

**Part 1. Matthew Emerton, *Aspects of  $p$ -adic categorical local Langlands for  $\mathrm{GL}_2(\mathbb{Q}_p)$***

Fix  $p \geq 5$  and  $G := \mathrm{GL}_2(\mathbb{Q}_p)$ . Let  $E/\mathbb{Q}_p$  be a finite extension, with ring of integers  $\mathcal{O}$  and residue field  $\mathbb{F}$ . Let  $\zeta : G_{\mathbb{Q}_p} \rightarrow \mathcal{O}^\times$  be trivial (or  $\varepsilon$ , the cyclotomic character, if you want). It will play the role of a central character.

Let  $\mathcal{A}$  denote the category of smooth representations of  $G$ , with central character  $\zeta$ , on locally torsion  $\mathcal{O}$ -modules, i.e., any element of the module is killed by some  $p^n$ .

We denote by  $\mathfrak{X}$  the “ $EG$  stack” parametrizing families of projective rank 2 étale  $(\varphi, \Gamma)$ -modules over Fontaine’s ring

$$\mathbf{A}_{\mathbb{Q}_p} = \widehat{\mathbb{Z}_p((\tau))},$$

where  $\varphi(1+\tau) = (1+\tau)^p$  and  $\gamma_a(1+\tau) = (1+\tau)^a$ . Here “family” means family over a  $p$ -adically complete  $\mathcal{O}$ -algebra. To match the central character condition, we’ll have a determinant condition. The determinant of our rank two  $(\varphi, \Gamma)$ -modules will be a rank one  $(\varphi, \Gamma)$ -module, and we want this determine to be to be  $\zeta\varepsilon^{-1}$ . Those are the objects in play.

**Theorem 1** (Dotto–Emerton–Gee). *There exists a fully faithful functor*

$$\mathfrak{a} : D_{f.g.}^b(\mathcal{A}) \hookrightarrow D_{\mathrm{coh}}^b(\mathfrak{X}).$$

**Remark 2.** It might be that what we’ve written doesn’t characterize the functor. We don’t claim that there is a unique functor. It is going to satisfy certain properties. We don’t know that they characterize it, but it is in some sense reasonable. Maybe at first it’s a bit hard to see what we mean at all, but a very rough idea is that an object in  $\mathcal{A}$ , which is a smooth representation, well, one could at least think of a smooth representation that is not irreducible (or at least admissible) as some family, or integral, of representations that are irreducible (or admissible). There is a classical mod  $p$  Langlands that relates irreducible mod  $p$  representations (the building blocks in  $\mathcal{A}$ ) to semisimple mod  $p$  2-dimensional Galois representations of  $G_{\mathbb{Q}_p}$ , which are somehow closed points in  $\mathfrak{X}$ . So the building block of the functor will be that the building blocks in  $\mathcal{A}$  go to some sort of skyscrapers in  $\mathfrak{X}$ , and a typical element of  $\mathcal{A}$  will be some direct integral of building blocks, while a typical element of  $\mathfrak{X}$  will be a direct integral of its fibers. That’s how you’re supposed to imagine this functor going.

When you have a derived category and put a subscript at the bottom (indicating cohomologies in a subcategory), this condition should be a reasonably behaved one for this notion to give you a reasonable subcategory. One would want to know that “finitely generated” is a reasonable condition to put on  $\mathcal{A}$ -modules. There’s kind of a fact which is that in  $\mathcal{A}$ , “finitely generated” is equivalent to “finitely presented”. So  $\mathcal{A}$  has a kind of “Noetherian” nature. We want to say a bit more about this, partly to introduce more notation and concepts, but also, it’s clear what it means for a representation to be finitely generated, but less so for finitely presented in this setting, since  $\mathcal{A}$  is not a category of modules over a ring. So let’s say a bit more about what “finitely presented” means here. The building blocks of the finitely generated modules are the *free* finite rank modules. The corresponding building blocks in  $\mathcal{A}$

