sqrt{d})$, one can certainly generate principal prime ideals $\mathfrak{p} = (a + b\sqrt{d})$ satisfying any congruence and (in addition) with a and b positive — the point is that such primes split completely in a certain ray class field with conductor $\mathfrak{m}_f \mathfrak{m}_\infty$ where the finite part $\mathfrak{m}_f$ of the conductor reflects any fixed choice of congruence, and the infinite part reflects the choice of local condition $\mathbf{R}^{\times 2} \subset \mathbf{R}^\times = F_v^\times$ for $v | \infty$. But there is no analogue of this in the complex case — there are no local conditions one can impose in $\mathbf{C}^\times$. Indeed,

the set of primes p with $a_p = 2a > 0$ is *not* cut out by Chebotarev conditions. Assume otherwise. Then, after passing to some finite extension $F/\mathbf{Q}$, we may assume that $\Re\{a_{\mathfrak{p}}\} > 0$ for all $\mathfrak{p}$ which split completely in $F(\sqrt{-1})$. (Alternatively, that $a_{\mathfrak{p}} < 0$ for all such $\mathfrak{p}$, but the resulting proof is exactly the same.) Now consider the Grossencharacter χ over $F(\sqrt{-1})$ associated to this elliptic curve. Now one simply applies the usual equidistribution results (as in [**Hec20**]) to deduce that the values of $b_{\mathfrak{p}} = a_{\mathfrak{p}}/N(\mathfrak{p})$ for $\mathfrak{p}$ in $F(\sqrt{-1})$ are equidistributed in $S^1 \subset \mathbf{C}^{\times}$.

I admit I do not know how to do the exercise even for this curve without using Hecke's results. I offer an Aperol spritz to anyone who can without using either Hecke's results or generalizations thereof (including all potential modularity theorems!)

For a general hyperelliptic curve X, let $A = \mathrm{Jac}(X)$. Suppose that A is potentially modular, in that over some number field F it is associated to an isobaric sum $\boxtimes \pi_i$ of cuspidal representations. By a weight argument, none of the π_i are given by a power of the norm character (and thus in particular not $|\cdot|^{1/2}$). Let $\widetilde{a}_{\mathfrak{p}} = a_{\mathfrak{p}}/N(\mathfrak{p})^{1/2} \in \mathbf{R}$ denote the normalized Hecke eigenvalues. The usual estimates of $L(\pi, s)$ and the Rankin–Selberg L-series $L(\pi_i \times \pi_j^{\vee}, s)$ (using results of Shahidi [**Sha81**]) give the estimates:

$$\frac{1}{\pi(x)} \sum_{N(\mathfrak{p}) < x} \widetilde{a}_{\mathfrak{p}} \sim 0,$$

$$\frac{1}{\pi(x)} \sum_{N(\mathfrak{p}) < x} \widetilde{a}_{\mathfrak{p}}^2 \sim \#\{(i,j) \mid \pi_i \simeq \pi_j\} = n > 0,$$

for some positive integer n. This immediately implies that $\widetilde{a}_{\mathfrak{p}}$ must often be both positive and negative. This is how one proves Theorem 1.13.3 when X has genus two, using Theorem 1.12.8. It seems hard, however, to imagine proving that A is potentially modular unless every simple factor of A (over $\mathbf{C}$, or possibly even over some totally real or CM number field) gives rise to compatible families of Galois representations of dimension at most 4. Presumably there are no such A when the genus of X is large enough. I am not sure it is known for what g there exist $A = \mathrm{Jac}(X)$ with $X/\mathbf{Q}$ which are isogenous to a product of elliptic curves. On the other hand, for curves of genus $g = 2$, one can extract more information from the fact that there is a known transfer from automorphic forms from $\mathrm{GSp}(4)$ to $\mathrm{GO}(5)$ and then $\mathrm{GL}(5)$. This gives a number of L-functions to work with of degrees 4, 5, 4×4, 4×5, and 5×5 which one can exploit using the standard Tauberian theorems to get more precise information about $\widetilde{a}_{\mathfrak{p}}$. This analysis was carried out by Noah Taylor, who proves the following optimized version of Theorem 1.13.3 (there is an analogous version over a totally real field):

THEOREM 1.13.5 ([**Tay18**]). *Let X be a curve of genus 2 over* **Q**. *There exist a positive density of primes p such that*

$$X(\mathbf{F}_p) > p + 1 + \left(\frac{2}{3} - \epsilon \right) \sqrt{p}$$

There also exist a positive density of primes p such that

$$X(\mathbf{F}_p) < p + 1 - \left(\frac{2}{3} - \epsilon \right) \sqrt{p}.$$

That is, $\widetilde{a}_{\max}(X) \geq 2/3$ and $\widetilde{a}_{\min}(X) \leq -2/3$.

Unfortunately, this is not enough to be able to distinguish (purely from these asymptotics) X from a curve of genus one, where $\widetilde{a}_{\max} = 2$ and $\widetilde{a}_{\min} = -2$. Thus, I am sad to report, there is nothing simple that I know but Nick Katz doesn't.

2. Part II: The Modified Taylor–Wiles method

2.1. The Taylor–Wiles method Part One: Formalism of Galois Deformations. The main goal of the paper [**CG18**] to explain an approach to modularity whose reach extends beyond the Taylor–Wiles method, or, perhaps more accurately, explains to how *extend* the Taylor–Wiles method to circumstances beyond its original applicability. To explain how to go beyond the Taylor–Wiles method, one should, at least in passing, say a little something about the Taylor–Wiles method itself. To this end, there are several approaches one can take. The first is to point the reader towards [**DDT97**], which is an excellent, detailed, and mostly self-contained exposition of the Taylor–Wiles method at least as it occurs in its original form. The second, which is what I shall do, is to try to explain both the Taylor–Wiles method (and its extension) to the case of one-dimensional representations. The advantage of this approach is that it exhibits many of the key features of the Wiles argument while at the same time suppressing many of the technical details which, while important, are irrelevant to the goals of this exposition. It does, of course, also obscure some of the *relevant* technical details, but we will come back to them later. I should stress that this exposition is not intended for someone who has not thought about deformation rings before, but someone who wants to see these ideas applied in a simpler setting.

2.1.1. *The standard constructions.* We begin by recalling some of the standard constructions (see [**Maz89, Maz97**]). Let p be prime, let k be a finite field of characteristic p, let F be a number field, let k finite field of characteristic p, let $\mathcal{O}$ be the ring of integers in a finite extension of $\mathbf{Q}_p$ with uniformizer ϖ and $\mathcal{O}/\varpi = k$. (For example, one could take $\mathcal{O} = W(k)$ and $\varpi = p$.) Let

$$\overline{\rho} : G_F \to \mathrm{GL}_n(k)$$

be an absolutely irreducible Galois representation (we shall also be interested in representations to $\mathrm{GSp}_4(k)$ as well, and the formalism there is very similar). Let $\mathcal{C}_{\mathcal{O}}$ denote the category of Artinian rings $(A, \mathfrak{m})$ for $\mathcal{O}$-algebras A

with a fixed isomorphism $\pi : A/\mathfrak{m} = k$. A deformation of $\bar{\rho}$ to A is a representation $\rho : G_F \to \mathrm{GL}_n(A)$ such that the following map commutes:

$$
\begin{array}{ccc}
& \mathrm{GL}_n(A) & \\
{\scriptstyle \rho} \nearrow & & \searrow {\scriptstyle \pi} \\
G_F \xrightarrow{\quad \bar{\rho} \quad} & & \mathrm{GL}_n(k).
\end{array}
$$

Say that two representations ρ are equivalent if they are conjugate by an element of $\ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(k))$. One is usually interested in the functor $\mathcal{D}$ from $\mathcal{C}_\mathcal{O}$ to sets which gives the set of all deformations satisfying some natural set of properties, for example, they are unramified outside a finite set Q, and they satisfy some p-adic Hodge theoretic condition at primes $v|p$. In many situations, these functors $\mathcal{D}$ are pro-representable by a universal deformation ring R which is a complete Noetherian $\mathcal{O}$-algebra (with a fixed isomorphism $R/\mathfrak{m}_R = k$). If we were considering deformations with no restrictions, we can deduce that the set $\mathcal{D}(k[\epsilon]/\epsilon^2)$ is in bijection with

$$
\mathrm{Ext}^1_{k[G_F]}(\bar{\rho}, \bar{\rho}) = H^1(F, \mathrm{Hom}(\bar{\rho}, \bar{\rho})) = H^1(F, \mathrm{ad}(\bar{\rho})).
$$

Alternatively, we can identify the set $\mathcal{D}(k[\epsilon]/\epsilon^2)$ with commutative diagrams as follows:

$$
\begin{array}{ccc}
\mathfrak{m}_R & \longrightarrow & \varepsilon k[\varepsilon]/\varepsilon^2 \simeq k \\
\downarrow & & \downarrow \\
R & \longrightarrow & k[\varepsilon]/\varepsilon^2 \\
\downarrow & & \downarrow \\
R/\mathfrak{m}_R & \xrightarrow{\quad = \quad} & k,
\end{array}
$$

from which we can see that

$$
\mathcal{D}(k[\epsilon]/\epsilon^2) = \mathrm{Hom}_R(\mathfrak{m}_R, k) = \mathrm{Hom}_R(\mathfrak{m}_R/(\mathfrak{m}_R^2, p), k) = \mathrm{Hom}_k(\mathfrak{m}_R/(\mathfrak{m}_R^2, p), k).
$$

For example, if we consider deformation rings in which the only condition we impose is that the representations are unramified outside some finite set Q, and let R_Q denote the corresponding deformation rings, then

$$
\dim_k(\mathfrak{m}_R/(\mathfrak{m}_R^2, p)) = \dim_k H^1_Q(F, \mathrm{ad}(\bar{\rho})),
$$

where H^1_Q denotes the subset of classes in H^1 which are unramified outside Q, equivalently, the kernel of

$$
H^1(F, \mathrm{ad}(\bar{\rho})) \to \bigoplus_{v \notin Q} H^1(I_v, \mathrm{ad}(\bar{\rho})).
$$

Note that, for a complete Noetherian local ring $(R, \mathfrak{m})$ with residue field k, the number $q = \dim_k(\mathfrak{m}_R/(\mathfrak{m}_R^2, p))$ is the minimal integer for which there exists a surjection

$$
\mathcal{O}[\![x_1, \ldots, x_q]\!] \to R.
$$

2.2. The Taylor–Wiles method Part Two: Galois Representations for $\mathrm{GL}(1)$**.** Let us now assume that F is either $\mathbf{Q}$ or is an imaginary quadratic field. These fields F are precisely the fields for which the unit group $\mathcal{O}_F^\times$ is finite. This will coincide with the fields such that the "usual" Taylor–Wiles formalism applies quite nicely. Let k be a finite field of characteristic p. Let

$$\overline{\rho} : G_F \to \mathrm{GL}_1(k) = k^\times$$

be the trivial representation. Let $\mathcal{D}$ denote the functor on $\mathcal{C}_\mathcal{O}$ such that $\mathcal{D}(A)$ is the set of deformations of $\overline{\rho}$ to A which are unramified outside Q. Here we suppose that Q is divisible only by primes $v \nmid p$. Let R_Q denote the universal deformation ring of $\overline{\rho}$. That is, R is the ring which pro-represents $\mathcal{D}$. We know, following Minkowski, that there are *no* everywhere unramified extensions of $\mathbf{Q}$. So if $F = \mathbf{Q}$ and $S = \emptyset$, we know that any ρ must be trivial. But that means there is exactly one ρ for every A, and $\mathcal{D}$ is represented by $\mathcal{O}$, so $R_\emptyset = \mathcal{O}$. Suppose we don't assume Minkowski's theorem, or, alternatively, we consider general imaginary quadratic fields F or non-trivial sets Q. Then we could let F_Q denote the maximal extension of F which is unramified outside Q. Certainly any ρ must factor through $\mathrm{Gal}(F_Q/F)$. In fact, more is true. Since $\overline{\rho}$ is trivial, the image of ρ lands in

$$1 + \mathfrak{m} \subset A^\times,$$

and it is easy to see that the left hand side is a finite abelian p-group. So, if Γ is the maximal pro-p abelian quotient of $\mathrm{Gal}(F_Q/F)$, then ρ factors through Γ. Suppose that $A = \mathcal{O}[\![\Gamma]\!]$. There is a representation

$$\rho^\Gamma : \mathrm{Gal}(F_Q/F) \to \Gamma \to \mathcal{O}[\![\Gamma]\!]^\times$$

which sends $\gamma \in \Gamma$ to $[\gamma]$. Moreover, given *any* deformation ρ to some A unramified outside Q, it's easy to see there is a map

$$\pi_A : \mathcal{O}[\![\Gamma]\!] \to A$$

sending $[\gamma]$ to $\rho(\gamma)$ such that the following diagram commutes:

$$
\begin{array}{ccc}
 & \mathcal{O}[\![\Gamma]\!]^\times & \\
{\scriptstyle \rho^\Gamma}\nearrow & & \searrow{\scriptstyle \pi_A} \\
R_Q \xrightarrow{\qquad \rho \qquad} & & A^\times.
\end{array}
$$

In other words, $R = \mathcal{O}[\![\Gamma]\!]$ and $\rho^\Gamma = \rho^{\mathrm{univ}}$. The group Γ is a pro-p group. The Frattini quotient $\Phi(\Gamma)$ is isomorphic to the Galois group of the maximal abelian extension of F unramified everywhere of exponent p. If we gave ourselves class field theory, then we could identify Γ with the p-part of the ray class group of F of conductor Q (since $v|Q$ by definition does not divide p, any extension unramified at v will be tamely ramified at v and so have conductor dividing Q.) Without class field theory, we can still use a theorem of Hermite (which says that there are only finitely many extensions of F of fixed degree and bounded discriminant) to conclude that there are only

finitely many everywhere unramified $\mathbf{Z}/p\mathbf{Z}$-extensions of F, and deduce that the Frattini quotient $\Phi(\Gamma)$ is finite. Assuming that $\Phi(\Gamma)$ has n generators, then we deduce that Γ has q topological generators for some integer q (either by using standard facts about p-groups, or, since we are in the commutative situation, using Nakayama's Lemma. It follows that Γ is a quotient of $\mathbf{Z}_p^q$, and thus R is a quotient of

$$\mathcal{O}[\![\mathbf{Z}_p^q]\!] \simeq \mathcal{O}[\![x_1, \ldots, x_q]\!].$$

Returning to the more general formalism, we see that $\mathrm{ad}(\overline{\rho}) = k$ is the trivial representation, and

$$H^1_Q(F, k) = \mathrm{Hom}(\mathrm{Gal}(F_S/F), k) = \mathrm{Hom}(\Gamma, k).$$

The dimension of $H^1_Q(F, k)$ is a fundamentally deep invariant of F which is not something that can be "computed." Of course, we can compute it in terms of the class group of F — for example, when $Q = \emptyset$, it is equal to the p-rank of the class group of F, but this is (in some sense, assuming class field theory) simply a description rather than an actual answer. The key way in which Wiles deals with this class group (a special case of a Selmer group) is to play it off against a related but quite distinct "dual" Selmer group. More generally, there is a deep relationship between the groups $H^1(F, M)$ and $H^1(F, M^\vee(1))$ for a G_F-module M. The origins of this duality go back quite far in number theory, and often (in the classical literature) go by the name "mirror theorems." They result, ultimately, from the fact that there are two fundamentally different ways to describe abelian $\mathbf{Z}/p\mathbf{Z}$-extensions of a field F, namely:

(1) Via class field theory, as coming from $\mathbf{Z}/p\mathbf{Z}$ quotients of the class group of F.
(2) Via Kummer theory, as coming from taking pth roots of elements $\alpha \in F(\zeta_p)$.

This relations are wrapped up in what (in Wiles' formulation) comes down to the Greenberg–Wiles formula, or alternatively comes down to Poitou–Tate duality. But it should really be emphasized that these complicated number theoretical statements are at their heart quite elementary and easy to explain, at least phenomenologically.

Let us suppose that we have a field K of characteristic 0 that contains a pth root of unity. Kummer theory gives a very nice characterization of degree p extensions of K as being of the form $K(\alpha^{1/p})$ for some element $\alpha \in K^\times/K^{\times p}$. Let us be a little more precise. Suppose that L/K is the maximal abelian extension of K of exponent p (it will have infinite degree). Then Kummer theory gives a canonical isomorphism:

$$K^\times/K^{\times p} \simeq \mathrm{Hom}(\mathrm{Gal}(L/K), \mu_p) = H^1(\mathrm{Gal}(L/K), \mu_p).$$

Thinking of μ_p as an $\mathbf{F}_p$-vector space, it is also convenient to write $\mu_p = \mathbf{F}_p(1)$. In fact, even if K does not contain a p-th root of unity, one still has

an isomorphism between the first and last terms. Here the map sends $\alpha \in K$ to the homomorphism (or cocycle)

$$\sigma \mapsto \frac{\sigma \alpha^{1/p}}{\alpha^{1/p}} \in \mu_p.$$

Let us now think about $\mathbf{Z}/p\mathbf{Z}$-extensions of K from the perspective of class field theory. We first consider the case when K is local and then when K is global.

When K is a local field, local class field theory provides an Artin map

$$K^\times \to \mathrm{Gal}(\overline{K}/K)^{\mathrm{ab}},$$

with dense image, and thus in particular a canonical isomorphism

$$K^\times/K^{\times p} \to \mathrm{Gal}(L/K).$$

Comparing this with the description above, we see that they are compatible in that they both imply that

$$\log_p([L:K]) = \dim K^\times/K^{\times p}.$$

On the other hand, they also can be combined to give an isomorphism

$$H^1(K, \mathbf{F}_p(1)) \simeq \mathrm{Hom}(H^1(K, \mathbf{F}_p), \mathbf{F}_p),$$

or a perfect pairing

$$H^1(K, \mathbf{F}_p(1)) \times H^1(K, \mathbf{F}_p) \to \mathbf{F}_p.$$

There is also a cup-product map between these groups to $H^2(K, \mathbf{F}_p(1))$, which can be identified with p-torsion in the Brauer group, which can be identified with $\mathbf{F}_p$, and this cup product is exactly the pairing constructed above. We can extend scalars of the pairing above to k to obtain a paring between the cohomology of k with the cohomology of $k(1)$. More generally, if M is any $k[G_K]$-module annihilated by p, then one may define

$$M^* = M^\vee(1) = \mathrm{Hom}_k(M, k(1)).$$

We then have the following:

THEOREM 2.2.1 (Tate local duality). *Let K be a local field of characteristic zero, let k be a finite field of characteristic p, and let M be a $k[G_K]$-module annihilated by p. Then there is a perfect pairing under the cup-product:*

$$H^i(K, M) \times H^{2-i}(K, M^*) \to H^2(K, k(1)) = k$$

with the following properties:

(1) *Let $G_v = \mathrm{Gal}(\overline{K}/K)$, and $I_v \subset G_v$ the decomposition group. Let*

$$H^1_{\mathrm{ur}}(K, M) = \ker H^1(K, M) \to H^1(I, M) \simeq H^1(G_v/I_v, M^{I_v})$$

denote the subspace of unramified classes. If the residue characteristic res.char(K) of K is different from p, then

$$H^1_{\mathrm{ur}}(K, M)^\perp = H^1_{\mathrm{ur}}(K, M^*).$$

(2) *The dimension of $H^1(K, M)$ is equal to*

$$\dim H^0(K, M) + \dim H^0(K, M^*) + \begin{cases} \dim(M), & p = \text{res.char}(K), \\ 0, & p \neq \text{res.char}(K). \end{cases}$$

For Global fields, things are more complicated, but yet surprisingly similar. An instructive example is the following. If ε is the cyclotomic character, then $\mathbf{F}_p(i) = \mu_p^{\otimes i}$ is the one dimensional $\mathbf{F}_p$ vector space on which $G_{\mathbf{Q}}$ acts by ε^i.