If you have a finitely generated element of  $\mathcal{A}$ , so it’s some  $G$ -representation with a finite number of generators, then what you can do is you can act using  $K = \mathrm{GL}_2(\mathbb{Z}_p)$

and the center  $Z \leq G$  (which just acts by the scalar  $\zeta$ ) and act on your finitely many generators with  $KZ$ . This will give you a finitely-generated  $KZ$  module, but since the representation is smooth, you'll actually get a finite  $\mathcal{O}$ -module. A typical finitely generated  $G$  module will be generated over  $G$  by some finite  $\mathcal{O}$ -module  $V$ , which is a representation of  $KZ$ . So what you can do is, your representation that you're thinking about will be a quotient of a compact induction

$$\mathrm{ind}_{KZ}^G V \twoheadrightarrow \pi. \quad (1)$$

Thus such compact inductions are the building blocks of the finitely generated representations. Now  $\pi$  being “finitely presented” is equivalent, by definition if you'd like, to the kernel of the map (1) being finitely generated. (We note that if it is true for one such  $V$ , it will be true for all  $V$ .) It's true for  $\mathrm{GL}_2(\mathbb{Q}_p)$  that finitely generated implies finitely presented (by Berthel–Livre (?) and Breuil), but that's not an obvious fact, nor a general one (e.g., it fails for  $\mathrm{GL}_2$  over a quadratic extension of  $\mathbb{Q}_p$ ).

We want to come back to these representations in a little bit, but first, let's say what we mean by coherent.  $\mathfrak{X}$  is a formal algebraic stack, and is in fact Noetherian, so you can write it as an increasing union, or colimit, of algebraic stacks  $\mathfrak{X}_n$ :

$$\mathfrak{X} = \mathrm{colim}_n \mathfrak{X}_n.$$

Here  $\mathfrak{X}_n \hookrightarrow \mathfrak{X}_{n+1}$  are closed immersions, which we can think of as a “thickening”. You can push forward a coherent sheaf on  $\mathfrak{X}_n$  to think of it as a coherent sheaf on  $\mathfrak{X}_{n+1}$ , so you can take the colimit of those maps at the level of derived categories:

$$D_{\mathrm{coh}}^b(\mathfrak{X}) := \mathrm{colim}_n D_{\mathrm{coh}}^b(\mathfrak{X}_n).$$

You have to be a bit careful thinking about such things, however. For instance, the structure sheaf  $\mathcal{O}_{\mathfrak{X}}$  is *not* coherent (by this definition). The structure sheaf  $\mathcal{O}_{\mathfrak{X}_n}$  is a coherent sheaf on  $\mathfrak{X}$ , and so  $\mathcal{O}_{\mathfrak{X}}$  is kind of an inverse limit  $\varinjlim_n \mathcal{O}_{\mathfrak{X}_n}$ , so in our framework, it will be a *pro*-coherent sheaf, but not itself a coherent sheaf:

$$\mathcal{O}_{\mathfrak{X}} = \varinjlim_n \mathcal{O}_{\mathfrak{X}_n} \in \mathrm{Pro} D_{\mathrm{coh}}^b(\mathfrak{X}).$$

So that's at least some slight orientation about what the categories are. Now what we'd like to say is, how would you produce such a functor?

We should have said earlier that one key property of this functor  $\mathfrak{a}$  is that it is *exact*, i.e., preserves distinguished triangles. We actually use the language of stable  $\infty$  categories rather than derived categories, which offers some technical advantages, but in the language of derived categories, we certainly want the property of exactness. This functor will not be *t* exact

There are abelian categories  $\mathcal{A}$  and  $\mathcal{X}$  underlying  $\mathcal{A}$  and  $\mathfrak{X}$ , but these will not be preserved by our functor, so working with derived categories is out of necessity. But if we were working with such abelian categories, say thinking of them as categories of modules over a ring, we could try to understand functors between them by understanding where the ring goes. What you discover is that any such functor is given by some sort of tensor, which you might call the *kernel* of such a functor (in the sense of an integral kernel). So if you're given a functor and want to find the kernel, you just apply the functor to the ring, and this gives you the kernel to describe the functor in general. This gives a clue as to how you should think of this functor, but like we said, it's not the category of modules over a ring. Why not? Well, roughly for two reasons. What would the ring be? Well, you're looking

at the category of representations of a group  $G$ , so it'd be the category of modules over the group ring of  $G$ , but why isn't it that? Well, roughly, we're looking at  $p$ -power torsion modules, but not with any bounded amount of torsion necessarily, so the only coefficient ring that would make sense is  $\mathcal{O}$ , but we're not allowed to take  $\mathcal{O}[G]$  itself or even the trivial representation of  $G$  on  $\mathcal{O}$  as a module of  $A$ , because  $\mathcal{O}[G]$  acts on modules that are too  $p$ -adically big to be in our category, but also there's no continuity condition built into that group ring, so that group ring acts on representations that are not continuous or smooth, so it doesn't take into account the topology.