We will now us compare the following cohomology groups:

(1) $H_p^1(\mathbf{Q}, \mathbf{F}_p(i))$, extensions unramified away from p and with no condition at p.
(2) $H_{p^*}^1(\mathbf{Q}, \mathbf{F}_p(1-i))$, extensions unramified everywhere and additionally totally split at p.

Let us try to give the Kummer theory description of $H_p^1(\mathbf{Q}, \mathbf{F}_p(i))$ and compare it to the class field theory description of $H_{p^*}^1(\mathbf{Q}, \mathbf{F}_p(1-i))$. Let $G = (\mathbf{Z}/p\mathbf{Z})^\times = \text{Gal}(\mathbf{Q}(\zeta_p)/\mathbf{Q})$. There is an inflation map

$$H_p^1(\mathbf{Q}, \mathbf{F}_p(i)) \simeq H_p^1(\mathbf{Q}(\zeta_p), \mathbf{F}_p)^{\varepsilon^{-i}} = H_p^1(\mathbf{Q}(\zeta_p), \mathbf{F}_p(1))^{\varepsilon^{1-i}}.$$

Ignoring the ramification conditions, Kummer theory gives a nice description of the group $H^1(\mathbf{Q}(\zeta_p), \mathbf{F}_p(1))$ — if we let $K = \mathbf{Q}(\zeta_p)$, it is precisely $K^\times/K^{\times p}$. Let $\mathcal{O} = \mathbf{Z}[\zeta_p]$. Now let us also impose the ramification conditions. Certainly classes in

$$\mathcal{O}[1/p]^\times/\mathcal{O}[1/p]^{\times p} \subset K^\times/K^{\times p}$$

give rise to extensions unramified outside p. However, these are not necessarily the only such classes. Namely, given $K(\alpha^{1/p})/K$, it is unramified if and only if the ideal (α) has valuation divisible by p for all primes $v \nmid p$. But that implies that, as ideals in $\mathcal{O}[1/p]$, $(\alpha) = I^p$ for some ideal I with I^p principal, and hence there is a map

$$0 \to \mathcal{O}[1/p]^\times/\mathcal{O}[1/p]^{\times p} \subset H_p^1(\mathbf{Q}(\zeta_p), \mathbf{F}_p(1))) \to \text{Cl}(\mathcal{O}[1/p])[p].$$

This map is easily seen to be surjective — if $I^p = (\alpha)$, then $K(\alpha^{1/p})/K$ is unramified outside p. Note that we also have an identification $\text{Cl}(\mathcal{O}[1/p]) = \text{Cl}(\mathcal{O})$. This is because the prime above p is given by $(1 - \zeta_p)$ is principal. Since G has order prime to p, we deduce that there is also an identification

$$H_p^1(\mathbf{Q}(\zeta_p), \mathbf{F}_p(i)) \simeq \left(\mathcal{O}[1/p]^\times/\mathcal{O}[1/p]^{\times p} \right)^{\varepsilon^{1-i}} \oplus \text{Cl}(K)[p]^{\varepsilon^{1-i}}.$$

One should not be surprised by the appearance of the class group here — of course, we know that $\mathbf{Z}/p\mathbf{Z}$ quotients of the class group give extensions everywhere unramified and thus certainly unramified outside p. However, a closer look shows that the description above is *not* compatible with this description given the action of G! Rather, it relates to the description of $H_{p^*}^1(\mathbf{Q}, \mathbf{F}_p(1-i))$. The latter group is isomorphic (again via inflation) to

the classes in $H^1(K, \mathbf{F}_p)^{G=\varepsilon^{i-1}}$ which are split at p. If we took the group of classes which were merely unramified at p (denoted $H^1_\emptyset$), then we would have

$$H^1_\emptyset(\mathbf{Q}, \mathbf{F}_p(1-i)) = \operatorname{Hom}(\operatorname{Cl}(K), \mathbf{F}_p)^{\varepsilon^{i-1}} = \operatorname{Hom}((\operatorname{Cl}(K)/p\operatorname{Cl}(K))^{\varepsilon^{1-i}}, \mathbf{F}_p).$$

On the other hand, as noted above, the prime p splits principally in $\mathbf{Q}(\zeta_p)$, and hence the prime above p splits completely in any unramified extension of $\mathbf{Q}(\zeta_p)$. Hence the description above coincides with H^1_{p*}. For an abelian group A of p-power order with an action of a cyclic group G of order prime to p, we have $A[p] \simeq A/pA$ as G-modules. (Proof: possibly after a finite flat extension, decompose into G-eigenspaces.) Thus we have now related both $H^1_p(\mathbf{Q}, \mathbf{F}_p(i))$ and $H^1_{p*}(\mathbf{Q}, \mathbf{F}_p(1-i))$ to the same group. In particular, we deduce that, for $p > 2$,

(1)
$$\dim H^1_p(\mathbf{Q}, \mathbf{F}_p(i)) - \dim H^1_{p*}(\mathbf{Q}, \mathbf{F}_p(1-i)) = \dim \left(\mathcal{O}[1/p]^\times / \mathcal{O}[1/p]^{\times p}\right)^{\varepsilon^{1-i}}$$

$$= \begin{cases} 0, & i \equiv 1 \mod 2, \quad i \not\equiv 1 \mod p-1, \\ 1, & i \equiv 0 \mod 2, \quad i \not\equiv 0 \mod p-1, \\ 1, & i \equiv 1 \mod p-1, \\ 1 & i \equiv 0 \mod p-1, \end{cases}$$

$$= \quad (\dim H^1_p(\mathbf{Q}_p, \mathbf{F}_p(i)) - \dim H^0(\mathbf{Q}_p, \mathbf{F}_p(i)))$$
$$+ (\dim H^1(\mathbf{R}, \mathbf{F}_p(i)) - \dim H^0(\mathbf{R}, \mathbf{F}_p(i))$$
$$- (\dim H^0(\mathbf{Q}, \mathbf{F}_p(i)) - \dim H^0(\mathbf{Q}, \mathbf{F}_p(1-i)).$$

This very classical computation has far reaching generalizations which have several manifestations, including the Poitou–Tate exact sequence. The formulation most useful for direct applications to modularity lifting is the following:

THEOREM 2.2.2 (Greenberg–Wiles formula). *For a number field F, and a finite G_F-module M annihilated by p, let $\mathcal{L}_v \subset H^1(F_v, M)$ denote a subspace such that, for all but finitely many v, $\mathcal{L}_v$ is precisely the subgroup of unramified classes. Let $\mathcal{L}^*_v$ denote the $\mathcal{L}^\perp_v \subset H^1(F_v, M^*)$ under Tate local duality. Let $H^1_{\mathcal{L}}(F, M)$ and $H^1_{\mathcal{L}^*}(F, M)$ denote the subgroup of cohomology classes which lie in $\mathcal{L}_v$ or $\mathcal{L}^*_v$ respectively for all v. Assume that $\mathcal{L}_v = H^1_{\mathrm{ur}}(F_v, M)$ for all but finitely many v. Then*

$$\frac{\#H^1_{\mathcal{L}}(F, M)}{\#H^1_{\mathcal{L}^*}(F, M^*)} = \frac{\#H^0(F, M)}{\#H^0(F, M^*)} \prod_v \frac{\#\mathcal{L}_v}{\#H^0(F_v, M)}.$$

Although this is certainly more general than the very special example we gave above, they are, in the end, quite similar. By inflation, one can pass to the extension of F where M is trivial, and then once more the result is (at its heart) a reflection of the two descriptions of abelian extensions coming from Kummer theory and Class Field Theory.

2.3. The Taylor–Wiles method Part Three: Automorphic forms for GL(1)**.** Let us now talk about the "automorphic" side. The automorphic side is given by class field theory. But let us only assume the parts of class field theory which we need. First consider the case $F = \mathbf{Q}$. Here the "automorphic forms" we are interested in are coming from finite order characters of $\widehat{\mathbf{Z}}$. We know that any such map factors through $(\mathbf{Z}/D\mathbf{Z})^\times$ for some D, and that associated to this we obtain a corresponding representation of $\mathrm{Gal}(\mathbf{Q}(\zeta_D)/\mathbf{Q}) = (\mathbf{Z}/D\mathbf{Z})^\times$ which is unramified outside $D\infty$. For an imaginary quadratic number field F, there is a similar description, but let us write everything more adelically. For a finite set of places Q, let

$$U_Q = \prod_v U_{Q,v},$$

$$U_Q^{(\infty)} = \prod_{v \nmid \infty} U_{Q,v} \subset \mathbf{A}_F^{(\infty),\times}, \qquad U_{Q,\infty} = \prod_{v|\infty} U_v \subset (F \otimes \mathbf{R})^\times.$$

where,
(1) For finite v,
   (a) $U_{Q,v} = \mathcal{O}_{F,v}^\times$ if $v \notin Q$,
   (b) $U_{Q,v} = 1 + \pi_v \mathcal{O}_{F,v} \subset \mathcal{O}_{F,v}^\times$ if $v \in Q$.
(2) For infinite v, U_v is a compact group defined as follows:
   (a) If $v|\infty$ is real, then U_v is trivial if $v|Q$ and $U_v = \{\pm 1\} \subset \mathbf{R}^\times$ otherwise.
   (b) If $v|\infty$ is complex, then $U_v = S^1 = \{z \subset \mathbf{C} \mid |z| = 1\} \subset \mathbf{C}^\times = (F \otimes \mathbf{R})^\times$,

Associated to Q, we may then define a space

$$X_Q = F^\times \backslash \mathbf{A}_F^\times / U_Q.$$

Let's unpack exactly what the space X_Q is equal to when $F = \mathbf{Q}$ or an imaginary quadratic field F. First, X_Q is a group, and there is a surjective norm map to $\mathbf{R}^+$. Since there is a natural inclusion of $\mathbf{R}^+$ to $(F \otimes \mathbf{R})^\times = \mathbf{R}^\times$ or $\mathbf{C}^\times$, this gives a splitting of this map. Via the logarithm map there is an isomorphism $\mathbf{R}^+ \simeq \mathbf{R}$, and so we can write

$$X_Q = \mathbf{R} \times (F^\times \backslash \mathbf{A}_F^{\times,|\cdot|=1} / U_Q).$$

Let's specialize to the case of $Q = \emptyset$. There is certainly a map

$$(F^\times \backslash \mathbf{A}_F^{\times,|\cdot|=1} / U_\emptyset) \to \mathrm{Cl}(F)$$

which sends (α_v) to the class of $\prod \mathfrak{p}_v^{v(\alpha_v)}$. The kernel is represented by adeles which are units at every finite place and such that

$$\prod_{v|\infty} |\alpha_v| = 1.$$

For $F = \mathbf{Q}$ or an imaginary quadratic field, this is precisely the set $U_{\emptyset,\infty}$, and so the map above is an isomorphism. Hence $X_\emptyset$ is none other than a copy of $\mathbf{R}$ times the class group. The fields F we are considering here are

precisely the fields for which $X_\emptyset$ is $\mathbf{R}$ times a finite group (as we shall see in more detail when we consider the general case below). Let us now consider the set X_Q. The same argument as above shows that X_Q is precisely a copy of $\mathbf{R}$ times the ray class group of conductor Q. Since the copy of $\mathbf{R}$ will be irrelevant for everything that follows, let us write $X_Q = \mathbf{R} \times Y_Q$ where

$$Y_Q = (F^\times \backslash \mathbf{A}_F^{\times, |\cdot|=1}/U_Q)$$

Note that Y_Q is as a group isomorphic to the ray class group $\mathrm{RCl}(Q)$. As a set, it is a union of points. There is an action of the group $\mathrm{RCl}(Q)$ on

$$H^*(X_Q, \mathcal{O}) = H^*(Y_Q, \mathcal{O}) = H^0(Y_Q, \mathcal{O}) = \mathcal{O}^{\#\mathrm{RCl}(Q)}$$

which realizes the latter as the free rank one module $\mathcal{O}[\mathrm{RCl}(Q)]$. For each $\alpha \in \mathbf{A}_F^{(\infty),\times}$, we can define a Hecke operator $T_\alpha \in \mathrm{End}_{\mathcal{O}} H^0(Y_Q, \mathcal{O})$ to be the endomorphism induced by α. Since α acts by the image of $\alpha \in \mathrm{RCl}(Q)$, the $\mathcal{O}$-algebra $\mathbf{T}_Q$ of endomorphisms generated by these Hecke operators is precisely $\mathbf{T}_Q \simeq \mathcal{O}[\mathrm{RCl}(Q)]$.

We require the following input:

ASSUMPTION 2.3.1. There exists a surjective map

$$R_Q \to \mathbf{T}_Q = \mathcal{O}[\mathrm{RCl}(Q)].$$

The corresponding Galois representation $G_F \to \mathrm{GL}_1(\mathbf{T}_Q) = \mathbf{T}_Q^\times$ has image $\mathrm{RCl}(Q)$, and has the following properties:

(1) If $v \notin Q$, the image of Frobenius is equal to T_{ϖ_v} for any uniformizer ϖ_v of F_v.
(2) If $v \in Q$, the composite map

$$\mathbf{Q}_v^\times \to (F^\times \backslash \mathbf{A}_F^\times/U_Q) \to \mathrm{RCl}(Q)$$

is equal to the map

$$\mathbf{Q}_v^\times \to \mathrm{Gal}(\overline{F}_v/F_v)^{\mathrm{ab}} \to \mathrm{Gal}(\overline{F}/F)^{\mathrm{ab}} \to \mathrm{RCl}(Q)$$

where the first map is the Artin map from local class field theory.

Let's unpack this assumption for $F = \mathbf{Q}$. Here $\mathrm{RCl}(Q) = (\mathbf{Z}/Q\mathbf{Z})^\times/\pm 1$, and the Galois representation is factoring through $\mathrm{Gal}(\mathbf{Q}(\zeta_Q)^+/\mathbf{Q})$, where $\mathbf{Q}(\zeta_Q)^+$ is the totally real subfield of $\mathbf{Q}(\zeta)$. The corresponding Dirichlet character is coming from the canonical map $(\mathbf{Z}/Q\mathbf{Z})^\times \to (\mathbf{Z}/Q\mathbf{Z})^\times/\pm 1$. There are then two compatibilities required. The first is that the Frobenius element at the prime p in $\mathrm{Gal}(\mathbf{Q}(\zeta_Q)^+/\mathbf{Q}) = (\mathbf{Z}/Q\mathbf{Z})^\times/\pm 1$ is $[p]$. This is really easy, since Frobenius at p sends ζ to ζ^p and so is $[p]$. The second is slightly more tricky, since, at the very least, it requires knowing what the local Artin map is. What is also a little tricky is that we think of Dirichlet characters as maps

$$\widehat{\mathbf{Z}} \to (\mathbf{Z}/Q\mathbf{Z})^\times,$$

and it's not so obvious to see how $\mathbf{Q}_v^\times$ for a prime v fits into this picture. So
we have to unravel things slightly and think about the isomorphism

$$\mathbf{Q}^\times \backslash \mathbf{A_Q}^\times / U_Q \simeq \mathbf{R} \times ((\mathbf{Z}/Q\mathbf{Z})^\times / \pm 1).$$

What happens to $\mathbf{Z}_v^\times$ and $v \in \mathbf{Q}_v^\times$ for a prime $v|Q$ under this map? Let's be
careful and imagine that $v \in \mathbf{Z}$ is a positive prime number. Let's also be nice
to ourselves and imagine that Q itself is a prime power. (The general case
then follows using the Chinese remainder theorem.) Then we want to com-
pute the image of $v \in \mathbf{Q}_v^\times$. This first maps to the idele $(1, 1, \ldots, 1, v, 1, \ldots)$
with v in the slot at v and 1 everywhere else. We can then multiply by
the element of U_Q which is v at all other places. Here is where it is con-
venient to assume that Q is a power of v so this element does actually lie
in U_Q. Since this element lies in U_Q, it maps to the same element of the
RHS. Finally, we adjust by the diagonal element $v \in \mathbf{Q}^\times$ to get the idele
which is 1 at all finite places and v^{-1} at the infinite place. But the map
from the connected group $\mathbf{R}^\times / U_{Q,\infty}$ to $((\mathbf{Z}/Q\mathbf{Z})^\times / \pm 1)$ is certainly trivial,
and thus $v \in \mathbf{Q}_v^\times$ also has trivial image. If, on the other hand, we take a
general $u \in \mathbf{Z}_p^\times$, then the map comes from the inclusion $\mathbf{Z}_p^\times \to \widehat{\mathbf{Z}}^\times$, and this
is sent to $u \in ((\mathbf{Z}/Q\mathbf{Z})^\times / \pm 1)$. In the end, we deduce that the map

$$\mathbf{Q}_v^\times \to ((\mathbf{Z}/Q\mathbf{Z})^\times / \pm 1)$$

sends the positive integer v to 1 and sends a unit u to the class of $u \mod Q$.
Now we must compare this to the map coming from local class field theory.
First of all, we need to recall some facts about Lubin–Tate groups. Fortu-
nately, I once had a senior thesis student who wrote a beautifully clear expo-
sition on Lubin–Tate groups. Hence, whenever a student asks me a question
about Lubin–Tate groups, I simply direct them to [**Rie06**]. We need to show
that the Artin map from $\mathbf{Q}_v^\times$ to $(\mathbf{Z}/v^n\mathbf{Z})^\times = \mathrm{Gal}(\mathbf{Q}_v(\zeta_{v^n})/\mathbf{Q}_v)$ does indeed
send v to 1 and a unit u to $[u]$. For this, we can choose the Lubin–Tate for-
mal group to be the one associated to $f(X) = (1 + X)^v - 1 = vX \mod X^2$,
which corresponds to the formal group law $F_f(X, Y) = XY + X + Y$. In
particular, for $u \in \mathbf{Z}_p^\times$, we have the endomorphism

$$[u]X = (1 + X)^u - 1.$$

It follows that the $Q = v^n$-torsion is precisely μ_Q, and the action of $u \in \mathbf{Z}_p^\times$
is given by

$$((\zeta - 1) + 1)^u - 1 = \zeta^u - 1,$$

which is precisely $[u] \in (\mathbf{Z}/Q\mathbf{Z})^\times = \mathrm{Aut}(\mu_Q) = \mathrm{Gal}(\mathbf{Q}_v(\zeta_Q)/\mathbf{Q}_v)$. The claim
that v maps to zero precisely comes from the construction of the correspond-
ing tower, since v is the coefficient of X in $f(X)$. To give a second argument,
the claim that v maps to zero is equivalent to the claim that v is a norm
in $\mathbf{Q}_v(\zeta_{v^n})/\mathbf{Q}_v$. But we can check directly that v is the norm of $(1 - \zeta_{v^n})$ (as
long as $v \neq 2$, in which case everything is trivial).

As for the case of an imaginary quadratic field F, things are a little more complicated. One can still find explicit proofs using the theory of CM elliptic curves, although we shall not do this.

REMARK 2.3.2. The main statements of global class field theory directly imply Assumption 2.3.1. The analysis above is just to show that one can verify Assumption 2.3.1 directly from knowledge of cyclotomic fields and local class field theory.

2.4. The Taylor–Wiles method Part Four: Taylor–Wiles gluing. Now we bring some of the threads together. For a moment, let F be a general field. Let us assume that $\zeta_p \notin F$. (For $[F : \mathbf{Q}] \leq 2$, this means that $p \neq 2$, and $p \neq 3$ when $F = \mathbf{Q}(\sqrt{-3})$, but otherwise is no restriction.) If the Selmer group is $H^1_Q(F, k)$ where the condition is that all the primes outside Q are unramified and with no condition at Q. The classes in $H^1(\mathbf{Q}_p, k(1))$ dual to the unramified classes in $H^1(\mathbf{Q}_p, \mathbf{F}_p)$ are the peu-ramifée classes. Hence the dual Selmer group is $H^1_{Q^*}(F, k(1))$, where all classes are unramified outside p where they are peu-ramifée, and are also totally split at Q. From Theorem 2.2.2, we deduce the following:

LEMMA 2.4.1. *Let $F/\mathbf{Q}$ be a field which does not contain ζ_p. Let $q = \dim H^1_\emptyset(F, k)$. Then*

$$\dim H^1_{\emptyset^*}(F, k(1)) = q + (r_1 + r_2) - 1,$$

where (r_1, r_2) is the signature of F. In particular, if $F = \mathbf{Q}$ or an imaginary quadratic field, then

$$\dim H^1_\emptyset(F, k) = \dim H^1_{\emptyset^*}(F, k(1)).$$

PROOF. (cf.§8.2 of [**CG18**].) We first prove this using the Greenberg–Wiles formula (Theorem 2.2.2). All terms are trivial away from $v|p$ and $v|\infty$. Hence we have

$$\frac{\#H^1_\emptyset(F, k)}{\#H^1_{\emptyset^*}(F, k(1))} = \frac{\#H^0(F, k)}{\#H^0(F, k(1))} \prod_v \frac{\#H^1_{\mathrm{ur}}(F_v, k)}{\#H^0(F_v, k)}.$$

If $v|p$, then $H^1_{\mathrm{ur}}(F_v, k)$ is 1-dimensional. The H^0 factor which has order $\#k$ since $\zeta_p \notin F$ and hence $H^0(F, k(1)) = 0$. Hence the only remaining factors are at $v|\infty$. Since $p \neq 2$, $H^1(F_v, k) = 0$. Thus there is a $1/\#k$ factor for all $v|\infty$, and since there are exactly $r_1 + r_2$ infinite places, the result follows. $\qquad\square$

REMARK 2.4.2. For $F = \mathbf{Q}$, we proved the required case of Theorem 2.2.2 directly. However, the careful reader will note that even this direct proof used the identification of $\mathbf{Z}/p\mathbf{Z}$-quotients of the class group with unramified extensions. Thus we will (ultimately) be giving a circular argument to show that $\mathbf{Q}$ has no unramified $\mathbf{Z}/p\mathbf{Z}$-extensions for $p > 2$, since we will have assumed in our argument that such extensions come from the class group, which is trivial. One way to avoid this, and which is used in the usual proof of Kronecker–Weber for $\mathbf{Q}$, is to invoke Minkowski to show that there are

no unramified extensions of $\mathbf{Q}$ at all. We are less concerned with avoiding circularity than we are in explaining the method, so we don't do that here, and simply note that invoking Minkowski leads to an alternate proof of Lemma 2.4.1 for $F = \mathbf{Q}$. (Both groups are trivial in this case.)

2.4.3. *Minimal Modularity Theorems.* The basic idea is now to choose auxiliary sets of primes Q of size q and consider R_Q. The goal is to choose these primes so that $\mathbf{T}_Q$ gets bigger, but the tangent space of R_Q does not get bigger. To do this, we want $H_Q^1(F,k)$, which contains $H_\emptyset^1(F,k)$, to still be of dimension q, but we also want the primes $q \in Q$ to be $1 \mod p^N$ for large N.

LEMMA 2.4.4 (Annihilating the Dual Selmer Group). *Let $F = \mathbf{Q}$ or suppose that F is an imaginary quadratic field, and suppose that $\zeta_p \notin F$. Let $q = \dim H_\emptyset^1(F,k)$. Let $N \geq 1$ be any integer. Then there exist infinitely many sets of primes $Q = \{v\}$ with $N(v) \equiv 1 \mod p^N$ and $|Q| = q$ such that:*

$$\dim H_Q^1(F,k) = q.$$

PROOF. A calculation with the Greenberg–Wiles formula (as above) shows that, for an set Q of primes $\{v\}$ with $N(v) \equiv 1 \mod p^N$ and $|Q|$ of any cardinality, one has

$$\dim H_Q^1(F,k) - \dim H_{Q^*}^1(F,k(1)) = |Q|.$$

In particular, as one starts adding primes to Q, either $H_Q^1(F,k)$ goes up in dimension by one or $H_{Q^*}^1(F,k(1))$ goes down in dimension by one. So it suffices to show that, for any non-zero class $[c] \in H_{\emptyset^*}^1(F,k(1))$ which is non-trivial, there exists a prime v with $N(v) \equiv 1 \mod p^N$ such that $[c]$ is non-trivial in $H^1(F_v,k(1))$. Then, if one adds v to Q, one annihilates the class $[c]$. Note that $H_\emptyset^1(F,k(1)) \subset H^1(F,k(1))$ is isomorphic to $F^\times/F^{\times p}$ (tensor with k). Thus the problem becomes equivalent to showing that, for any $\alpha \in F^\times$ which is not a pth power, there is a prime v with $N(v) \equiv 1 \mod p$ such that α is not a pth power in $F_v^\times$. Assume otherwise. Then $\alpha \in F(\zeta_{p^n})_v^\times$ is a pth power for *every* prime which splits completely, since all such primes have norm $\equiv 1 \mod p$. But then, by Cebotarev applied to $F(\zeta_{p^n}, \alpha^{1/p})$, it follows that α is a pth power in $F(\zeta_{p^n})$. By Kummer theory, the extension $F(\alpha^{1/p}, \zeta_p)$ has degree p over $F(\zeta_p)$. But if this extension is contained in $F(\zeta_{p^n})$, it must coincide with $F(\zeta_{p^2})$, and then, up to a pth power, it follows that $\alpha = \zeta_p$. But we assumed that $\zeta_p \notin F$, so we are done.   □

With this in hand, we now consider such sets Q, and for these sets we the following diagram:

$$
\begin{array}{ccc}
\mathcal{O}[\![x_1,\ldots,x_q]\!] \longrightarrow R_Q & \longrightarrow\!\!\!\!\!\rightarrow & \mathbf{T}_Q \\
\downarrow & & \downarrow \\
R_\emptyset & \longrightarrow\!\!\!\!\!\rightarrow & \mathbf{T}_\emptyset
\end{array}
$$

All of the Galois representations involved can be restricted to $G_{\mathbf{Q}_v}$ for $v|Q$. Hence they are all deformations of the trivial representation of $G_{\mathbf{Q}_v}$, and thus deformations of $G_{\mathbf{Q}_v}^{\mathrm{ab}}$. The local deformation ring will be precisely

$$\mathcal{O}[\![\mathbf{Q}_v^\times \otimes \mathbf{Z}_p]\!]$$

If we choose Q so that v is prime to p, then there is an exact sequence

$$0 \to \mathbf{Z}_v^\times \to \mathbf{Q}_v^\times \to \mathbf{Z} \to 0,$$

where the last map is the valuation map, and the inclusion corresponds to the image of inertia at v in $G_{\mathbf{Q}_v}^{\mathrm{ab}}$. Thus, restricting to inertia, all of the rings involved are modules over

$$\mathcal{O}[\![\mathbf{Z}_v^\times \otimes \mathbf{Z}_p]\!] \simeq \mathcal{O}[\![(\mathbf{Z}/v\mathbf{Z})^\times \otimes \mathbf{Z}_p]\!].$$

Since Q contains q primes which are of the form $1 \mod p^N$, this latter ring is a quotient of

$$S_\infty = \mathcal{O}[\![t_1, \ldots, t_q]\!]$$

and it acts through a quotient which surjects onto

$$\mathcal{O}[\![t_i]\!]/((1 + t_i)^{p^N} - 1).$$

LEMMA 2.4.5. *If $\mathfrak{a}$ denotes the augmentation ideal $(t_1, \ldots, t_q)$ of S_∞, then $R_Q/\mathfrak{a} \simeq R_\emptyset$, and $\mathbf{T}_Q/\mathfrak{a} \simeq R_\emptyset$. Indeed, the ring $\mathbf{T}_Q$ is free over the quotient*

$$\bigotimes_{v|Q} \mathcal{O}[\![(\mathbf{Z}/v\mathbf{Z})^\times \otimes \mathbf{Z}_p]\!]$$

of S_∞.

PROOF. The claim for R_Q is obvious, since we are exactly imposing on our deformations that they be unramified at Q. The claim for $\mathbf{T}_Q$ follows from our identification $\mathbf{T}_Q \simeq \mathcal{O}[\![\mathrm{RCl}(Q) \otimes \mathbf{Z}_p]\!]$, and that $(\mathbf{Z}/v\mathbf{Z})^\times \otimes \mathbf{Z}_p$ is naturally a subgroup of $\mathrm{RCl}(Q)$. Note that we are crucially using here the fact that the unit group is finite and there are no units of order p, so that the exact sequence of class field theory

$$\mathcal{O}_F^\times \otimes \mathbf{Z}_p \to (\mathcal{O}_F/Q\mathcal{O}_F)^\times \otimes \mathbf{Z}_p \to \mathrm{RCl}(Q) \otimes \mathbf{Z}_p \to \mathrm{Cl}(\mathcal{O}_F) \otimes \mathbf{Z}_p \to 0$$

becomes a short exact sequence, because the first term vanishes. $\square$

Now, the magic of patching happens. Namely, one considers these diagrams as Q varies with N getting larger and larger, and then patch them formally with respect to the actions of the two power series rings ignoring everything except for the action of the two power series rings and the fact that the action of S_∞ can be lifted through the action of $\mathcal{O}[\![x_i]\!]$. The original argument is still pretty much the one that still works now. Alternatively, one can use the variation of ultrapatching as introduced by Scholze. The poor man's version of ultrapatching is just to imagine that one chooses these auxiliary primes $v \equiv 1 \mod p^\infty$. This doesn't literally make sense, but if you imagine it does make sense then you don't go wrong. However, I guess it is

incumbent on me to give at least one argument. Let me consider the special case when $q = 1$, which is no easier mathematically but significantly easier notationally. Let us choose any appropriate sequence $Q = \{q_N\}$ of primes such that $q_N \equiv 1 \mod p^N$. Let $\mathfrak{a}_N = ((1+t)^{p^N} - 1)$. Then, for each $M \geq N$, we obtain a diagram as follows:

$$\begin{array}{ccc}
\mathcal{O}/\varpi^N \mathcal{O}[\![x]\!]/\mathfrak{a}_N \longrightarrow\!\!\!\!\!\rightarrow R_{Q_M}/(\mathfrak{a}_N, \varpi^N) \longrightarrow\!\!\!\!\!\rightarrow \mathbf{T}_{Q_M}/(\mathfrak{a}_N, \varpi^N) \\
\nearrow \qquad\qquad\qquad \downarrow \qquad\qquad\qquad\qquad \downarrow \\
S_\infty/(\mathfrak{a}_N, \varpi^N) \qquad\qquad R_\emptyset/\varpi^N \longrightarrow\!\!\!\!\!\rightarrow \mathbf{T}_\emptyset/\varpi^N
\end{array}$$

Call this diagram $D_{M,N}$. For $N' \leq N$, there is is a commutative diagram $D_{M,N} \to D_{M,N'}$. Now every single object here is, after fixing N, of uniformly bounded length over the finite ring $S_\infty/(\mathfrak{a}_N, \varpi^N)$. The same applies to $D_{M,N'}$ for all $N' \leq N$, as well as all the maps $D_{M,N} \to D_{M,N'}$. Thus every possible piece of data (thinking only of these objects as S_∞ and $\mathcal{O}[\![x]\!]$-modules, together with all the corresponding morphisms between them) is finitary, and thus, by the pigeonhole principle, there exists an infinite subsequence of q_M for which the finite collection of commutative diagrams $D_{M,N} \to D_{M,N'}$ are all the same. Repeating this with larger M, we end up with patched rings R_∞ and T_∞ such that

$$R_\infty/(\mathfrak{a}_N, \varpi^N) \simeq R_{Q_M}/(\mathfrak{a}_N, \varpi^N), \qquad \mathbf{T}_\infty/(\mathfrak{a}_N, \varpi^N) \simeq \mathbf{T}_{Q_M}/(\mathfrak{a}_N, \varpi^N)$$

for all N and M. Then we end up with a diagram (dropping our notational ruse and returning to the case $q = q$):

$$\begin{array}{ccc}
\mathcal{O}[\![x_1, \ldots, x_q]\!] \longrightarrow\!\!\!\!\!\rightarrow R_\infty \longrightarrow\!\!\!\!\!\rightarrow \mathbf{T}_\infty \\
\uparrow \qquad\qquad\qquad \downarrow \qquad\qquad \downarrow \\
S_\infty \simeq \mathcal{O}[\![t_1, \ldots, t_q]\!] \qquad R_\emptyset \longrightarrow \mathbf{T}_\emptyset
\end{array}$$

Here we choose a lift of $S_\infty \to R_\infty$ to $\mathcal{O}[\![x_1, \ldots, x_q]\!]$ which is possible because S_∞ is free. Since the rings $\mathbf{T}_{Q_M}/(\mathfrak{a}_N, \varpi^N)$ were all free (of the same rank) over $S_\infty/(\mathfrak{a}_N, \varpi^N)$, we deduce that $\mathbf{T}_\infty$ is free as a module over S_∞. But that gives a regular sequence of length $q + 1$ of $\mathbf{T}_\infty$ considered as an $W[\![x_1, \ldots, x_q]\!]$-module. But that forces it to have positive rank over this module, which forces R_∞ to have positive rank, which forces them both to be free of rank one, which forces them to be isomorphic, which then, taking the quotient by the augmentation ideal $\mathfrak{a}$, shows that $R_\emptyset = \mathbf{T}_\emptyset$. Hence we prove that all unramified abelian exponent p extensions come from $\mathrm{Cl}(\mathcal{O}_F) \otimes \mathbf{Z}_p$!

2.5. The Taylor–Wiles method Part Five: the Calegari–Geraghty modification. Suppose we now try to do the same thing with a general field F. At first, thing seem even easier. We get

LEMMA 2.5.1 (Annihilating the Dual Selmer Group, Version Two). *Let $F = \mathbf{Q}$ or suppose that F is an imaginary quadratic field, and suppose that $\zeta_p \notin F$.*

Let $q = \dim H^0_\emptyset(F, k)$. Let $N \geq 1$ be any integer. Then there exist infinitely many sets of primes $Q = \{v\}$ with $N(v) \equiv 1 \mod p^N$ and $Q = q + l_0$ where $l_0 = r_1 + r_2 - 1$ such that:

$$\dim H^1_Q(F, k) = q.$$

But now, when we introduce the auxiliary sets Q, it is no longer the case that $\mathbf{T}_Q$, which should be $\mathcal{O}[\mathrm{RCl}(Q) \otimes \mathbf{Z}_p]$, is free over $\mathcal{O}[(\mathcal{O}_F/Q\mathcal{O}_F)^\times]$. This is because the presence of units mean that $\mathbf{T}_Q$ will generally be smaller than this. On the other hand, this is a good thing — if $\mathbf{T}_Q$ was free, then running the same argument with sets of size $|Q| = q + l_0$, we would get a diagram

$$
\begin{array}{ccccc}
\mathcal{O}[\![x_1, \ldots, x_q]\!] & \longrightarrow\!\!\!\!\rightarrow & R_\infty & \longrightarrow\!\!\!\!\rightarrow & \mathbf{T}_\infty \\
\uparrow & & \downarrow & & \downarrow \\
S_\infty \simeq \mathcal{O}[\![t_1, \ldots, t_q, \ldots t_{q+l_0}]\!] & & R_\emptyset & \longrightarrow\!\!\!\!\rightarrow & \mathbf{T}_\emptyset.
\end{array}
$$

And now we would get a regular sequence of $\mathbf{T}_\infty$ as an $\mathcal{O}[\![x_1, \ldots, x_q]\!]$ module of length $q + l_0 > q$, which is not possible if $\mathbf{T}_\emptyset$ is p torsion free (which it is). So the key point is to show that $\mathbf{T}_Q$ is somehow "big enough" as a module over S_∞ so that the patched ring $\mathbf{T}_\infty$ has a regular sequence of length q, which is enough to win.