So again, any finitely generated module is killed by some  $\mathcal{O}/p^n$ , and then you might have your group  $K$  and in fact  $KZ$  acting (with  $Z$  acting via  $\zeta$ ), and then some  $H$  that is open and normal in  $K$  might be acting trivially because the action is smooth, so the group algebra

$$\mathcal{O}/p^n[KZ/H]_\zeta$$

would be acting, but then the group ring of  $G$  would act compatibly with the group ring of  $KZ$ , leading to

$$\mathcal{O}[G]_\zeta \otimes_{\mathcal{O}[KZ]_\zeta} \mathcal{O}/p^n[KZ/H]_\zeta = \text{ind}_{KZ}^G \mathcal{O}/p^n[KZ/H]_\zeta,$$

and then we might take an inverse limit, giving

$$\mathcal{O}[[G]]_\zeta = \varprojlim \mathcal{O}[G]_\zeta \otimes_{\mathcal{O}[KZ]_\zeta} \widehat{\mathcal{O}/p^n[KZ/H]_\zeta} = \text{ind}_{KZ}^G \widehat{\mathcal{O}[[KZ]]_\zeta}. \quad (2)$$

So it's this inverse limit of the gadgets that were used for checking "finitely presented". You can either think of this as literally the inverse limit of a topological ring, or as a formal inverse limit – not an object of  $\mathcal{A}$ , but an object of  $\text{Pro } \mathcal{A}$ .

So again, if we actually were looking at the category of modules over a ring, this would be projective over that, but instead it's a projective pro-generator. In fact, it is the case that this category fully faithfully embeds in the category of finitely presented modules over this last ring (2). The key thing is that, to the extent that our category  $\mathcal{A}$  is a category of modules over anything, it is a category of modules over this ring (2), and so to find a kernel, we should apply our functor to this ring, leading us to consider

$$L_\infty := \mathfrak{a}(\mathcal{O}[[G]]_\zeta).$$

Applying our functor to this pro-object should lead to another pro-object. One of the first properties (not obvious) is that what this will give will be a pro-coherent sheaf (rather than a pro-complex). The name  $L_\infty$  goes back to a paper (with Caraiani, Gee, Shin, others) studying patching, where we posited a hypothetical true  $p$ -adic local Langlands object; " $L$ " was chosen because it came next to " $M$ " in the alphabet.

The Morita-theoretic yoga we were just recalling is that  $\mathfrak{a}$  is given by

$$\pi \mapsto L_\infty \otimes_{\mathcal{O}[[G]]_\zeta} \pi.$$

Here  $\pi$  is an object of  $\mathcal{A}$ , or more generally  $D_{\text{fg}}^b(\mathcal{A})$ . You can resolve  $\pi$  by considering free modules over this ring or via compact inductions at finite levels. If you just resolve  $\pi$  by finitely presented modules over this ring (which you can do), then you just get copies of  $L_\infty$ , which is a pro-object. So something you have to prove is that you actually land in coherent sheaves when you put in a  $\pi$ , rather than just pro-coherent sheaves. Note that  $L_\infty$  is pro-coherent (rather than coherent), so this is something special about  $\pi$ .

Okay, but first, we have to know what  $L_\infty$  actually is. For this, you have to guess. It's given by the ideas of Colmez, who explained how to convert 2-dimensional Galois representations into representations of  $G_{\mathbb{Q}_p}$  (a rather small object of  $\mathcal{A}$ ). He did it by a formula: take your Galois rep, turn it into a  $(\varphi, \Gamma)$ -module  $V$ , then take

$$D^{\natural} \otimes_{\zeta_{\varepsilon^2}} \mathbb{P}^1.$$

(This is notation Colmez introduced.) Here  $\mathbb{P}^1$  has an action of  $\mathrm{GL}_2(\mathbb{Q}_p)$ . We do this for the universal  $(\varphi, \Gamma)$ -module over  $\mathfrak{X}$ . That is to say, our stack has a universal  $(\varphi, \Gamma)$ -module, and we check that we can implement Colmez's construction naturally, producing this object.