The key is to look closer at the space

$$X_Q = (F^\times \backslash \mathbf{A}_F^\times / U_Q).$$

When $F = \mathbf{Q}$ or an imaginary quadratic field, this space was just a copy of $\mathbf{R}$ times the finite group $\mathrm{RCl}(Q)$. This is not the case in general. We know that it surjects onto $\mathrm{RCl}(Q)$, and this is indeed its largest finite quotient and hence the component group. But what do the fibres look like? Let's consider the fibre at the connected component. Since $\mathbf{A}_F^{(\infty),\times}$ is totally disconnected, there is a surjection from the connected component of the identity of $(F \otimes \mathbf{R})^\times / U_{Q,\infty}$, which is $(F \otimes \mathbf{R})^{Q,\times} / U_{Q,\infty}$, where

$$(F \otimes \mathbf{R})^{Q,\times} = \prod_{v | \infty} \begin{cases} \mathbf{R}^+ \subset \mathbf{R}^\times \simeq F_v^\times, & v \text{ real and } v | Q, \\ F_v^\times & \text{otherwise.} \end{cases}$$

Hence we should examine the image of the map

$$(F \otimes \mathbf{R})^{Q,\times} / U_{Q,\infty} \to X_Q = (F^\times \backslash \mathbf{A}_F^\times / U_Q).$$

Anything in $\mathcal{O}_F^\times \cap U_Q$ maps to zero in X_Q. A diagram chase identifies the image with

$$(\mathcal{O}_F^\times \cap U_Q) \backslash (F \otimes \mathbf{R})^{Q,\times} / U_{Q,\infty}.$$

As in the usual proof of Dirichlet's theorem, we may identify this space via the logarithm of absolute values map precisely with a copy of $\mathbf{R}$ times $(S^1)^{l_0} = (S^1)^{r_1 + r_2 - 1}$. When $Q = \emptyset$, then $(\mathcal{O}_F^\times \cap U_\emptyset) = \mathcal{O}_F^\times$. Now when we have a map

from $X_Q \to X_\emptyset$, the induced map on the connected component of the identity is precisely the surjection

$$(\mathcal{O}_F^\times \cap U_Q)\backslash(F \otimes \mathbf{R})^{Q,\times}/U_{Q,\infty} \to \mathcal{O}_F^\times\backslash(F \otimes \mathbf{R})^\times/U_{\emptyset,\infty}.$$

We shall also need to account for the intersection $\mathcal{O}_F^\times \cap U_Q$ and $U_{Q,\infty}$ inside $(F \otimes \mathbf{R})^\times$. By Kronecker's theorem, any unit in $U_{Q,\infty}$ is precisely a root of unity. Hence the degree of the map from X_Q to $X_\emptyset$ on the connected component of the identity is exactly

$$\frac{[\mathcal{O}_F^\times : (\mathcal{O}_F^\times \cap U_Q)]}{[\mu_F : \mu_F \cap U_Q]}.$$

For example, if $F = \mathbf{Q}$ or an imaginary quadratic field, then $\mathcal{O}_F^\times = \mu_F$ and both the numerator and denominator are the same, and so the degree of the map is one — this is not surprising given the fact that the fibres are contractible and the map is a covering map. Remember that there is also an exact sequence

$$0 \to \mathcal{O}_F^\times/(\mathcal{O}_F^\times \cap U_Q) \to U_\emptyset/U_Q \to \mathrm{RCl}(Q) \to \mathrm{Cl}(\mathcal{O}_F) \to 0.$$

The map $X_Q \to X_\emptyset$ is a covering map with Galois group Δ_Q, where Δ_Q is the quotient

$$0 \to \mu_F/(\mu_F \cap U_Q) \to U_\emptyset/U_Q \to \Delta_Q \to 0,$$

and the analysis above shows that this covering breaks up as a disconnected part and a covering map of tori.

Let us examine the cohomology of X_Q and the concomitant action of Hecke operators T_α for $\alpha \in \mathbf{A}_F^{(\infty)}$. The component group is isomorphic to $\mathrm{RCl}(Q)$. Since the connected components are copies of $\mathbf{R}$ times a torus of dimension $l_0 = r_1 + r_2 - 1$, we deduce that

$$H^*(X_Q, \mathcal{O}) \simeq H^0(X_Q, \mathcal{O}) \otimes \bigwedge\nolimits^* \left(\mathbf{Z}^{r_1+r_2-1}\right).$$

The Hecke operator T_α on cohomology just comes from the action of α on X_Q. On $H^0(X_Q, \mathcal{O})$, the operator T_α acts by the class $[\alpha]$ on the component group $\mathrm{RCl}(Q)$. To determine the geometric action on higher cohomology classes, we describe the action of α on the connected components. Suppose that $[\alpha]$ is trivial in $\mathrm{RCl}(Q)$. It follows that the projection of α to $\mathbf{A}_F^{(\infty),\times}$ can be written as xu where $x \in F^\times$ and $u \in U_Q^{(\infty)}$. But then $\alpha \in X_Q$ is equivalent to the image of x^{-1} in $(F \otimes \mathbf{R})^\times$. The corresponding action of α on the connected component is thus via multiplication by x^{-1}. But the identification of this space with a torus (times $\mathbf{R}$) is via the logarithm map, and hence α acts by *translation* on the tori, and T_α acts trivially on the corresponding cohomology classes. It follows that the $\mathcal{O}$-linear endomorphisms of $H^*(X_Q, \mathcal{O})$ generated by T_α are all coming via the action on the component group, and if $\mathbf{T}_Q \subset \mathrm{End}_\mathcal{O} H^*(X_Q, \mathbf{Z})$ denotes the ring generated by these operators, then the action of $\mathbf{T}_Q$ on $H^0(X_Q, \mathcal{O})$ factors precisely through the group ring $\mathcal{O}[\mathrm{RCl}(Q)]$, and hence $\mathbf{T}_Q \simeq \mathcal{O}[\mathrm{RCl}(Q)]$, as before.

So now comes the crucial question: how does one compare $\mathbf{T}_Q$ to $\mathbf{T}_\emptyset$? There is certainly a map of geometric spaces

$$X_Q \to X_\emptyset$$

which is Galois with covering group Δ_Q. But this covering can either come from increasing the number of connected components, or from some *unwinding* of the circles. What information, if any, can we deduce about the component group of the cover? The first thing to note is that X_Q has cohomological dimension l_0. Thus we can represent the cohomology of $X_\emptyset$ by a perfect complex $M_\emptyset^\bullet$ of $\mathcal{O}$-modules of amplitude $[0, l_0]$. But pulling this back to X_Q, we may represent the cohomology of the cover by a perfect complex $M_Q^\bullet$ of $\mathcal{O}[\Delta_Q]$-modules of amplitude $[0, l_0]$, and moreover, we will have

$$M_Q^\bullet \otimes_{\mathcal{O}[\Delta_Q]}^L \mathcal{O} = M_\emptyset^\bullet.$$

Does this give any information at all? What one should imagine is that we can choose our auxiliary set Q, now of order $q + l_0$, to consist of primes $v \equiv 1$ mod p^∞ (and not divisible by any other primes! — thus we really consider not all of X_Q but the intermediate quotient from $X_Q \to X_\emptyset$ whose degree has p-power order, and replace Δ_Q with its maximal p-power order quotient). Then we end up with a complex $M_\infty^\bullet$ with the following properties:

(1) $M_\infty^\bullet$ is a perfect complex of $S_\infty = \mathcal{O}[\![t_1, \ldots, t_{q+l_0}]\!]$-modules.
(2) The action of S_∞ factors through the map

$$\mathcal{O}[\![x_1, \ldots, x_q]\!] \to R_\infty \to \mathbf{T}_\infty = \operatorname{End}_{\mathcal{O}} M_\infty^\bullet.$$

It's immediately obvious that the cohomology of $M_\infty^\bullet$ over S_∞ can't contain any free modules, because S_∞ has to act via a ring of much smaller dimension. In fact, all the cohomology terms must have codimension at least l_0. However, we have the following:

LEMMA 2.5.2 (cf. Proposition 6.2 of [**CG18**]). *Let $M_\infty^\bullet$ be a perfect complex of $S_\infty = \mathcal{O}[\![t_1, \ldots, t_{q+l_0}]\!]$ modules of amplitude $[0, l_0]$. Suppose that $H^*(M_\infty^\bullet)$ has codimension at least l_0 as an S_∞-module. Then $M_\infty^\bullet$ is a resolution of a module M_∞ of codimension exactly l_0. In particular, in the derived category, $M_\infty^\bullet \simeq M_\infty[l_0]$ is concentrated in degree l_0.*

We won't prove this here, but it is basically an easy generalization of the fact that the kernel of any map of regular local rings is either zero or has positive rank. The conclusion, however, is that we now have a single module M_∞ which has exactly the right size — co-dimension l_0 over S_∞, and so, by some standard commutative algebra, of positive rank over R_∞, which forces $R_\infty = \mathbf{T}_\infty$ and $R_\emptyset = \mathbf{T}_\emptyset$.

2.6. The Taylor–Wiles method Part Six: Abstract Formalism. Let us extract the following lessons from the previous sections. In order to prove that $R_\emptyset = \mathbf{T}_\emptyset$, what we really need is the following:

DESIDERATA 2.7. *We would like to construct a sequence of compatible complexes $M_Q^\bullet$ for auxiliary sets Q with a suitable action of $\mathbf{T}_Q$ with the following properties:*

(1) *$M_Q^\bullet$ is a perfect complex of $\mathcal{O}[\Delta_Q]$-modules of amplitude $[0, l_0]$ for Δ_Q the pro-p quotient of $U_\emptyset/U_Q$.*

(2) *$H^*(M_Q^\bullet)$ has an action of $\mathbf{T}_Q$, and there exists a corresponding Galois representation $R_Q \to \mathbf{T}_Q$ such that:*

    (a) *This representation on inertia at Q gives rise to a map $\mathcal{O}[\Delta_Q] \to R_Q$ such that, if $\mathfrak{a}_Q$ is the augmentation ideal of this ring, then $R_Q/\mathfrak{a}_Q = R_\emptyset$, and the action of $\mathcal{O}[\Delta_Q]$ on $\mathbf{T}_Q$ is compatible with the action of this ring on $M_Q^\bullet$.*

    (b) *For the deformations we are considering, there is the corresponding relation between the dimension of the Selmer group and dual Selmer group:*

$$l_0 = \dim H_\emptyset^1(F, \mathrm{ad}(\overline{\rho})) - \dim H_{\emptyset^*}^1(F, \mathrm{ad}(\overline{\rho})).$$

In particular, if the dimensions of the Selmer group and dual Selmer group are q and $q+l_0$ respectively, there should also exist sets Q of size $q+l_0$ consisting of primes $\equiv 1 \mod p^N$ for any N such that the dual tangent space of R_Q vanishes and the tangent space of R_Q has dimension q.

For representations $\overline{\rho}$ in higher dimensions, one has to be a little more careful about the local condition $\mathcal{L}_v$ at primes $v|p$, since "unramified" will not typically be the desired local condition, rather, one wants some p-adic Hodge–Theoretic condition. Usually this requires the modification of the Taylor–Wiles method due to Kisin, which is compatible with the other modifications considered here (and outside the scope of this idiosyncratic discussion). In the sequel, we will only ever consider semistable ordinary deformation problems, since they are the most amenable to a treatment which works on the level of integral deformation rings.

2.8. The invariant l_0. It is useful to know in any situation how to compute what the invariant l_0 should be. We give the answer here in a fairly general case. Consider a representation

$$\overline{\rho} : G_F \to G^\vee(k),$$

where $G^\vee$ is a split reductive group over $\mathbf{Q}$ and such that $G^\vee(\mathbf{C})$ is the dual group of another split reductive group G. Associated to this, we may attach a universal global deformation ring R^{univ}, a universal local deformation ring R^{loc}, and a local Kisin deformation ring $R^{\mathbf{v}}$ which is a closed subscheme of R^{loc} cutting out semistable representations of some fixed type and Hodge–Tate weights. The map $\mathrm{Spec}(R^{\mathrm{univ}}) \to \mathrm{Spec}(R^{\mathrm{loc}})$ should under reasonable hypotheses be a finite morphism (see, for example, [**AC14**, Theorem 1]), and for the purposes of the next few lines one can replace R^{univ} by its image and consider it as a closed subscheme of $\mathrm{Spec}(R^{\mathrm{loc}})$. The global deformation

ring R^{glob} of $\overline{\rho}$ which cuts out deformations of type $\mathbf{v}$ is simply the ring

$$R^{\mathrm{glob}} = R^{\mathrm{univ}} \otimes_{R^{\mathrm{loc}}} R^{\mathbf{v}}.$$

Note that R^{glob} is an $\mathcal{O}$-algebra and (for some choices of $\overline{\rho}$ and $\mathbf{v}$) does have global points and so often has relative dimension at least 0 over $\mathrm{Spec}(\mathcal{O})$. However, the ring R^{glob} also has a "virtual dimension over $\mathrm{Spec}(\mathcal{O})$" given by

$$-l_0 = \dim R^{\mathrm{univ}} + \dim R^{\mathbf{v}} - \dim R^{\mathrm{loc}}.$$

(Here the dimension denotes the relative dimension over $\mathcal{O}$, or the dimension of the corresponding rigid analytic space.) The fact that $-l_0$ always turns out to be ≤ 0 reflects the expectation (given the Fontaine–Mazur conjecture) that, having fixed a weight, an inertial type, and a set of primes of bad reduction, then R^{glob} should have only finitely many $\overline{\mathbf{Q}}_p$-valued points. The fact that l_0 is often quite large even when R^{glob} *does* have $\overline{\mathbf{Q}}_p$-valued points reflects the fact that the intersection between the local and global deformation rings will not (in general) be transverse. This suggests taking into account not only the intersection R^{glob} but also the derived intersections. This is precisely what is considered in the papers [**Ven17, GV18**].

The actual computation of l_0 is now an exercise in Galois cohomology, at least under the assumption (which one expects to be true) that the local and global deformation rings are complete intersections, and that the dimension of $R^{\mathbf{v}}$ depends only on the Hodge–Tate weights. (This is known in many cases if not in complete generality.) The Galois cohomology computation reduces, using Euler characteristic formulas, into local terms depending only on the places $v|p$ and $v|\infty$. This allows us to split l_0 into three terms which depend on the weight, the group G, and the conjugacy class of complex conjugation c_v at $v|\infty$ respectively. More precisely:

(1) $l_{0,p}$ will be the difference between the dimension of the local Kisin deformation ring at a regular weight with the dimension in the weight corresponding to $\mathbf{v}$,

(2) $l_{0,G}$ will be a factor that depends only on G,

(3) $l_{0,\infty}$ will be a factor that depends only on the conjugacy classes of $\overline{\rho}(c_v)$ acting on the adjoint representation, where c_v for $v|\infty$ is complex conjugation in G_{F_v} as v ranges over all real places v of F.

We have effectively already computed the invariant l_0 for $\mathrm{GL}(1)$ (it is $r_1 + r_2 - 1$), so let us assume that G is semisimple. (Alternatively, we could assume that G is reductive, and then make sure to fix determinants for all our deformation rings.) Let us assume that $p > 2$ is totally split in F, and restrict ourselves to ordinary deformations. More specifically, fix a choice $T^{\vee}$ and $B^{\vee}$ of a torus and a Borel subgroup. Then we consider deformations such that $\rho|G_{F_v}$ can be conjugated to land inside some fixed Borel subgroup $B^{\vee}(\overline{\mathbf{Q}}_p)$, and we furthermore fix the corresponding representation on inertia

$$\psi : I_v \to B^{\vee}(\overline{\mathbf{Q}}_p) \to T^{\vee}(\overline{\mathbf{Q}}_p).$$

Finally, we additionally require that the representation be semistable. This implies that ψ is a direct sum of integral powers of the cyclotomic character. We may view ψ as coming from an algebraic map from $\mathbf{G}_m$ to $T^\vee$, and hence as an element $\eta^\vee$ of the cocharacter group $X_*(G^\vee)$. We further demand that $\eta^\vee$ is dominant with respect to $B^\vee$. Let $W_{\eta^\vee} \subseteq W$ denote the stabilizer of $\eta^\vee$ under the action of the Weyl group. The group $W_{\eta^\vee}$ is the Weyl group of some Levi subgroup $L^\vee$ of $G^\vee$ with unipotent subgroup $U^\vee \subset L^\vee$. (Warning: $U^\vee$ is the unipotent subgroup of the Levi, *not* the unipotent of the parabolic subgroup $P^\vee$ with Levi $L^\vee$!) We now define the contribution to l_0 coming from $v|p$, namely,

$$l_{0,p} = \sum_{v|p} \dim U_v^\vee.$$

The number $\dim U_v^\vee$ is exactly the dimension of the local framed ordinary deformation ring at regular weight minus the dimension of the deformation ring at the weight we are considering. (The regular weights correspond precisely to $\eta^\vee$ lying in the interior of the Weyl chamber, or equivalently weights such that $W_{\eta^\vee}$ is trivial.)

We next turn to the contribution at $v|\infty$. For each $v|\infty$, either $F_v \simeq \mathbf{R}$, and there exists a complex conjugation $c_v \in \mathrm{Gal}(\overline{F}_v/F_v)$ of order two, or $F_v \simeq \mathbf{C}$. For $F_v \simeq \mathbf{R}$, the contribution at this place is

$$\dim B^\vee(\mathbf{R}) - \dim \mathrm{ad}(\overline{\rho})^{c_v=-1}.$$

Since $p > 2$, the involution c_v lifts to involution on $\mathfrak{g}^\vee$ in characteristic zero, and we may re-write this formula as

$$\dim B^\vee(\mathbf{R}) - \frac{1}{2}\left(\dim \mathfrak{g}^\vee - \mathrm{Trace}(c|\mathfrak{g}^\vee)\right).$$

If $K(\mathbf{R}) \subset G(\mathbf{R})$ is a maximal compact subgroup, then there is an inequality ([**BV13**, Prop. 6.1])

$$\mathrm{Trace}(c|\mathfrak{g}^\vee) \geq \mathrm{rank}(G(\mathbf{R})) - 2 \cdot \mathrm{rank}(K(\mathbf{R})),$$

with equality for a unique conjugacy class (which one calls "odd" involutions). Hence, writing $\dim G(\mathbf{R}) = 2 \dim B(\mathbf{R}) - \mathrm{rank}(G(\mathbf{R}))$, this factor can also be written in the form:

$$\mathrm{rank}(G(\mathbf{R})) - \mathrm{rank}(K(\mathbf{R})) + \frac{1}{2}\left(\mathrm{Trace}(c|\mathfrak{g}) - (\mathrm{rank}(G(\mathbf{R})) - 2 \cdot \mathrm{rank}(K(\mathbf{R})))\right),$$

which exhibits this factor as the sum of two non-negative quantities. The first is trivial if and only if $G(\mathbf{R})$ admits discrete series, and the second is trivial if and only if c_v is odd. Finally, when $F_v \simeq \mathbf{C}$, the quantity is

$$\dim B(\mathbf{C}) - \dim \mathrm{ad}(\overline{\rho}) = 2 \dim B(\mathbf{R}) - \dim G(\mathbf{R}) = \mathrm{rank}(G(\mathbf{R})),$$

which is always positive. Writing rank to mean the real rank, we have equalities $\mathrm{rank}(G(\mathbf{C})) = 2 \cdot \mathrm{rank}(G(\mathbf{R}))$ and $\mathrm{rank}(K(\mathbf{C})) = \mathrm{rank}(G(\mathbf{R}))$, and hence we may also write this quantity above as:

$$\mathrm{rank}(G(\mathbf{C})) - \mathrm{rank}(K(\mathbf{C})).$$

Thus we may break the remaining contribution to l_0 into two factors $l_{0,G}$ and $l_{0,\infty}$ corresponding to the group G and the collection of complex conjugations $\{c_v\}$ at real places, namely:

$$l_{0,G} := \sum_{v \mid \infty} \mathrm{rank}(G(F_v)) - \mathrm{rank}(K(F_v)),$$

$$l_{0,\infty} := \frac{1}{2} \sum_{v \text{ real}} (\mathrm{Trace}(c \mid \mathfrak{g}) - (\mathrm{rank}(G(\mathbf{R})) - 2 \cdot \mathrm{rank}(K(\mathbf{R}))).$$

For semisimple G, we work under the hypothesis that $\bar{\rho}$ has big image and so $H^0(F, \mathrm{ad}(\bar{\rho})) = 0$. Then we have

(2) $$l_0 = l_{0,p} + l_{0,G} + l_{0,\infty}.$$

2.8.1. *Examples of l_0.* The first example is really a non-example — the case of $\mathrm{GL}(1)$, for which (since there are no unipotent elements and every involution is odd on the (trivial) adjoint representation), we have $l_{0,p} = l_{0,\infty} = 0$, and $l_0 = l_{0,G}$. The formula for semisimple groups would yield

$$l_0 = l_{0,G} \overset{?}{=} \sum_{v \mid \infty} 1 = r_1 + r_2.$$

However, $\mathrm{GL}(1)$ is not semisimple. If G admits a surjection $G \to \mathrm{GL}(1)$ with semisimple kernel, then $\mathrm{ad}(\bar{\rho})$ contains a trivial summand k and, if we are allowing the determinant to vary, the formula for l_0 needs to be adjusted by the factor $-\dim H^0(F, \mathrm{ad}(\bar{\rho})) = -1$. Hence $l_0 = r_1 + r_2 - 1$ in this case.

Consider the case when $G = G^\vee = \mathrm{GL}(2)$ and $F = \mathbf{Q}$. Since $l_0 = 0$ for one-dimensional representations over $\mathbf{Q}$, we may equally consider $G = \mathrm{SL}(2)$ and $G^\vee = \mathrm{PGL}(2)$, or equivalently $\mathrm{GL}(2)$-representations with fixed determinant. The local condition we are imposing is that

$$\rho|_{I_p} \simeq \begin{pmatrix} \varepsilon^n & * \\ 0 & \varepsilon^m \end{pmatrix},$$

where ε is the cyclotomic character, and n and m are fixed integers such that $n \geq m$, and such that $*$ is arbitrary when $n > m$ and trivial if $n \leq m$. The Weyl group has order 2. When $m > n$, then η is regular, and $U^\vee \subset L^\vee$ is trivial. When $m = n$, then η is trivial, and U is the rank one unipotent subgroup of $\mathrm{GL}(2)$. After twisting, we may assume that $m = 0$. Note that ordinary semistable representations in trivial weight are then precisely unramified. For the infinite places, there are precisely two conjugacy classes of involution on the adjoint representation, given by the non-trivial (odd) involution and the trivial (even) involution, corresponding to $\det(\bar{\rho}(c)) = -1$ or $+1$ respectively, with trace -1 and 3 respectively, and so $l_{0,\infty} = 0$ or $\frac{1}{2}(3 - (-1)) = 2$. The term $l_{G,0} = \mathrm{rank}(\mathrm{SL}_2(\mathbf{R})) - \mathrm{rank}(\mathrm{SO}_2(\mathbf{R})) = 0$. Hence we have the following four possibilities:

LEMMA 2.8.2.
(1) *Suppose that $\bar{\rho}$ is odd.*

(a) *For ordinary deformations with distinct Hodge–Tate weights,*
$l_0 = l_{0,p} + l_{0,\infty} = 0 + 0.$
(b) *For unramified deformations,* $l_0 = l_{0,p} + l_{0,\infty} = 1 + 0 = 1.$
(2) *Suppose that $\overline{\rho}$ is even.*
(a) *For ordinary deformations with distinct Hodge–Tate weights,*
$l_0 = l_{0,p} + l_{0,\infty} = 0 + 2 = 2.$
(b) *For unramified deformations,* $l_0 = l_{0,p} + l_{0,\infty} = 1 + 2 = 3.$

The same calculation shows:

LEMMA 2.8.3. *For a general field F of signature (r_1, r_2) and with $G^\vee = \mathrm{PGL}(2)$, then*

$$
l_0 = \sum_{v|p} \begin{cases} 1, & \text{distinct HT weights} \\ 0, & \text{otherwise} \end{cases} + r_2 + \sum_{v|\infty} \begin{cases} 2, & c_v \text{ trivial and } v \text{ real,} \\ 0, & \text{otherwise} \end{cases}.
$$

For general G, we have $l_0 = 0$ only if $l_{0,p} = l_{0,G} = l_{0,\infty} = 0$. Hence:

LEMMA 2.8.4. *For a split semisimple group G and representation $\overline{\rho}$ for $p > 2$ splitting completely in F, we have $l_0 = 0$ if and only if:*

(1) *We consider deformations which are regular for all $v|p$.*
(2) *F is totally real, and complex conjugation is odd for all $v|\infty$.*
(3) *$G(\mathbf{R})$ admits discrete series.*

For our last example, we consider $G = \mathrm{GSp}(4)$ and F a totally real field in which p splits completely. Note that $\mathrm{GSp}_4(\mathbf{R})$ does have discrete series, and so $l_{0,G} = 0$. Suppose that

$$\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_p)$$

is the mod-p representation associated to an abelian variety A which is ordinary at all $v|p$. There are two conjugacy classes of involution on the adjoint representation; one trivial and one odd. The action of complex conjugation on $A[p]$ is not given by a scalar, as can be seen from the Galois invariance of the Weyl pairing. Thus $l_{0,\infty} = 0$ as well for such $\overline{\rho}$. Now let us consider deformations of $\overline{\rho}$ which are semistable and ordinary. Taking the standard embedding of $\mathrm{GSp}(4) \to \mathrm{GL}(4)$, ordinary abelian varieties have Galois representations of the shape:

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

If we choose the symplectic form to be given by

$$
S = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},
$$

then the cocharacter group of PGSp_4 is generated by $t \mapsto \mathrm{diag}(t, 1, 1, t^{-1})$ and $t \mapsto \mathrm{diag}(t, t, 1, 1)$, and where the Weyl group $W = \mathbf{Z}/2\mathbf{Z} \wr \mathbf{Z}/2\mathbf{Z} = D_8$ is generated by $\sigma : (a, b, v/b, v/a) \mapsto (b, a, v/a, v/b)$ and $\tau : (a, b, v/b, v/a) \mapsto (v/a, v/b, b, a)$. In particular, for our η, the stabilizer is $\langle \sigma \rangle \simeq \mathbf{Z}/2\mathbf{Z}$, and the corresponding Levi $L^\vee$ is equal to $\mathrm{GL}(2)$ realized as

$$\mathrm{GSp}(4) \cap \begin{pmatrix} * & * & 0 & 0 \\ * & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

with maximal unipotent subgroup generated by

$$\begin{pmatrix} 1 & x & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -x \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

of dimension 1. Hence:

LEMMA 2.8.5. *For semistable ordinary deformations with Hodge–Tate weights $[0, 0, 1, 1]$ of an odd representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_p)$, we have*

$$l_0 = \sum_{v|p} 1 = [F : \mathbf{Q}].$$

2.8.6. *l_0 in more general settings.* It's not too hard to guess how to compute l_0 in slightly more general settings. For example, one can replace the semistable ordinary condition with a potentially semistable Kisin deformation ring $R^{\tau, \mathbf{v}}$ of inertial type τ and Hodge–Tate weights $\mathbf{v}$, and the answer should (roughly) be the same, with the proviso that one needs to give a Hodge-theoretic definition of the cocharacter $\eta^\vee$. This will also allow one to drop the condition that $F_v = \mathbf{Q}_p$ for all $v|p$ taking into account the weights for all maps $F_v \to \overline{\mathbf{Q}}_p$. One could also start with an algebraic automorphic form π for G and try to predict l_0 for the corresponding Galois representation (when they are conjectured to exist). If π is of cohomological type, then one always has $l_{0,v} = 0$ and $l_{0,\infty} = 0$. The infinitesimal character λ of π_∞ gives a dominant character of $X^*(T)$ and thus a cocharacter $\eta^\vee$ of $X_*(T^\vee)$, but — as one always must do in this business — one must first shift λ by ρ where ρ is half the sum of the positive weights, and this forces the correct $\eta^\vee$ to always lie in the interior of the Weyl chamber and thus $W_{\eta^\vee} = W_{\lambda+\rho}$ is always trivial. The fact that $l_{0,\infty}$ should be 0 is more subtle. For $\mathrm{GL}(n)$ over totally real fields, where the existence of Galois representations is known [**Sch15b, HLTT16**], the equality $l_{0,\infty} = 0$ has been established by Caraiani and Le Hung [**CLH16**]. Conjecturally, it should be true that any regular G-motive should give rise to a Galois representation such that complex conjugation is odd, but I confess I do not know how to prove this (even assuming [**Kle94**], of course) — looks like another Aperol spritz is up for grabs!

Another case which it might be worthwhile writing down the formula for is when one replaces $G^\vee$ by ${}^L G$. The reason one should do this is that there are some very interesting examples where $l_{G,0} = 0$ coming from quasi-split groups such as unitary groups. However, since the main emphasis of these notes are in cases where $l_0 > 0$, there are enough examples coming from split groups so that I don't feel the need to compute l_0. (Not that I really "computed" l_0 anyway, I just wrote something down.)

REMARK 2.8.7 (A battle over terminology). Some authors (see [**BV13**]) use a different terminology from l_0, despite this notation being well established in this context in [**ACC$^+$18, BCGP18, CG16, CG18**] and other places, following [**CE09**], and ultimately coming from [**BW00**]. It is clear that l_0 is the superior choice, in part because of its historical pedigree. The objection to the terminology is actually its strength. The original source of l_0 in [**BW00**] is an invariant which depends only on the group G. Namely, for a semisimple group H over $\mathbf{Q}$, one has

$$l_0 = \operatorname{rank}(H(\mathbf{R})) - \operatorname{rank}(H(\mathbf{R})).$$

if $H = \operatorname{Res}_{F/\mathbf{Q}}(G)$, then this l_0 is exactly equal to the invariant $l_{G,0}$ defined above. The objection is that l_0 is not always equal to $l_{G,0}$, which is denoted by l_0 in [**BW00**]. But note that potentially semistable Galois representations $\overline{\rho} : G_F \to G^\vee(\overline{\mathbf{Q}}_p)$ are expected to come (more or less) from automorphic forms from H. Moreover (as discussed above) the Galois representations associated to automorphic forms in Betti cohomology are all (conjecturally, and sometimes provably) in regular weight and totally odd. Hence $l_0 = l_{G,0}$ for all these Galois representations, and there is no ambiguity. The fact that l_0 may be non-zero for representations when $G = \operatorname{GL}(2)$ and $F = \mathbf{Q}$ (in either the irregular or even case) even though $l_{\operatorname{GL}(2),0} = 0$ should also not cause any confusion, since this situation will only occur when considering structures other than the Betti cohomology of arithmetic groups. Long may l_0 reign!

2.9. Extras: Kisin's formalism and Taylor's Ihara Avoidance. There are a few enrichments we would like to include going forward, but only want to describe in passing. The first is the Kisin formalism, which allows (amongst other things) one to prove non-minimal modularity lifting theorems using the same basic approach as the minimal case. For $\operatorname{GL}(1)$, this amounts (for example) to allowing (from the beginning) some tame level N. Not much is lost by first considering the case when N is a single auxiliary prime away from p. The solution is to work with R_∞ as an algebra not over $\mathcal{O}$, but over a local deformation ring at N. We have already computed the local deformation ring at N, which is

$$\mathcal{O}[\![(\mathbf{Z} \oplus (\mathbf{Z}/N\mathbf{Z})^\times) \otimes \mathbf{Z}_p]\!] \simeq \mathcal{O}[\![X, Y]\!]/((1 + Y)^{p^m} - 1),$$

where m is the largest power of p dividing $N - 1$. Let's imagine we are back in the $F = \mathbf{Q}$ case. The conclusion we draw in this case (all numerology

accounted for) is that there is a map

$$\mathcal{O}[\![x_1, \ldots, x_q, Y]\!]/((1+Y)^{p^m} - 1) \to R_\infty \to \mathbf{T}_\infty,$$

and we know that $\mathbf{T}_\infty$ as an R_∞-module has the property that its support is a union of components of $\mathrm{Spec}(R_\infty)$ of dimension $\dim R_\infty$. If the latter was connected, we would be done, but it is not — however, one "only" has to show that there are indeed modular points on each component. And this one can do (in the case $F = \mathbf{Q}$, for example, by looking at extensions of the type $\mathbf{Q}(\zeta_N)$). Related to this problem is a key idea due to Taylor, which allows one to (sometimes) pass between deformation rings with local conditions which are connected to deformation rings with local conditions that have multiple components. There is no analogue of this trick for GL(1), because it is related to unipotent representations, in particular, to deformations where at some auxiliary prime q, the representations for fields of characteristic zero either take the shape:

$$\rho|I_v \simeq \begin{pmatrix} \chi & 0 \\ 0 & \chi^{-1} \end{pmatrix},$$

where $\chi : G_{F,v} \to \overline{\mathbf{Q}}_p^\times$ is a fixed character (via local class field theory) on $\mathcal{O}_{F,v}^\times$ of order p, or

$$\rho|I_v \simeq \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

(Here v has residue characteristic different from p.) These can more integrally be characterized by demanding that the image of inertial elements satisfies the characteristic polynomial $(X - \chi)(X - \chi^{-1})$ or $(X - 1)^2$. Since χ has order p, these deformation rings are the same in characteristic p, on the other hand, the latter deformation ring has two components in characteristic zero depending on whether $*$ is zero or not. This is the starting observation of the paper [**Tay08**]. It (in some sense) totally resolves the problem of multiple components for deformation rings away from p, at a slight cost. To work, it requires that there exists a character χ on $\mathcal{O}_{F,v}^\times$ of order p, and (since v has residue characteristic different from p, this is the same as asking that p divides the order of $k_v^\times$, or $N(v) - 1$. But p need not divide $N(v) - 1$ in general. Taylor's solution is to make a cyclic base change in which v is inert to ensure this divisibility. We discuss the implications for modularity lifting in cases when $l_0 > 0$ in the next section.

2.10. Base Change and Potential Automorphy. All automorphy lifting theorems are contingent on a residual modularity assumption on $\overline{\rho}$. It is typically the case that knowing a general residual modularity conjecture (Serre's conjecture and its variants) one can then deduce modularity in characteristic zero for a compatible system $\mathcal{R}$ by varying the prime p. (Serre himself explains how his conjecture implies the Shimura–Taniyama conjecture in his original paper, see [**Ser03**, Théorème 4].) One solution is to establish cases of Serre's conjecture for small p (as Wiles [**Wil95**] does

for $p = 3$ using the result of Langlands–Tunnell [**Lan80, Tun81**]). This idea seems to run into a wall very quickly. An alternative approach is to prove *potential* modularity results by using a generalization of the p-q trick. This idea is due to Taylor [**Tay03**]. Taylor's argument involves an application of a result of Moret-Bailly [**MB89**], during which one loses control over the resulting field F except locally at any chosen finite set of places (including ∞). This is one reason why, even for studying abelian varieties over **Q**, we need to prove general modularity lifting results over (say) totally real fields F in which p splits completely. But there is a second issue which arises even *assuming* residual modularity over **Q**. The Taylor Ihara's avoidance argument mentioned in the last section requires replacing **Q** by some cyclic extension $F/\mathbf{Q}$. Even though one has more control over F in this case, it still means that we cannot avoid having to work over fields other other than **Q**. When one makes a base change, the invariant l_0 (if it is already > 0) invariably increases. Although the method (to some extent) treats all cases with $l_0 > 0$ the same, there are some special extra tricks that can be applied when $l_0 = 1$ which will be important later. It might also be worth mentioning here that it is sometimes the case that certain functorialities (especially when one allows quasi-split groups) *decrease* the value of l_0. One particular case of interest consists of deformations of representations $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ which are of regular weight. For this problem, we have $l_0 = 1$. However, for any imaginary quadratic field $E/\mathbf{Q}$, there is a corresponding representation $\overline{\varrho} : G_E \to \mathrm{GL}_3(k)$ (given by $\mathrm{ad}^0(\overline{\rho})$) which extends to a representation from $G_{\mathbf{Q}} \to \mathcal{G}_3(k)$ where $\mathcal{G}$ is the group of [**CHT08**] — and in this (quasi-split) context the corresponding value of l_0 is 0. This is the peculiar phenomena underlying the arguments of [**Cal09, Cal12**].

3. Examples

In order to produce interesting examples (in cases when $l_0 > 0$) we need to find contexts in which we can construct interesting complexes $M_Q^\bullet$ of amplitude $[0, l_0]$ in situations where the corresponding Galois representations (possibly conjectural) are also of type l_0. At the moment, there are two natural sources of complexes $M_Q^\bullet$:

(1) Complexes $M_Q^\bullet$ computing the Betti cohomology of arithmetic locally symmetric spaces. In these cases, we have $l_{0,p} = l_{0,\infty} = 0$ but $l_0 = l_{0,G}$ can be non-zero.

(2) Complexes $M_Q^\bullet$ computing the coherent cohomology of Shimura varieties in irregular weight. In these cases, we have $l_{0,\infty} = l_{0,G} = 0$, but $l_0 = l_{0,p}$ can be non-zero.

Our emphasis here will be on complexes on the second kind. (For those of the first, see [**ACC$^+$18**]). The most natural first example (beyond the case $l_0 = 0$) is the case of modular forms of weight one.

3.1. Weight One Modular Forms. Let $p > 2$ be prime, and let $N \geq 5$ be prime to p. Let $\mathcal{O}$ be the ring of integers in a finite extension K of $\mathbf{Q}_p$ with uniformizer ϖ. Let $X_1(N)$ be the modular curve considered as a smooth proper curve over $\mathrm{Spec}(\mathcal{O})$, and let ω be the pushforward of the relative dualizing sheaf along the universal elliptic curve. The coherent cohomology group $H^0(X_1(N), \omega)$ may be identified with the space of modular forms of weight one. This is precisely the space of forms where one expects to see characteristic zero lifts of $\overline{\rho}$. For general m, however, one knows that the map:

$$H^0(X_1(N), \omega) \to H^0(X_1(N), \omega/\varpi^m)$$

need not be surjective. (This was first observed by Mestre for $N = 1429$ and $p = 2$, see [**Edi06**, Appendix A]), and many examples for larger p have been subsequently computed by Buzzard and Schaeffer [**Buz14, Sch15a**].) The invariant $l_0 = 1$ in this case, and there is a very natural choice of complex $M_Q^\bullet$, namely

$$M_Q^\bullet = \mathrm{R}\Gamma(X_H(Q), \omega)_{\mathfrak{m}},$$

where $X_1(Q) \to X_H(Q) \to X_0(Q)$ is the intermediate cover which is Galois over $X_0(Q)$ with Galois group U/U_Q. Let us now go though (revisiting Desiderata 2.7) and see what it is we need to prove about these complexes.

(1) $M_Q^\bullet$ *is a perfect complex of $\mathcal{O}[\Delta_Q]$-modules of amplitude $[0, l_0]$ for Δ_Q* the pro-p quotient of $U_\emptyset/U_Q$. To check the amplitude, it suffices to show that $H^2(X, \omega)_{\mathfrak{m}}$ vanishes. This follows from the stronger claim that $H^2(X, \omega)$ vanishes, which follows from the fact that X is a curve (since one can reduce further to the vanishing of the corresponding special fibre.) The $\mathcal{O}[\Delta_Q]$-structure of the complex comes from the fact that the map $X_H(Q) \to X_0(Q)$ is étale with Galois group Δ_Q.