What this object will look like is sort of, if this formal stack  $\mathfrak{X}$  were some formal spectrum  $\mathrm{Spf} A$  of some ring  $A$  (say locally in the fppf topology), then the  $(\varphi, \Gamma)$ -module is some finite rank module over something like  $\mathbb{A}_{\mathbb{Q}_p}$ , but with  $\mathbb{Z}_p$  replaced by this other algebra  $A$ . Then, roughly,  $\mathbb{P}^1$  looks like a power series ring. So if you're on a chart on  $\mathfrak{X}$  of the form  $\mathrm{Spf} A$ , then  $L_\infty$  is roughly  $A[[T]]$ . So we're already seeing the coefficients  $A$  (which is a bit like the structure sheaf of  $\mathfrak{X}$ , pro just along the formal directions in  $\mathfrak{X}$ ) but also this  $[[T]]$ , which is another pro direction. That's the pro nature of this  $L_\infty$ .

Having produced this  $L_\infty$ , it's not clear that the resulting functor does anything that you want, e.g., that it lands in  $D_{\mathrm{coh}}^b(\mathfrak{X})$  rather than some more pro version, and it's not clear that it's fully faithful. So to study this functor, we need some tools. We want to conclude the talking by saying what those tools are. For that, we need to come back to  $\mathfrak{X}$ , which we recall is this stack of  $(\varphi, \Gamma)$ -module. Fontaine tells us that if we take a point of  $\mathfrak{X}$  over some (literally) finite  $\mathcal{O}$ -algebra, that's a rank two  $(\varphi, \Gamma)$ -module, then this gives a Galois representation. So if you probe  $\mathfrak{X}$  with small rings, you get Galois representations. But if you probe  $\mathfrak{X}$  with *big* rings, you get stuff that's too big. But somehow

$$\mathfrak{X}[\overline{\mathbb{F}_p}] \leftrightarrow \bar{\rho} : G_{\mathbb{Q}_p} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_p}).$$

Now  $\mathfrak{X}$  is an algebraic stack, so it has these  $\overline{\mathbb{F}_p}$ -points (what you might call “finite type” points), which, if it were a variety, would be the closed points, but here we have an Artin rather than Deligne–Mumford stack, so such points are not in general closed, but instead specialize to semisimplifications. Thus the semisimple  $\bar{\rho}$ 's correspond to the closed  $\overline{\mathbb{F}_p}$ -points:

$$\mathfrak{X}(\overline{\mathbb{F}_p})_{\mathrm{closed}} \leftrightarrow \bar{\rho}^{\mathrm{ss}}.$$

There's another thing, a scheme  $X$  (that will in the minute be a formal scheme over  $\mathrm{Spf} \mathcal{O}$ ), such that  $X(\overline{\mathbb{F}_p})$  is in bijection with  $\mathfrak{X}(\overline{\mathbb{F}_p})_{\mathrm{closed}}$ . The underlying reduced scheme  $X_{\mathrm{red}}$  will be a chain of copies of the projective line, where the number of them is  $\frac{p+1}{2}$ , where the sign is given by  $\zeta(-1) = \pm 1$ . What we claim is that semisimple Galois representations naturally arrange themselves in this way. It's an observation that in some sense goes back to computations in the beginning of this century that were later brought out in conversations of Berget and Breuil, and partly pointed out by Kisin, but kind of carefully noted and promoted to thinking you could really think about by Emerton and Gee, as part of this *EG* stack, and so we'll maybe tell you how this goes, at least briefly.

We'll be looking at the reduction  $\bar{\varepsilon}$  of our cyclotomic character mod  $p$ . Maybe we're looking at a direct sum of two characters, say

$$\begin{pmatrix} w_\alpha \cdot \bar{\varepsilon}^i & \\ & w_{\alpha^{-1}} \cdot \bar{\varepsilon}^{-(i+1)} \end{pmatrix}. \quad (3)$$

Here  $i$  is a discrete variable, while  $\alpha$  lives in some  $\mathbb{G}_m$ . This kind of a thing is a common thing you'll see in number theory talks on Galois representations. A general principle in mathematics, going back at least to Arabic algebraists (pre-Renaissance), is that a piece of notation can refer either to a number or a variable, and this leads to the idea of the *EG* stack, which is that (3) should really be a family of things. Now, what happens if you let  $\alpha$  go to  $\infty$ ? There's a natural way to think about it coming from the theory of reduction of crystalline theorems (or Serre's conjectures, etc). If you send  $\alpha$  to 0, then the above will go to an irreducible induction, then take the inverse to get what happens near  $\infty$ .

So what we said is that the closed points of  $\mathfrak{X}$  are in bijection with these semisimple guys, and once you know that, there's a unique way to produce a morphism of underlying topological spaces,

$$|\mathfrak{X}| \rightarrow |X|,$$

once you've matched the closed points. Then you can essentially push forward the structure sheaf on  $\mathfrak{X}$  to get a structure sheaf  $\mathcal{O}_X = f_* \mathcal{O}_{\mathfrak{X}}$  under  $f : \mathfrak{X} \rightarrow X$ . By Paskunas, semisimple  $\bar{\rho}$  correspond to *blocks* of  $A$  that are locally admissible (colimits of admissible). One can make the following:

**Definition 3.** Given a closed subset  $Y \hookrightarrow |X|$ , we define

$$\mathcal{A}_Y := \left\{ \begin{array}{l} \pi \in \mathcal{A} : \text{all nonzero subquotients} \\ \text{of } \pi \text{ lie in blocks} \\ \text{induced by } Y \in \mathcal{Y} \end{array} \right\} \subseteq \mathcal{A}.$$

We then define, for  $U := |X| - Y$ , the quotient

$$\mathcal{A}_U := \mathcal{A} / \mathcal{A}_Y.$$

We then have the following result [?, Theorem 1.1.1]:

**Theorem 4** (DEG).  $\tilde{\mathcal{A}} : U \mapsto \mathcal{A}_U$  is a stack of abelian categories over  $|X|$ , with  $\Gamma(X, \tilde{\mathcal{A}}) = \mathcal{A}$ .

Now for  $x \in X(\overline{\mathbb{F}_p})$ , we have a formal substack  $D_{\text{fg}}^b(\mathcal{A}_x) \xrightarrow{\mathfrak{a}} D_{\text{fg}}^b(\mathfrak{X}_x)$ . In characteristic  $p$ , you can thicken up pseudorepresentations, but not deform them nontrivially. These sort of stacks were first studied by Carl(?)–Wang–Erickson. Using Colmez–Paskunas's work, and then the work of Johansson–Newton–Wang–Erickson, we can rewrite

$$\mathfrak{a}|_{D_{\text{fg}}^b(\mathcal{A}_x)}$$

in terms of Colmez's functors, and then using the mentioned work in this rewritten form, we deduce that this functor is fully faithful. This version of the functor was thus already studied in somewhat disguised form by those authors. We thus have some “fully faithfulness” available. With that in hand, our arguments need

to combine with this localization theory and diagrams like

$$\begin{array}{ccc} D_{\text{fg}}^b(\mathcal{A}) & \xrightarrow{\alpha} & D_{\text{coh}}^b(\mathfrak{X}) \\ \downarrow & & \downarrow \\ D_{\text{fg}}^b & \xrightarrow{\alpha_U} & D_{\text{coh}}^b(f^{-1}(U)). \end{array}$$

to get to the whole setting. That's roughly how we prove the theorem.

We note that nearly everything in the proof is specific to  $\text{GL}_2(\mathbb{Q}_p)$ , in a very fundamental way.

**Remark 5.**  $\mathfrak{X}$  makes sense in great generality, but  $X$  does not.

[next talk at 11:00!!!!]

## Part 2. Andrew Wiles (University of Oxford), *Non-abelian descent and modularity*

So I've been thinking for a long time now about how to prove modularity more generally. Modularity lifting is in very good shape, whereas proving residual modularity is much more hard, and besides the theorem of Khare and Wintenberger, there isn't something very general. Seems extremely difficult to generalize Khare–Wintenberger directly, so I've been trying to take the approach using potential modularity, which is known quite generally. Suppose you have a representation ( $\ell$ -adic, or a compatible system)  $\rho$ , over some totally real field  $F_0$ . You know that if you take some not-necessarily-solvable totally real field  $F$  of  $F_0$ , passing from  $\rho$  to  $\tilde{\rho}$ , then you can find an automorphic form  $f$  that corresponds to  $\tilde{\rho}$ . This  $f$  will be invariant (since  $\tilde{\rho}$  comes from  $F_0$ ), and the question then arises: does  $f$  equal to the lift  $\hat{f}_0$  of some  $f_0$ ? Namely, can we find the  $f_0$  over  $F_0$  such that the natural diagram commutes? That will be a way of finding modularity. Thus, potential modularity reduces the problem to descent, but unfortunately, it's nonsolvable descent.