(2) $H^*(M_Q^\bullet)$ *has an action of $\mathbf{T}_Q$, and there exists a corresponding Galois representation $R_Q \to \mathbf{T}_Q$.* The existence of Galois representations is a standard exercise in this case, as it can be reduced to higher weight by using powers of the Hasse invariant. However, it is very important (in order for the numerology to work out) that we use the deformation ring R_Q corresponding with the correct local property at p, which in this case corresponds to representations which are unramified at p. We return to this below.

(3) *This representation on inertia at Q gives rise to a map $\mathcal{O}[\Delta_Q] \to R_Q$ such that, if $\mathfrak{a}_Q$ is the augmentation ideal of this ring, then $R_Q/\mathfrak{a}_Q = R_\emptyset$, and the action of $\mathcal{O}[\Delta_Q]$ on $\mathbf{T}_Q$ is compatible with the action of this ring on $M_Q^\bullet$.* The argument for this in [**CG18**] is along the lines of corresponding arguments in higher weight. However, there is an alternative argument found in [**KT17**] (and in a different guise in [**Ven17, GV18**]) which is more robust.

(4) *For the deformations we are considering, there is the corresponding relation between the dimension of the Selmer group and dual Selmer*

group related to the invariant l_0. This is satisfied as long as we can take the local condition at p to be unramified.

The key problem is thus to show that the Hecke action on the cohomology of $M_Q^\bullet$ gives rise to a Galois representation which is unramified at p. It suffices to prove the result for the cohomology of $M_Q^\bullet \otimes \mathcal{O}/\varpi^m$, and thus for

$$H^i(X_1(Q), \omega \otimes \mathcal{O}/\varpi^m), \qquad i = 0, \ldots, l_0$$

for all m. (Here $l_0 = 1$, but we want to anticipate other contexts with more general values of l_0.) Equivalently, we may work with the direct limit of these groups, which is $H^i(X_1(Q), \omega \otimes E/\mathcal{O})$. For formal reasons, the last non-zero cohomology group of this kind ($i = l_0 = 1$) is necessarily divisible. Hence the classes in H^{l_0} all come from forms of characteristic zero. Thus it suffices to prove that:

(1) All torsion classes in $H^i(X_1(Q), \omega \otimes \mathcal{O}/\varpi^m)$ have Galois representations which are unramified at p for $i < l_0$,
(2) All classes contributing to the characteristic zero cohomology of $H^{l_0}(X_1(Q))$ also contribute in lower degrees with the same Hecke eigenvalues, and are thus also unramified at p.

For $l_0 = 1$, it means one only has to understand the Galois representations associated to classes contributing to H^0, and these tend to be much easier to understand. In particular, in the case at hand, the remaining work comes down to the following:

THEOREM 3.1.1. *Let* $M = H^1(X_1(Q), \omega/\varpi^m)$, *and let* $\mathbf{T}$ *denote the* $\mathcal{O}$-*endomorphisms of this module generated by Hecke operators* T_n *for* n *prime to* Q. *Let* $\mathfrak{m}$ *be a non-Eisenstein maximal ideal of* $\mathbf{T}$. *Then the corresponding Galois representation:*

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}})$$

is unramified at p.

(This is more or less [**CG18**, Theorem 3.11], although see [**CS17**] for a simplification of this argument which works both for $p = 2$ and — suitably interpreted — in the reducible case.) The basic idea, which goes back at least as far as [**Gro90**] after replacing $\mathbf{T}_{\mathfrak{m}}$ by $\mathbf{T}_{\mathfrak{m}}/\mathfrak{m} = k$, is as follows. Suppose that $T_p \bmod \mathfrak{m} = a_p$ and $\langle p \rangle \bmod \mathfrak{m} = \chi(p)$. (Here χ is some finite order Nebentypus character which we now fix.) Let α and β be the roots of the polynomial $X^2 - a_p X + \chi(p)$. (Extend scalars of k if necessary.) Let us concentrate on the case when $\alpha \neq \beta$, since that assumption is in effect in [**BCGP18**] (although not in [**CG18**] or in [**Wie14**]). The goal is to show that the representation ρ can be viewed in two different ways, call them ρ_α and ρ_β (these will all be the same representation). The point is that $\rho_{\tilde{r}}$ should have the following property — it is ordinary at p, and

$$\rho_{\tilde{r}}|_{D_p} \simeq \begin{pmatrix} \chi \cdot \lambda^{-1} & * \\ 0 & \lambda \end{pmatrix},$$

where χ is the fixed Nebentypus character which is unramified at p, and $\lambda :$ $G_{\mathbf{Q}_p} \to \mathbf{T}_{\mathfrak{m}}^{\times}$ is an unramified character such that $\lambda(\mathrm{Frob}_p) \equiv \dot{r} \mod \mathfrak{m}$. If there exists $\rho_{\dot{r}}$ with $\dot{r} = \alpha$ or β (and $\alpha \neq \beta$), then these two simultaneous descriptions of ρ imply that ρ is a direct sum of two unramified characters which evaluate on Frobenius to α and β modulo $\mathfrak{m}$ respectively. In particular, it is unramified. (The case when $\alpha = \beta$ requires a different argument.)

In order to construct $\rho_{\dot{r}}$, we want to realize the space $M_{\mathfrak{m}}$ of weight one forms localized at $\mathfrak{m}$ inside a space of ordinary modular forms where the operator U_p acts by $\dot{r} \mod \mathfrak{m}$. In characteristic zero this procedure is familiar — one passes from weight $\Gamma_1(Q)$ to $\Gamma_1(Q) \cap \Gamma_0(p)$, and the space of old forms generated by an eigenform f is given by f and Vf, where V acts (say) on q-expansions via $q \mapsto q^p$. There is a formal relationship between the three operators $T = T_p$, $U = U_p$, and $V = V_p$ given in weight one by $T = U + \chi(p)V$, as well as a formal identity $UV = 1$. Hence, if $Tf = a_p f$, and $U(Vf) = f$, the action of U on the space $\{f, Vf\}$ is given by

$$\begin{pmatrix} a_p & 1 \\ -\chi(p) & 0 \end{pmatrix}$$

If α and β are the eigenvalues of this matrix (the roots of $X^2 - a_p X + \chi(p)$), then $U - \alpha$ and $U - \beta$ project onto the β and α eigenspaces respectively. If $\alpha = \beta$, then there is a generalized eigenspace of dimension two (certainly the matrix for U is not a scalar). Now consider doing this argument at the level of torsion classes. Instead of passing to level $\Gamma_0(p)$, we may equally pass to higher level by multiplying by a power of the Hasse invariant, and then projecting the resulting image to the corresponding $\dot{r}$-ordinary classes for $\dot{r} = \alpha$ or β. In order for this to work, we need this projection map to be injective on the source. By Nakayama's lemma, we may reduce to classes in $M[\mathfrak{m}]$. The problem then reduces (as it does in characteristic zero) to showing that f and Vf are linearly independent. If they are not linearly independent, then (applying U) we deduce that f is a simultaneous eigenform for T, U, and V. This is immediately seen to be impossible by examining the action on q-expansions.

This argument does not directly generalize to the symplectic case (even over $\mathbf{Q}$). The paper [**CG16**] employs a rather labyrinthian argument involving q-expansions to prove an analogous result for $\mathrm{GSp}_4 / \mathbf{Q}$.

In the rank one setting of $\mathrm{GL}(2)$, the Hecke operator T can be thought of (in the correspondence world) as coming from two operators U and V coming from Frobenius and Verschiebung respectively. The operator U controls ordinarity, but the operator V has a very simple description. Viewing the level structure at p as either having good reduction (corresponding to spherical level Sph) or coming from a choice of degree p subgroup $H \subset E[p]$ of order p (corresponding to Iwahori level Iw), we have a decomposition

$$U_{\mathrm{Sph}} = U_{\mathrm{Iw}} + \chi(p)V$$

on weight one forms with Nebentypus character χ. For GSp(4), we have more level structures at p. Relevant for this immediate discussion is the chain of level structure Sph $\supset$ Kli $\supset$ Iw at p where Sph denotes trivial level structure at p, Kli denotes Klingen level structure which corresponds to a choice of subgroup $H \subset A[p]$ of order p, and Iw denotes Iwahori level structure which corresponds to a choice of filtration $H \subset L \subset A[p]$ where H has order p and L is an isotropic plane of order p^2. We then (very roughly) have a decomposition of the appropriate Hecke operator into *three* operators, corresponding (roughly) to whether the intersection of the corresponding isotropic plane $L \subset A[p]$ with the kernel of Frobenius has rank 0, 1, or 2 respectively. In [**CG16**], there is a decomposition in weight $(2, 2)$ (and trivial Nebentypus) of the form

$$(3) \qquad\qquad U_{\mathrm{Sph},1} = U_{\mathrm{Iw},1} + Z_{\mathrm{Sph},1} + pV$$

where the analogue of the operator V doesn't play a role after reduction modulo p. The subscripts 1 here on the Hecke operators reflect the fact that, since the rank of Sp(4) is two, there are two types of Hecke operator at p. The operator $U_{\mathrm{Sph},1}$ is the one corresponding to the double cosets of $\mathrm{diag}(1, 1, p, p)$. Given an oldform at Iwahori level associated to a characteristic zero eigenform of spherical level with crystalline Frobenius eigenvalues $\{\alpha, \beta, p/\beta, p/\alpha\}$, then $U_{\mathrm{Iw},1}$ has eigenvalue (up to reordering) α.) Just as in the abecedarian context of the operators T, U, and V for $\mathrm{GL}(2)$[1], one can reduce the problem to showing the lack of existence of simultaneous eigenforms for $U_{\mathrm{Iw},1}$ and $Z_{\mathrm{Sph},1}$. The paper [**CG16**] relies on what is ultimately a fairly circuitous argument to establish this fact — while effective, it is certainly not enlightening. In [**BCGP18**], however, for reasons related to p-adic deformations in the weight aspect (which we return to later) we do not work at Spherical level, but rather at Klingen level (which in this context doesn't really have an analogue for $\mathrm{GL}(2)$). The corresponding identity of formal Hecke operators then becomes:

$$(4) \qquad\qquad U_{\mathrm{Kli},1} = U_{\mathrm{Iw},1} + Z_{\mathrm{Kli},1} + pV$$

Unlike in [**CG18**] or even [**CG16**], there do appear to be formal q-expansions which are simultaneously eigenforms for $U_{\mathrm{Iw},1}$ and $Z_{\mathrm{Kli},1}$. Hence a different argument is required, which was ultimately a good thing, because the resulting argument is both more geometric and more robust, and no longer requires q-expansions.

The argument of [**BCGP18**] ultimately uses traces of maps on differentials, and is new even for modular forms of weight one (although antecedents of the argument occur in [**Joc82, Ser73, Cai14**]). It is instructive to give

---

[1]abecedarian both in the sense of being simpler but also in the sense of being alphabetical. For some reason, my coauthors of [**BCGP18**] objected to this usage in the introduction (or anywhere else) in [**BCGP18**], so I am delighted to have an opportunity to reintroduce it here. See also §3.7.

(at least) the much simpler version of this argument in the case of GL(2) (as we also do in [**BCGP18**]).

Let X_1 denote the special fibre of $X_1(Q)$. If $f \in H^0(X_1, \omega)$, we may think of Uf as a section of $H^0(X_1 \setminus SS, \omega)$ for the finite set SS of supersingular points of X_1. There is a commutative diagram

(5)
$$
\begin{array}{ccc}
H^0(X_1, \omega) & \xrightarrow{\mathrm{Ha}\cdot U} & H^0(X_1, \omega^p) \\
\downarrow & & \downarrow \\
H^0(SS, \omega) & \longrightarrow & H^0(SS, \omega^p)
\end{array}
$$

where the vertical maps are the natural restriction maps, Ha is the Hasse invariant, and the lower horizontal map is an isomorphism given as multiplication by the form denoted B in [**Edi92**, Prop 7.2].

LEMMA 3.1.3. *If $f \in H^0(X_1, \omega)$ and $Uf = \alpha f$ for $\alpha \neq 0$ and $Tf = a_p f$, then $p = 2$ and $f = \mathrm{Ha}$.*

PROOF. Suppose that f is a U-eigenform in $H^0(X_1, \omega)$ with non-zero eigenvalue. Considering the commutative diagram (5), we see that since $\mathrm{Ha} \cdot Uf$ maps to zero in $H^0(SS, \omega^p)$, the restriction of f to SS must vanish. Thus $f = \mathrm{Ha} \cdot g$ for some $g \in H^0(X_1, \omega^{2-p})$, and this cohomology group vanishes if $p > 2$, so $f = 0$ in this case. If $p = 2$, the only non-zero sections of $H^0(X_1, \mathcal{O}_X)$ are constants, and we deduce that f is a multiple of the Hasse invariant. $\square$

3.2. Even Artin Representations. Another (very) interesting case is the case of two dimensional representations $\bar{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ which are *even*, and where we look for lifts that are unramified at p and so (conjecturally) have finite image by the Fontaine–Mazur conjecture [**FM95**]. The case of even Galois representations with distinct Hodge–Tate weights is also interesting, but was mostly addressed using trickery in [**Cal09, Cal12**]. The first admission is that there is no known construction of a complex $M_Q^\bullet$. We can, however, make predictions on what properties it should have. The first remark is that $l_0 = 3$. Hence $M_Q^\bullet$ should have amplitude $[0, 3]$. Let us compare this with one attempt to construct Galois representations associated to the relevant Maass forms due to Carayol [**Car98**]. The basic idea is that Maass forms π for $\mathrm{GL}(2)/\mathbf{Q}$ for eigenvalue $1/4$ may be transferred to a form of $U(2, 1)$ so that the transferred form is now a totally degenerate limit of discrete series. This transfer can then, in turn, be related to the cohomology of a non-algebraic complex 3-dimensional Griffiths–Schmid variety $\Gamma \backslash \Omega$ (for an arithmetic group Γ and a complex symmetric domain Ω) for some vector bundle $\mathscr{E}$ in degrees H^1 and H^2. The hope is that one can find some integral structure (perhaps related to various cup product maps) in order to prove (at least) that the Hecke eigenvalues are algebraic. The most optimistic version of this would enable one to make sense of $H^*(\Gamma \backslash \Omega, \mathscr{E})$ where $\mathscr{E}$ has integral structure, and then (continuing in this vein) show that, after localizing at

some $\mathfrak{m}$, the resulting Galois representations are unramified at p. (Functoriality can certainly change l_0, but in fact we still have $l_0 = 3$ in this case considering conjugate self dual representations $\rho : G_E \to \mathrm{GL}_3(k)$ which are unramified at p). However, it turns out that it no such integral structure exists — at least in strong enough form to give rise to complexes $M_Q^\bullet$ satisfying all of Desiderata 2.7. We now explain why. Any such collection of complexes $M_Q^\bullet$ patch to give a complex $M_\infty^\bullet$. From this complex, we can recover the original cohomology groups at minimal level by taking the derived tensor product $M_\infty^\bullet \otimes^L S_\infty/\mathfrak{a}$. For formal reasons, the support of this tensor product as a $\mathbf{T}_\mathfrak{m}$-module will be the same as the support of the naïve tensor product $M_\infty^\bullet \otimes S_\infty/\mathfrak{a}$. Since $l_0 = 3$, the complex must have amplitude $[0, 3]$. Thus the existence of a Maass form (which contributes to cohomology in H^1 and H^2) would force H^0 to also be non-zero, but this would contradict the known vanishing result $H^0(\Gamma \backslash \Omega, \mathscr{E} \otimes \mathbf{C}) = 0$. In conclusion, although this argument is far from showing that Carayol's approach is doomed, it does put limits on exactly what can be true in this situation.

3.3. Weight Two Modular Forms and (higher) Hida Theory.
We now discuss the case of modular forms of weight two. In contrast to the discussion of §3.1, things seem very much easier here. There is, however, an interesting subtlety which is easy to resolve here but gets to the crux of (one of) the issues in the next section. The invariant $l_0 = 0$ in this setting — indeed, this is the original setting of Wiles (although Wiles uses the étale cohomology of modular curves rather than coherent cohomology). The issues relating to level structure at the Taylor–Wiles primes Q are incidental to the main difficulties, so, in the sequel, let us write X for the completed modular curve of some level prime to p over $\mathrm{Spec}(\mathcal{O})$. When we also want to add level structure of type $\Gamma_0(p)$ at p, we shall simply write $X_0(p)$. A natural choice (in the context of coherent cohomology) is to take $M^\bullet = \mathrm{R}\Gamma(X, \Omega_X^1)_\mathfrak{m}$. We now would like to show that this complex has amplitude $[0, l_0] = [0, 0]$, or equivalently, that $H^1(X, \Omega_X^1)_\mathfrak{m} = 0$. One easy way to do this would be to show that $H^1(X, \Omega_X^1) = 0$. It turns out, however, that $H^1(X, \Omega_X^1)$ is *not* zero. Indeed, it is a consequence of Serre duality that this cohomology group is free of rank one. Note that the behavior here is different from what happens in weights $k \geq 3$, even though (on the Galois side) the case of weight $k = 2$ and weight $k \geq 3$ are very similar. One way to resolve this issue is to compute the action of the Hecke operators on this space. By passing to characteristic zero, one can compute that the Hecke action is Eisenstein (Verdier duality, which relates $H^1(X, \Omega_X^1)$ to $H^0(X, \mathcal{O}_X)$, is compatible with the action of Hecke up to twist — see the discussion in the proof (part (3)) of [**CG18**, Theorem 3.30]). This is a problem which will come up again for $\mathrm{GSp}(4)$ (and other groups) where it will not be so easy to resolve. Namely, in order to prove that $M^\bullet$ has the correct amplitude, we would like to know that

$$H^{l_0+1}(X, \mathscr{E})_\mathfrak{m} = 0$$

vanishes for suitably non-Eisenstein $\mathfrak{m}$ and the corresponding automorphic vector bundle $\mathscr{E}$. We will also usually be in a context in which we know that the cohomology groups

$$H^{l_0+1}(X, \mathscr{E} \otimes \mathbf{C})$$

can be computed (being careful about boundary issues) in terms of automorphic forms following work of Harris and and Zucker [**Har90, HZ94, HZ01**]. For most weights of the appropriate irregularity, one expects these groups to be zero, but not always, and often not in the most interesting cases. At least one can check that the automorphic forms which contribute in these degrees give rise to Galois representations which should be reducible, and hence do not contribute to the localization of our complex at a non-Eisenstein maximal ideal $\mathfrak{m}$. (For example, it is only the trivial representation which contributes to $H^1(X, \Omega_X^1 \otimes \mathbf{C})$ for the modular curve.) Unfortunately, this is not enough to get vanishing of the cohomology integrally, in general, and this presents a genuine stumbling block. For example, this was the main obstacle in [**CG16**] for unconditional minimal modularity lifting results for $\mathrm{GSp}(4)/\mathbf{Q}$. An approach to resolving this obstacle is exactly presented in the work of Pilloni on higher Hida Theory [**Pil17**]. We begin with a somewhat anachronistic way of motivating this theory in the case under consideration.