One example where we can do this is cyclic base change, where we do get descent, but cyclic base changes depends upon the trace formula, and this works in contexts where you have (inner) forms or a group, or when you have a subgroup given by invariance under one element, as is the case for cyclic base change. For non-solvable base change, invariance is much more subtle, so we trade one problem for a much harder one.

What we're going to do is cyclic base change in a slightly different context from normal and see how to replace the arguments in that.

Let's take  $F$  and  $F_0$  totally real number fields. We're going to assume that  $F/F_0$  is a cyclic extension, with Galois group generated by some element  $\sigma$ . (This generality is sufficient for what we want to use it for.) We assume moreover that  $F/F_0$  is unramified. Let  $D_0$  be a quaternion algebra over  $F_0$  that is ramified at all the infinite places and unramified at the finite ones. To do this, we need to assume that  $[F_0 : \mathbb{Q}] \equiv 0 \pmod{2}$ . This restriction on the finite places is not really essential, but as we'll explain later, there's a reason why we wanted to include it.

We then form a Shimura variety (really a set)

$$X_0 := X_0(U_0) = D_0^\times \backslash D_{0,f}^\times / U_0,$$

where

- $D_0^\times$  is the multiplicative group of our quaternion algebra,

- $D_{0,f}^*$  is its group of finite adelic points, and
- $U_0$  is some level structure, say  $U_0 = U_0(\mathfrak{n}_0)$  (“ $T_0(\bullet)$ ” structure) with  $\mathfrak{n}_0$  square-free.

We then have a space

$$S(U_0) = \{f : X_0(U_0) \rightarrow \mathbb{C}\}.$$

It might seem like this is a very simple structure, but it carries the action of the Hecke operators in the usual way, and moreover, by the theorem of Jacquet–Langlands and Shimizu, this  $S(U_0)$  is isomorphic as a Hecke module to the space of weight two Hilbert modular forms of level  $U_0$ . I’m going to work with these spaces instead of with  $\mathrm{GL}_2$  and Hilbert modular forms. From the point of view of the trace formula, basically we’re focusing on the elliptic elements in the trace formula. I think one could do this for  $\mathrm{GL}_2$  instead, but it’s simpler and easier to handle if we do it in this context.

Let’s turn to the trace formula. Trace of what? Let  $(\pi_0) \subseteq \mathcal{O}_{F_0}$  be a prime that splits completely in  $F$ , thus  $\pi_0 = \prod_i \pi^{\sigma^i}$ , with  $(\pi) \subseteq \mathcal{O}_F$ . What we want to compute is the trace on the space of forms here. We’re going to consider the vector space  $\mathbb{C}[X_0]$ , which has two important bases:

- (1)  $X_0$  itself, given characteristic functions of points  $t_{0,i}$ , and
- (2) a basis of eigenforms  $\mathcal{F}_0$  for the Hecke operators.

What we want to compute is the trace of  $T_{\pi_0}$ . We can express this analytically in terms of  $\mathcal{F}_0$ , and geometrically in terms of the points. We write this as

$$\mathrm{trace} T_{\pi_0}|_{\mathcal{F}_0} = |T_{\pi_0} \cdot \Delta_{X_0}|, \quad \Delta_{X_0} = \Delta \subset X_0 \times X_0.$$

What we need to do is to relate the trace of  $T_{\pi_0}$  on  $\mathcal{F}_0$  to  $T_\pi$  acting on the space of invariant forms over  $F$ . But the way of getting from  $X_0$  to  $X$  is going to be via the geometric side. So we want to understand what this set  $T_{\pi_0} \cdot \Delta_{X_0}$  is.

So, what is an element  $Q \in T_\pi$ ? We pick some  $t_{0,i}$ , some coset representative  $\varpi_{\pi_0}$ , and we write

$$t_{0,i} \varpi_{\pi_0} = \gamma_0 t_{0,i} u_0,$$

where  $\gamma_0 \in D_0^\times$  and  $u_0 \in U_0$ . So we get some expression like that, and reading off from this, we can see for instance that  $\gamma_0$  has determinant (or norm) of the same ideal as  $\pi_0$ .

So associated to this  $Q$ , we have some  $t_{0,i}$  (some representative of  $X_0$ ) and an element  $\gamma_0$  of  $D_0^\times$ . It moreover satisfies a quadratic equation over  $F_0$ , so it actually lies in some quadratic extension

$$L_0 = F_0(\gamma_0).$$

That quadratic extension is necessarily CM (coming from the fact that  $D_0$  is ramified at all infinite places).

Now, what we want to do is associate a similar kind of thing over  $F$ . So, to get over  $F$  from the element  $\gamma_0$ , we choose

$$\gamma \in L = F(\gamma_0)$$

so that

$$\mathcal{N}_{L/L_0}(\gamma) = \gamma_0. \tag{4}$$

(Because it’s a cyclic extension, class field theory tells you that you only need to check this locally. It’s easy to see from our construction and Hasse’s theorem that

such a  $\gamma$  exists.) In view of (4), it's easy to see that

$$(\gamma) = \mathfrak{b}^{\sigma-1} \mathfrak{p},$$

where  $n(\mathfrak{p}) = (\pi)$  is a prime divisor of  $\pi$ , lying above the corresponding factorization of  $\gamma_0$ . So we have a prime factorization like this and can turn it back into adelic language. If we pick adelic representatives for the ideals, then we can write this as

$$\beta^\sigma \omega_\pi = \gamma \beta u.$$

Now, what does this say? Thinking about it, it gives us an element of  $T_\pi \sigma \cdot \Delta_X$ , where  $X = X(U_0)$  is defined like  $X_0$ , but now over  $F$ , in terms of  $D = D_0 \otimes_{F_0} F$ .

So what we've ended up with is a map

$$T_{\pi_0} \cdot \Delta_{X_0} \hookrightarrow T_{\pi^\sigma} \cdot \Delta_X. \quad (5)$$

Now we want to analyze surjectivity. In the trace formula (for cyclic base change, anyway), enormous use is made of the norm operator, which is defined like the usual norm operator, but defined on the non-abelian group  $D^\times$  rather than just on the multiplicative group of some field such as  $F^\times$ . This requires making a choice, such as

$$\mathcal{N}_\sigma : D \ni y \mapsto yy^\sigma y^{\sigma^2} \cdots$$

Obviously the norm of  $\gamma$  is  $\gamma_0$ . What you find is that

$$\mathcal{N}_\sigma(\gamma_1) \sim \mathcal{N}_\sigma(\gamma_2) \implies \gamma_1 \sim^\sigma \gamma_2.$$

Here

- $\sim$  denotes conjugacy,
- $\sim^\sigma$  denotes  $\sigma$ -conjugacy.

You can show in this way that there's actually a converse map to (5). That's the mechanism that the proof of cyclic base change will use. Actually, this refines (5) to an isomorphism.

Now in the usual development of cyclic base change, it's important to note that the norm map only gives you a norm map on conjugacy classes, i.e. it's not well-defined on elements. You can thus compare orbits and, more generally, orbital integrals, using this. But we're not going to get into it, because this method doesn't work in the non-solvable case at all.

We get that far, and with what we have, we can now complete the proof of base change in this case. We have

$$\text{trace } T_{\pi_0}|_{\mathcal{F}_0} = |T_{\pi_0} \cdot \Delta_{X_0}| = |T_{\pi^\sigma} \cdot \Delta_X| = \text{trace } T_{\pi^\sigma}|_{\mathcal{F}} = \text{trace } T_\pi|_{\mathcal{F}^{\text{inv}}}.$$

Here

- the second equality comes from the geometric argument that we just gave,
- the third equality comes from the trace formula for  $X$ ,
- the fourth equality uses that  $\sigma$  permutes the non-invariant forms, but is trivial on the invariant ones.

Okay, we've only done this for certain  $\pi$  (splitting completely into principal ideals), but now Cebotarev, using the associated Galois representations, tells us that we get this for all places.

So that's what we want to model this argument about generalizing the trace formula on.



The first key problem in doing so is that, as I said, we don't have one  $\sigma$ , so we can't apply this argument. But also, the cohomological method simply doesn't work more generally.

We replace this now, as in class field theory, we find an analytic argument to replace this cohomological argument. Let me explain that first, since this is much simpler. Let's focus on the case of a non-solvable extension, with no abelian subquotient. We'll continue to use  $F$  and  $F_0$  for the two fields, and we write  $d = [F : F_0]$  for their degree. We again choose primes  $\pi_0$  in  $F_0$  that split completely into principal primes in  $F$ , and consider the following sum

$$\sum_{\pi_0} \frac{1}{d} \sum_{\pi|\pi_0} \sum_{f \in \mathcal{F}^{\text{inv}}} a_{\pi}(f) \cdot \text{Norm}(\pi)^{-s}. \quad (6)$$

Here, again,  $\mathcal{F}^{\text{inv}}$  is (a basis for) the space of forms invariant under the action of the Galois group.