Our problem is that we would like all the cohomology (giving rise to interesting weight two modular forms) to be concentrated in degree zero, but we are studying the cohomology of a smooth projective curve, which has the misfortune of having cohomology (*a priori*) in degrees 0 and 1. Let us suppose we restrict our attention to considering residual representations

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

of weight two which have the additional property that they are *ordinary* at p. The usual way to think about ordinary modular forms is to allow functions to have poles at supersingular points. If X_1 is the special fibre of X, then one can think of ordinary modular forms as sections of the *open* variety $X_1^{\geq 1} := X_1 \setminus SS$. Since this variety is no longer proper, the cohomology group $H^0(X_1^{\geq 1}, \omega^2)$ is now infinite dimensional. However, there does exist an operator U $(= U_{\mathrm{Iw}})$ on this space together with a Hida idempotent $e(U)$, with the property that

$$e(U)H^0(X_1^{\geq 1}, \omega^2)$$

is finite dimensional, and corresponds to the space of classical ordinary forms of level $\Gamma_0(p)$. Instead of taking the special fibre of X and removing the supersingular points, we can take the special fibre $X_{\mathrm{Iw},1}$ of $X_{\mathrm{Iw}} = X_0(p)$. Points on X_{Iw} are of the form (E, H), and we may let $X_{\mathrm{Iw},1}^I$ denote the space where the p-rank of H is ≥ 1 (equivalently in this case, $= 1$). Note that $X_{\mathrm{Iw},1}^I$ is not the entire ordinary locus of $X_{\mathrm{Iw},1}$, but rather the component of the ordinary locus where the corresponding subgroup H is also multiplicative. This description of $X_{\mathrm{Iw},1}$ inside the special fibre of $X_0(p)$ coincides with the description of $X_1^{\geq 1}$ inside the special fibre of X from the well–known

description of the special fibre of $X_{\mathrm{Iw}} = X_0(p)$ as two copies of X intersecting transversally at the supersingular points, with one copy generically having H étale and the other with H generically multiplicative. Note that all of this is taking place on the special fibre, and that ultimately we also want information about characteristic zero. Thus one also has to consider the corresponding formal scheme $\mathfrak{X}_{\mathrm{Iw}}^I$ with special fibre $X_{\mathrm{Iw},1}^I$. (This is just the formal scheme whose rigid analytic space is the usual component of the ordinary locus containing ∞.) Many of the arguments comparing complexes reduce via Nakayama's lemma to statements that are ultimately checked on the special fibre. We now have the following theorem:

THEOREM 3.3.1. *Let $e(U)$ denote the Hida projector associated to U. There is an isomorphism*

$$e(U)\mathrm{R}\Gamma(X_1(p),\Omega^1) \simeq e(U)\mathrm{R}\Gamma(\mathfrak{X}_{\mathrm{Iw}}^I,\Omega^1).$$

Let us unpack what this Theorem says in a number of steps. First, because $\mathfrak{X}_{\mathrm{Iw}}^I$ is affine, the complex on the RHS has amplitude $[0,0]$, which is exactly what we want for $M^\bullet$. On the other hand, the H^0 term is, *a priori*, infinite dimensional. Hence the equality with the LHS at the level of H^0 terms shows that the complex on the RHS does compute a finite dimensional space of classical forms. Second, because $X_1(p)$ is projective, the LHS *a priori* has amplitude $[0,1]$. Thus this theorem also says that the Hida projector $e(U)$ annihilates the higher cohomology groups. Note that implicit in this theorem is even that there is a way to *define* the Hecke operator U on higher cohomology, which is not at all transparent. None the less, once we have a theorem of this kind, it leads to a natural definition of $M_Q^\bullet$ which has all the requisite properties concerning amplitudes, and the only thing remaining to prove a modularity lifting theorem is to know that the Galois representations associated to classes in $e(U)H^1(X_1(p),\Omega^1)$ are indeed ordinary in the usual sense.

3.4. Siegel Modular Forms of parallel weight two over Q. Let us now move on to the penultimate case we shall consider, which is the case of Siegel modular forms of weight $(2,2)$ over $\mathbf{Q}$, which is exactly where one should expect to find forms associated to Abelian surfaces over $\mathbf{Q}$. By the $F = \mathbf{Q}$ case of Lemma 2.8.5, the invariant $l_0 = 1$ in this case. Let X denote a smooth toroidal compactification of the Siegel 3-fold over $\mathrm{Spec}(\mathcal{O})$. There exists a line bundle ω^2 (which is of the form $(\det \mathscr{E})^{\otimes 2}$ for a rank two bundle $\mathscr{E} = \pi_*\Omega^1_{A/Y}$ over the interior Y of X) such that Siegel modular forms of scalar weight 2 over $\mathbf{C}$ are precisely given by classes in $H^0(X,\omega^2 \otimes \mathbf{C})$. As in the case of weight two modular forms over $\mathbf{Q}$, we naturally want to take $M^\bullet = \mathrm{R}\Gamma(X,\omega^2)_{\mathfrak{m}}$ for suitable $\mathfrak{m}$, and we immediately run in to the problem that we do not know how to prove the non-vanishing of

$$H^2(X,\omega^2)_{\mathfrak{m}}.$$

As in the case of classical modular forms of weight two considered in §3.3, the group $H^2(X, \omega^2)$ will (in general) be non-zero and see contributions from certain non-tempered automorphic forms (whose associated Galois representations will be reducible). However, we currently have very little idea how to prove integral vanishing statements. If (k, l) is a pair of integers with $k \geq l \geq 2$, there exists a automorphic vector bundle whose coherent cohomology in degree zero gives the space of classical (vector valued) Siegel modular forms of weight (k, l) — it is given by $(\mathrm{Sym}^{k-l} \mathscr{E}) \otimes (\det \mathscr{E})^{\otimes l}$ over $Y \subset X$. When $k = l$, this is a line bundle and one recovers the classical space of scalar forms. The corresponding Galois representations (when irreducible) are conjecturally associated to deformation problems of type $l_0 = 0$ when $k \geq l \geq 3$ and of type $l_0 = 1$ when $k \geq l = 2$. We can moreover consider $l_0 = 1$ modularity problems for $\kappa = (k, 2)$ of this type, in which case we denote the vector bundle by ω^κ. Curiously enough, although the considerations on the Galois side are very similar in weights $(2, 2)$ and $(k, 2)$ for $k > 2$, the automorphic situation is often simpler in the later case, because at least in these settings there are no longer any non-tempered forms contributing to $H^2(X, \omega^\kappa) \otimes \mathbf{C}$. In fact, in favourable circumstances when $p - 1 > k \geq 4$, one can even prove these groups vanish integrally [**LS13**]. (This is analogous to the distinction between what happens for GL(2) in weights $k = 2$ and weights $k > 2$.) It will be useful in the sequel to consider the these spaces and the corresponding complexes $M_\kappa^\bullet$ (of various flavours).

The special fibre X_1 has a filtration according to the rank of $A[p]$, which is an integer between 0 and 2. The ordinary locus $X_1^{=2}$ is precisely the A for which $A[p]$ has rank 2, and its complement $X_1^{\leq 1}$ is the vanishing locus of the Hasse invariant. Along this locus, there is a higher Hasse invariant (these higher Hasse invariants constructed in [**Box15, GK15**] are generalizations of the B which appeared in equation 5), whose vanishing locus in this case is the supersingular locus $X_1^{=0}$, and which has codimension 2 inside X_1. We let $X_1^{\geq 1} = X \setminus X_1^{=0}$ denote the complement of the supersingular locus. A first approximation to the complexes we want, at least modulo p, are to consider $\mathrm{R}\Gamma(X_1^{\geq 1}, \omega^\kappa)$, to construct a certain Hecke operator T on this cohomology, and then to take the projection to the part where T is invertible. There is, unfortunately, a menagerie of Hecke operators and various spaces considered in [**BCGP18**], with filtrations coming from the p-rank as well as the degrees of various choices of isotropic subspaces which are related but different. There are also Hecke operators at $v | p$ at spherical level, Klingen level, and Iwahori level, some of which are the same and some of which are different. In fact, as mentioned previously, there are also two types of Hecke operators for each level. Note, however, that it is crucial in our setting to have T act on the entire complex, which raises the following problem:

PROBLEM 3.4.1. *How does one define Hecke operators on higher coherent cohomology?*

Note that even for H^0 things are not always straightforward. Already it is non-trivial to give a good integral definition of T_p for classical modular forms. This is especially true for torsion classes in weight 1. Gross [**Gro90**] uses the q-expansion principle and (implicitly) the fact that the map $X_0(p) \to X$ is a finite morphism, which is no longer true in general. The problem is that the "obvious" definition in terms of correspondences does not have the correct normalization, which may and indeed does involve dividing by some power of p, which takes some care when considering torsion classes. Conrad [**Con07**] gives an alternative construction which is still restricted to the modular curve setting. The paper [**ERX17**] gives a construction of T using a dualizing trace map. The paper [**Pil17**] introduces a new approach to constructing Hecke operators in higher cohomology using duality in coherent cohomology and fundamental classes (see §3.8.7 of [**BCGP18**]). Note that a key point in all of these cases is that the "correct" definition of T involves first defining a map coming from a correspondence and then showing that it is "divisible" by the correct power of p. In order to do this, however, the Hecke operator *must* be given (up to some normalization) as a composition of correspondences, because the arguments which prove that one can normalize these operators correctly are geometric. In particular, if Hecke operators A and B are defined separately using different correspondences, there is no obvious way for proving that (for example) their sum is divisible by p. Thus it is miraculous that, in the setting of $\mathrm{GSp}(4)$, that one does have access to such constructions for enough relevant Hecke operators, which was one of the crucial observations of [**Pil17**]. The first important Hecke operator (denoted T) is one which is closely related to double cosets of $\mathrm{diag}(1, p, p, p^2)$, although it is not literally what one would call $U_{\mathrm{Sph},2}$ (in contrast to $U_{\mathrm{Sph},1}$ occurring in equation 3). It is defined by considering the composite of two correspondences: one from $X = X_{\mathrm{Sph}}$ to X_{Par} (where Par denotes paramodular level structure at p), and the other from a correspondence between X_{Par} and X, both of these maps corresponding to $\mathrm{diag}(1, 1, 1, p)$ and $\mathrm{diag}(1, p, p, p)$.

If k is sufficiently large ($k \geq 3$), then the action of T on a classical Siegel modular form of weight $\kappa = (k, 2)$ with crystalline Frobenius eigenvalues $\{\alpha, \beta, p^{k-1}/\alpha, p^{k-1}/\beta\}$ will be (up to reordering) $\alpha\beta \mod p$. Hence, at least in these weights, the idempotent associated to T cuts out the ordinary space (at least for large enough k).

In the previous section, the spaces $X_1^{\geq 1}$ were affine and thus had cohomology only in degree zero. The space $X_1^{\geq 1}$ here is no longer affine, but it *almost* has the property that it is covered by two affines, namely, it is the complement of the zero locus of two sections. The weasel word almost is required because this is only literally true on the minimal compactification (where the higher Hasse invariants are defined). But that turns out to be just as good, since the higher derived images of the pushforward of automorphic sheaves under the map $\pi : X \to X^{\mathrm{min}}$ all vanish ([**HLTT16**, Theorem 5.4], see also [**AIP15**].) So, for all intents and purposes (and for the coherent

sheaves we consider) $X_1^{\geq 1}$ has cohomological dimension 1 (as does the related space $X_{\mathrm{Kli},1}^I$ which occurs below). The next step is to consider the map

$$(6) \qquad e(T)\mathrm{R}\Gamma(X_1^{\geq 1}, \omega^\kappa) \to e(T)\mathrm{R}\Gamma(X_1, \omega^\kappa).$$

Again we are in the situation where the LHS has amplitude $[0,1]$ and the RHS is finite. What one actually proves is a comparison between these complexes in Equation 6 for weights $(k,2)$ where k is sufficiently large. One of the main ideas of [**Pil17**] is that the amplitude $[0,1]$ complex on the LHS can be p-adically interpolated over weights $(k,2)$ as k varies. Note that in usual Hida theory, one interpolates a single amplitude $[0,0]$ complex over all weights, and the specialization of that single term to a regular weight recovers the classical space. Here the idea is to interpolate the complex of amplitude $[0,1]$ over a space of irregular weights of the form $(k,2)$, such that the specialization recovers a complex which, at least for large enough k, returns the classical ordinary complex. In order to interpolate ordinary modular forms in the usual setting, one has to work at p^∞-power level (at the Igusa tower), and similarly in this setting one has to consider spaces at (higher) Klingen level, where the infinite Klingen level structure amounts to choosing a cyclic subgroup $H \subset A[p^\infty]$. Here one can p-adically interpolate the sheaves ω^κ into a big sheaf Ω^κ. Replacing X now by X_{Kli}, where one imposes level p Klingen structure at p, there is a corresponding space $X_{\mathrm{Kli},1}^I(p^\infty)$ where the corresponding rank one p-divisible group H is multiplicative. Now the desired complex $M_\Lambda^\bullet$ is constructed as follows. First one takes the ordinary projection $e(U)\mathrm{R}\Gamma(\mathfrak{X}_{\mathrm{Kli}}^I(p^\infty), \Omega^\kappa)$, where $e(U)$ is an ordinary projector associated to the corresponding Hecke operator $U = U_{\mathrm{Kli},2}$ at Klingen level in contrast with the operator T (which was related to $U_{\mathrm{Sph},2}$) above, and Ω^κ is a big sheaf (of the flavour considered in [**AIP15**]) which interpolates weights of the form $(k,2)$ as k varies. The operator $U_{\mathrm{Kli},2}$ in contrast to T is well behaved in weight $(2,2)$. This complex is now a complex over the rank one Iwasawa algebra, and a specialization to weight $\kappa = (2,2)$ gives the complex which we ultimately want to use in our modularity lifting arguments.

In order to avoid missing the forest through the trees, let us recap in reverse:

(1) We construct complexes $M^\bullet$ of amplitude $[0,1]$ by first constructing big complexes $M_\Lambda^\bullet$ over an Iwasawa algebra parametrizing irregular weights $(k,2)$ and then specializing to weight $\kappa = (2,2)$. We know enough about this complex in H^0 to compare H^0 of this specialization to a space of classical forms.

(2) Under appropriate assumptions on $\mathfrak{m}$ and $\bar\rho$, by doubling we can prove the required local–global compatibility for H^0.

(3) We prove comparisons between the three complexes $M^\bullet$ and its specializations with the other complexes $\mathrm{R}\Gamma(X_1^{\geq 1}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_1, \omega^\kappa)$.

Note that $X_1^{\geq 1}$ is related but not quite the same as the special fibre $X_{\mathrm{Kli},1}^I$ at Klingen level (unlike for modular curves where the analogous objects were literally the same.)

(4) We can prove enough about the H^1 terms of these complexes in characteristic zero (since we are in the $l_0 = 1$ situation) to apply the Taylor–Wiles method. Using language from [**CG18**], we want to prove that the complex $M^\bullet$ is balanced, which amounts to showing that the Euler characteristic $(\dim H^0 - \dim H^1)$ is ≤ 0, which ultimately follows from specialization of the big complex $M_\Lambda^\bullet$ not to weight $\kappa = (2,2)$ but to higher irregular weights $(k,2)$ (where the H^2 terms vanish in characteristic zero, and where our comparison theorems allow us to compare to $\mathrm{R}\Gamma(X_1, \omega^\kappa)$), and then using local constancy of Euler characteristics.

A key ingredient is to construct families of modular forms (or at least complexes) of varying irregular weights, which is related to notions of overconvergence first studied by Coleman and recently vastly generalized by [**AIP15**]. We freely admit that this is all sufficiently new so that our results are probably not optimal as far as either the method of proofs or the strength of the conclusions. However, we were motivated by proving a particular theorem so optimization and elegance of arguments was not always the highest priority.

We are now ready to pass from $\mathbf{Q}$ to F. (Ready or not, here we go.)

3.5. Siegel Modular Forms of parallel weight two over totally real fields F. Let us now pass to totally real fields F, where we shall assume that p is totally split. One (of several) advantages of working with p totally split is that the local geometry at all $v|p$ is approximated by a product of what happens for $F = \mathbf{Q}$. The invariant l_0 in the case of abelian surfaces is (Lemma 2.8.5) equal to $\sum_{v|p} 1 = [F : \mathbf{Q}]$. Just as we did above, with $U_v = U_{\mathrm{Kli}(v),2}$, we can construct a complex

$$(7) \qquad \prod_{v|p} e(U_v)\mathrm{R}\Gamma(\mathfrak{X}_{\mathrm{Kli}}^I(p^\infty), \Omega^\kappa)_{\mathfrak{m}}$$

of amplitude $[0, l_0]$ which "should," by all measure, be the "correct" complex after specialization to parallel weight $(2,2)$. (To be fair, the corresponding complex $\mathrm{R}\Gamma(X, \omega^2)_{\mathfrak{m}}$ should also be the correct complex as well.) The difficulty is that the only arguments we have for proving local–global compatibility apply to H^0, and to say anything about the complex of Equation 7 we would need local-global compatibility results for H^i for all $i < l_0$ (as well as some extra information about H^{l_0} in characteristic zero.) This is something we have no idea how to do! Instead, what we actually do is a dirty trick. Let us work backwards to motivate our construction. We want to construct a classical form f of weight $(2,2)$ which is ordinary for all $v|p$. It turns out to be equivalent to construct a corresponding classical form f whose level at all $v|p$ has Klingen level structure. This form will be an eigenvector for a Hecke operator $U_{\mathrm{Kli}(v),1}$ (and $U_{\mathrm{Kli}(v),2}$) for $v|p$. The eigenvalue of $U = U_{\mathrm{Kli}(v),2}$

will be $\alpha_v \beta_v$, but the eigenvalue of $U_{\mathrm{Kli}(v),1}$ will be $\alpha_v + \beta_v$. Associated to any set $I \subset v|p$, we can consider the oldforms which have Iwahori level at $v \in I$. For each of these forms, and for each choice $\mathfrak{r}_v \in \{\alpha_v, \beta_v\}$ (which we assume are distinct modulo p), there should be an old form with the following properties:

(1) At primes $v \in I$, the eigenvalue of $U_{\mathrm{Iw}(v),1}$ is $\mathfrak{r}_v \in \{\alpha_v, \beta_v\}$.
(2) At primes $v \notin I$, the eigenvalue of $U_{\mathrm{Kli}(v),1}$ is $\alpha_v + \beta_v$.

For a fixed I, there are $2^{|I|}$ such forms. These collections of oldforms may be familiar to readers who have seen a proof of the Artin conjecture for totally real fields in which p splits completely. The strategy in the Artin case is that, simply using classical ordinary modularity lifting theorems, one constructs the $2^{[F:\mathbf{Q}]}$ forms f for $I = \{v|p\}$ over the ordinary locus, and then uses analytic continuation and étale descent and the q-expansion principle to glue these forms together and extend over the entire Hilbert modular variety to obtain the classical form (following the original approach of [**BT99**].) It is well known that this approach doesn't work in the Siegel case, because the q-expansions cannot be recovered from the Hecke eigenvalues. Instead, what we first do is construct the $2^{[F:\mathbf{Q}]}$ forms $f_{\mathfrak{r}}$ on the ordinary locus (and their analytic continuations) with $I = \{v|p\}$ using $l_0 = 0$. But *then* we construct the $2^{[F:\mathbf{Q}]-1}$ forms $g_{w,\mathfrak{r}}$ for every $J = \{v|p\} \setminus \{w\}$ of size $[F : \mathbf{Q}]-1$. (So in total there are $[F : \mathbf{Q}]2^{[F:\mathbf{Q}]-1}$ such forms $g_{w,\mathfrak{r}}$ as w and $\mathfrak{r}$ vary.) The construction of these forms $g_{w,\mathfrak{r}}$ (which have analytical continuations into regions which depend on whether $v \in J$ or not) comes by applying modularity results in the $l_0 = 1$ setting in a manner completely analogous to what was described in the $l_0 = 1$ case over $\mathbf{Q}$, although now using complexes which are a mix of Klingen tower level structure at $w|p$ and Iwahori level structure at all other $v|p$. (There is a certain amount of horrible indexing pain that occurs in [**BCGP18**] but seems somewhat unavoidable.) By multiplicity one results for the ordinary forms (coming out of the usual Taylor–Wiles method) we have that $g_{w,\mathfrak{r}}$ lie inside the dimension two space generated by $f_{\mathfrak{r}}$ where $\mathfrak{r}_v$ is determined by g for $v \in J$, and $\mathfrak{r}_w \in \{\alpha_w, \beta_w\}$. In particular, our new $l_0 = 1$ modularity theorem, allows us (in effect) to glue together these pair of forms $f_{\mathfrak{r}}$ to get $g_{w,\mathfrak{r}}$ without having to use q-expansions (because $g_{w,\mathfrak{r}}$ has been constructed by the $_0 = 1$ modularity lifting theorem instead of gluing), and thus to be able to analytically continue them deeper into our space than one can achieve simply in the classical ordinary case. This crucial step then turns out to be enough (combined with étale descent and induction for all $J \subset \{v|p\}$) to obtain the desired form f over a suitably large region of X so that it can then be analytically continued across all of X_{Kli} using the approach first found by Kassaei [**Kas06**]. Hence we obtain classicality of f which completes the proof.

3.6. A few extras in passing. It's always a little hard to describe a 300 page paper in 3 pages. One remark that might be worth making is why

70 of those pages are taken up with the Taylor–Wiles method. One point is that we do need to explain why Taylor's Ihara avoidance argument works for $\mathrm{GSp}(4)$ rather than $\mathrm{GL}(n)$ — there are a few twists which require slightly different treatment here. The main point, however, is as follows. If our desired goal was simply to prove the potential modularity of abelian surfaces, there is enough flexibility in the choice of p to ensure that all the local ordinary deformation rings which arise are smooth. This would have simplified a number of issues both related to the properties of these local deformation rings as well as some local representation theory. However, we made the extra effort to relax the local conditions in order that our theorems apply to ordinary abelian surfaces when $p = 3$. One reason for doing so is that it is possible to produce many (infinitely many) abelian surfaces over $\mathbf{Q}$ with $\mathrm{End}^0(A_{\mathbf{C}}) = \mathbf{Z}$ which we can prove are modular, rather than potentially modular. This is a consequence of the happy accident that the moduli space $\mathcal{A}_2(3)$ with principal level structure at 3 is rational, since this space is actually birational to the Burkhardt quartic ([**Bur91, Cob06, Bak46, Hun96, BN17**]). Not only that, but given any representation

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$$

whose similitude character is cyclotomic, the corresponding twist $\mathcal{A}_2(\overline{\rho})$ is always unirational over $\mathbf{Q}$ of degree at most 6 and thus contains infinitely many rational points. (Curiously enough, this twist is typically *not* rational over $\mathbf{Q}$ and often not even unirational for any map of degree less than 6 — this happens, for example, when $\overline{\rho}$ has surjective image [**CC18**]).

3.7. While you sit here (as I am leaving). My hope is that these lectures leave you in a slightly better mood than that of Captain Haddock.

FIGURE 1. Captain Haddock, from *The Calculus affair* by Hergé. This probably won't make it into the published version for copyright reasons.

References

[AC14]   Patrick B. Allen and Frank Calegari, *Finiteness of unramified deformation rings*, Algebra Number Theory **8** (2014), no. 9, 2263–2272.

[ACC+18]   Patrick Allen, Frank Calegari, Ana Caraiani, Toby Gee, David Helm, Bao Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack Thorne, *Potential automorphy over CM fields*, in preparation, 2018.

[AIP15]   Fabrizio Andreatta, Adrian Iovita, and Vincent Pilloni, *p-adic families of Siegel modular cuspforms*, Ann. of Math. (2) **181** (2015), no. 2, 623–697.

[Bak46]   H. F. Baker, *A Locus with* 25920 *Linear Self-Transformations*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 39, Cambridge, at the University Press; New York, The Macmillan Company, 1946.

[BCDT01]   Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* **Q**: *wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

[BCGP18]   George Boxer, Frank Calegari, Toby Gee, and Vincent Pilloni, *Abelian surfaces over totally real fields are potentially modular*, preprint, 2018.

[BK14]   Armand Brumer and Kenneth Kramer, *Paramodular abelian varieties of odd conductor*, Trans. Amer. Math. Soc. **366** (2014), no. 5, 2463–2516.

[BL04]   Christina Birkenhake and Herbert Lange, *Complex abelian varieties*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 302, Springer-Verlag, Berlin, 2004.

[BLGGT14] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), no. 2, 501–609.

[BN17]   Nils Bruin and Brett Nasserden, *Arithmetic aspects of the Burkhardt quartic threefold*, preprint, 2017.

[Box15]   G. Boxer, *Torsion in the Coherent Cohomology of Shimura Varieties and Galois Representations*, preprint, July 2015.

[Bra47]   Richard Brauer, *On Artin's L-series with general group characters*, Ann. of Math. (2) **48** (1947), 502–514.

[BT99]   Kevin Buzzard and Richard Taylor, *Companion forms and weight one forms*, Ann. of Math. (2) **149** (1999), no. 3, 905–919.

[Bur91]   Heinrich Burkhardt, *Untersuchungen aus dem Gebiete der hyperelliptischen Modulfunctionen*, Math. Ann. **38** (1891), no. 2, 161–224.

[Buz14]   Kevin Buzzard, *Computing weight one modular forms over* **C** *and* $\overline{\mathbf{F}}_p$, Computations with modular forms, Contrib. Math. Comput. Sci., vol. 6, Springer, Cham, 2014, pp. 129–146.

[BV13]   Nicolas Bergeron and Akshay Venkatesh, *The asymptotic growth of torsion homology for arithmetic groups*, J. Inst. Math. Jussieu **12** (2013), no. 2, 391–447.

[BW00]   A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, second ed., Mathematical Surveys and Monographs, vol. 67, American Mathematical Society, Providence, RI, 2000.

[Cai14]   Bryden Cais, *On the U_p operator in characteristic p*, Math. Res. Lett. **21** (2014), no. 2, 255–260.

[Cal09]   Frank Calegari, *Even Galois Representations and the Fontaine-Mazur conjecture*, preprint, 2009.

[Cal12]   ———, *Even Galois Representations and the Fontaine-Mazur conjecture II*, J. Amer. Math. Soc. **25** (2012), no. 2, 533–554.

[Car98]   Henri Carayol, *Limites dégénérées de séries discrètes, formes automorphes et variétés de Griffiths-Schmid: le cas du groupe* U$(2,1)$, Compositio Math. **111** (1998), no. 1, 51–88.

[CC18]   Frank Calegari and Shiva Chidambaram, *Rationality of twists of $\mathcal{A}(3)$*, in preparation, 2018.

[CDT99]   Brian Conrad, Fred Diamond, and Richard Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), no. 2, 521–567.

[CE09]   Frank Calegari and Matthew Emerton, *Bounds for multiplicities of unitary representations of cohomological type in spaces of cusp forms*, Ann. of Math. (2) **170** (2009), no. 3, 1437–1446.

[CG16]   Frank Calegari and David Geraghty, *Modularity lifting for non-regular symplectic representations*, preprint, 2016.

[CG18]   Frank Calegari and David Geraghty, *Modularity lifting beyond the Taylor–Wiles method*, Inventiones mathematicae **211** (2018), no. 1, 297–433.

[CHT08]   Laurent Clozel, Michael Harris, and Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Pub. Math. IHES **108** (2008), 1–181.

[CLH16]   Ana Caraiani and Bao V. Le Hung, *On the image of complex conjugation in certain Galois representations*, Compos. Math. **152** (2016), no. 7, 1476–1488.

[Cob06]   A. B. Coble, *An Invariant Condition for Certain Automorphic Algebraic Forms*, Amer. J. Math. **28** (1906), no. 4, 333–366.

[Con07]   Brian Conrad, *Arithmetic moduli of generalized elliptic curves*, J. Inst. Math. Jussieu **6** (2007), no. 2, 209–278.

[CS17]   Frank Calegari and Joel Specter, *Pseudo-representations of weight one*, preprint, 2017.

[DDT97]   Henri Darmon, Fred Diamond, and Richard Taylor, *Fermat's last theorem*, Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), Int. Press, Cambridge, MA, 1997, pp. 2–140.

[Del74]   Pierre Deligne, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. (1974), no. 43, 273–307.

[Edi92]   Bas Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594.

[Edi06]   ———, *Comparison of integral structures on spaces of modular forms of weight two, and computation of spaces of forms mod 2 of weight one*, J. Inst. Math. Jussieu **5** (2006), no. 1, 1–34, With appendix A (in French) by Jean-François Mestre and appendix B by Gabor Wiese.

[ERX17]   Matthew Emerton, Davide Reduzzi, and Liang Xiao, *Unramifiedness of Galois representations arising from Hilbert modular surfaces*, Forum Math. Sigma **5** (2017), e29, 70.

[Fal83]   G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366.

[FKRS12]   Francesc Fité, Kiran S. Kedlaya, Víctor Rotger, and Andrew V. Sutherland, *Sato-Tate distributions and Galois endomorphism modules in genus 2*, Compos. Math. **148** (2012), no. 5, 1390–1442.

[FM95]   Jean-Marc Fontaine and Barry Mazur, *Geometric Galois representations*, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993), Ser. Number Theory, I, Int. Press, Cambridge, MA, 1995, pp. 41–78.

[Fon82]   Jean-Marc Fontaine, *Sur certains types de représentations p-adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti-Tate*, Ann. of Math. (2) **115** (1982), no. 3, 529–577.

[Fon90]   ———, *Représentations p-adiques des corps locaux. I*, The Grothendieck Festschrift, Vol. II, Progr. Math., vol. 87, Birkhäuser Boston, Boston, MA, 1990, pp. 249–309.

[Fon94a]   ———, *Le corps des périodes p-adiques*, Astérisque (1994), no. 223, 59–111, With an appendix by Pierre Colmez, Périodes *p*-adiques (Bures-sur-Yvette, 1988).

[Fon94b]   ———, *Représentations p-adiques semi-stables*, Astérisque (1994), no. 223, 113–184, With an appendix by Pierre Colmez, Périodes *p*-adiques (Bures-sur-Yvette, 1988).

[GK15]   W. Goldring and J.-S. Koskivirta, *Strata Hasse invariants, Hecke algebras and Galois representations*, preprint, July 2015.

[Gro90]   Benedict H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. **61** (1990), no. 2, 445–517.

[GV18]   S. Galatius and A. Venkatesh, *Derived Galois deformation rings*, Adv. Math. **327** (2018), 470–623.

[Har90]   Michael Harris, *Automorphic forms and the cohomology of vector bundles on Shimura varieties*, Automorphic forms, Shimura varieties, and *L*-functions, Vol. II (Ann Arbor, MI, 1988), Perspect. Math., vol. 11, Academic Press, Boston, MA, 1990, pp. 41–91.

[Hec20]   E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen*, Math. Z. **6** (1920), no. 1-2, 11–51.

[HLTT16]   Michael Harris, Kai-Wen Lan, Richard Taylor, and Jack Thorne, *On the rigid cohomology of certain Shimura varieties*, Res. Math. Sci. **3** (2016), 3:37.

[How96]   Everett W. Howe, *Constructing distinct curves with isomorphic Jacobians*, J. Number Theory **56** (1996), no. 2, 381–390.

[Hun96]   Bruce Hunt, *The geometry of some special arithmetic quotients*, Lecture
            Notes in Mathematics, vol. 1637, Springer-Verlag, Berlin, 1996.
[HZ94]   Michael Harris and Steven Zucker, *Boundary cohomology of Shimura vari-
            eties. I. Coherent cohomology on toroidal compactifications*, Ann. Sci. École
            Norm. Sup. (4) **27** (1994), no. 3, 249–344.
[HZ01]   _____, *Boundary cohomology of Shimura varieties. III. Coherent cohomol-
            ogy on higher-rank boundary strata and applications to Hodge theory*, Mém.
            Soc. Math. Fr. (N.S.) (2001), no. 85, vi+116.
[Joc82]   Naomi Jochnowitz, *Congruences between systems of eigenvalues of modular
            forms*, Trans. Amer. Math. Soc. **270** (1982), no. 1, 269–285.
[Kas06]   Payman L. Kassaei, *A gluing lemma and overconvergent modular forms*, Duke
            Math. J. **132** (2006), no. 3, 509–529.
[Kat12]   Nicholas Katz, *Simple things we don't know*, notes from the Colloquio De
            Giorgi at Scuola Normale Superiore on March 21, 2012.
[Kle94]   Steven L. Kleiman, *The standard conjectures*, Motives (Seattle, WA, 1991),
            Proc. Sympos. Pure Math., vol. 55, Amer. Math. Soc., Providence, RI, 1994,
            pp. 3–20.
[KT17]   Chandrashekhar B. Khare and Jack A. Thorne, *Potential automorphy and
            the Leopoldt conjecture*, Amer. J. Math. **139** (2017), no. 5, 1205–1273.
[Lan80]   Robert P. Langlands, *Base change for* GL(2), Annals of Mathematics Stud-
            ies, vol. 96, Princeton University Press, Princeton, N.J.; University of Tokyo
            Press, Tokyo, 1980.
[LS13]   Kai-Wen Lan and Junecue Suh, *Vanishing theorems for torsion automorphic
            sheaves on general PEL-type Shimura varieties*, Adv. Math. **242** (2013), 228–
            286.
[Maz89]   B. Mazur, *Deforming Galois representations*, Galois groups over **Q** (Berkeley,
            CA, 1987), Math. Sci. Res. Inst. Publ., vol. 16, Springer, New York, 1989,
            pp. 385–437.
[Maz97]   Barry Mazur, *An introduction to the deformation theory of Galois repre-
            sentations*, Modular forms and Fermat's last theorem (Boston, MA, 1995),
            Springer, New York, 1997, pp. 243–311.
[MB89]   L. Moret-Bailly, *Groupes de Picard et problèmes de Skolem II*, Ann. Scient.
            Ec. Norm. Sup. **22** (1989), 181–194.
[Mum08]   David Mumford, *Abelian varieties*, Tata Institute of Fundamental Research
            Studies in Mathematics, vol. 5, Published for the Tata Institute of Funda-
            mental Research, Bombay; by Hindustan Book Agency, New Delhi, 2008,
            With appendices by C. P. Ramanujam and Yuri Manin, Corrected reprint of
            the second (1974) edition.
[Pil17]   Vincent Pilloni, *Higher coherent cohomology and p-adic modular forms of
            singular weight*, preprint, 2017.
[PT15]   Stefan Patrikis and Richard Taylor, *Automorphy and irreducibility of some
            l-adic representations*, Compos. Math. **151** (2015), no. 2, 207–229.
[Rie59]   B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*,
            Monatsberichte der Berliner Akademie (1859), 671?680.
[Rie06]   Emily Riehl, *Lubin–tate formal groups and local class field theory*, Senior
            Thesis, Harvard University, 2006.
[Sch15a]   George J. Schaeffer, *Hecke stability and weight 1 modular forms*, Math. Z.
            281 (2015), no. 1-2, 159–191.
[Sch15b]   Peter Scholze, *On the p-adic cohomology of the Lubin-Tate tower*, preprint,
            June 2015.
[Ser70]   Jean-Pierre Serre, *Facteurs locaux des fonctions zêta des variétés algébriques
            (définitions et conjectures)*, Séminaire Delange-Pisot-Poitou. Théorie des
            nombres **11** (1969-1970), no. 2, 1–15 (fre).

[Ser73]   _____ , *Formes modulaires et fonctions zêta p-adiques*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), Springer, Berlin, 1973, pp. 191–268. Lecture Notes in Math., Vol. 350.

[Ser03]   _____ , *On a theorem of Jordan*, Bull. Amer. Math. Soc. (N.S.) **40** (2003), no. 4, 429–440 (electronic).

[Sha81]   Freydoon Shahidi, *On certain L-functions*, Amer. J. Math. **103** (1981), no. 2, 297–355.

[Tat67]   J. T. Tate, *p-divisible groups*, Proc. Conf. Local Fields (Driebergen, 1966), Springer, Berlin, 1967, pp. 158–183.

[Tay03]   R. Taylor, *Remarks on a conjecture of Fontaine and Mazur*, Journal of the Institute of Mathematics of Jussieu **1** (2003), no. 01, 125–143.

[Tay08]   Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations. II*, Publ. Math. Inst. Hautes Études Sci. (2008), no. 108, 183–239.

[Tay18]   Noah Taylor, *Sato-tate distributions on abelian surfaces*, preprint, 2018.

[Tsu99]   Takeshi Tsuji, *p-adic étale cohomology and crystalline cohomology in the semi-stable reduction case*, Invent. Math. **137** (1999), no. 2, 233–411.

[Tun81]   Jerrold Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. (N.S.) **5** (1981), no. 2, 173–175.

[TW95]   Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[Ven17]   Akshay Venkatesh, *Derived hecke algebra and cohomology of arithmetic groups*, submitted, 2017.

[Voi02]   Claire Voisin, *A counterexample to the Hodge conjecture extended to Kähler varieties*, Int. Math. Res. Not. (2002), no. 20, 1057–1075.

[Wie14]   Gabor Wiese, *On Galois representations of weight one*, Doc. Math. **19** (2014), 689–707.

[Wil95]   Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Zuc77]   Steven Zucker, *The Hodge conjecture for cubic fourfolds*, Compositio Math. **34** (1977), no. 2, 199–209.

E-mail address: `fcale@math.uchicago.edu`

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