**Remark 6.** "Invariant" is in a suitably defined sense – there is a choice here regarding whether to take things up to twist. The thing that's confusing is that invariant cyclic Galois representations do not descend, but that's the only obstruction. For instance, projective representations always descend.

We're going to compare that, on the one hand, to

$$\sum_{\pi_0} \sum_{f \in \mathcal{F}_0} a_{\pi_0}(f_0) \text{Norm}(\pi_0)^{-s}. \quad (7)$$

Here, as before,  $\mathcal{F}_0$  is a basis for eigenforms on  $F_0$ .

In the above, we further restrict  $\pi_0$  to primes of degree 1 that split completely in  $M$ , a totally real field that contains the Hilbert class field  $H_F$  of  $F$ . The first sum (6) would then be the main part contributing to these  $L$ -functions for the supposed base changes to  $M$  of these forms. That is to say, if we actually had base changes to  $M$ , then this would be the main part, at least as far as convergence goes; the rest over  $M$  (if they existed) would have Dirichlet series that converge at  $\Re(s) = 3/2$ .

By potential automorphy, we can choose  $M$  such that  $\mathcal{F}_0$  and  $\mathcal{F}$  are simultaneously automorphic over  $M$ .

If we make a similar construction as in the discussion of cyclic base change, then the coefficients here in the non-solvable case would be

$$\frac{1}{d} \sum_{\pi_0} \sum_{\pi|\pi_0} \sum_{f \in \mathcal{F}^{\text{inv}}} a_{\pi}(f)$$

will be greater than or equal to

$$\sum_{\pi_0} \sum_{f_0 \in \mathcal{F}_0} a_{\pi_0}(f_0),$$

so the difference is nonnegative. Now they're not cuspidal these two, but we can match up the Eisenstein parts of (6) and (7) easily. We thus obtain an  $L$ -function that is cuspidal (i.e., no pole) and has positive coefficients. It's not difficult to deduce that for almost all  $\pi_0$ , we get equality.

That side of the argument, also as in class field theory, can be done relatively simply. The problem is the construction. How do we get this map between the two? If we go back to the same kind of argument here (try and construct a  $\gamma$  in

the same way), we can find that the norm of  $\gamma$  is  $\gamma_0$ , but what happens is that you get the identity, of ideals in  $L = F(\gamma_0) = F(\gamma)$ ,

$$(\gamma) = \prod_t \mathfrak{b}_t^{\sigma_t - 1} \cdot \mathfrak{p}.$$

(This uses some class field theory. You have to set it up right. It's not trivial, but you can put yourself in this situation. It's not just Hasse's theorem, as in the cyclic case.) We don't have a  $\sigma$  that is picking out the invariant forms. What I want to do now is, instead of what I did before, let's view this as one entry of

$$(\prod T_{b_t} \sigma_t) T_\pi t_{0,i} = \prod T_{b_t} \cdot t_{0,i}, \quad (8)$$

where  $b_t = n_{L/F}(\mathfrak{b}_t)$ .

**Step 1.** Replace  $T_\pi$  by  $T_\pi T_\alpha$ , where  $T_\alpha$  projects to the invariant part. Once you've done that, because  $T_\alpha$  projects, the identity (8) becomes

$$T_\pi T_\alpha \prod_t T_{b_t} t_{0,i} = \prod_t T_{b_t} t_{0,i}. \quad (9)$$

Now what we do is choose our  $T_{b_t}$  so that they're invertible. We have to make sure we can do that in the original thing. Then we view the products over  $t$  in (9) as new basis elements. Now we're getting something that looks more like the trace of of  $T_\pi T_\alpha$ . For this, we need that  $\prod T_{b_t}$  is invertible in  $\mathbb{C}[X]$ .

Okay, so that looks nice, but the issue is that in doing (8), I only gave you one coset representative for the Hecke operator, but we need the whole Hecke operator, not just one entry. This is where it becomes more complicated. I'm not going to go into it in detail, but the idea is that you make the  $b_t$ 's complicated ideals (products of many ideals) and you consider different factorizations of  $b_t$  and you get different representatives, and you have to do it in such a way that you can make sure you get all the representatives on both sides of (9). So you choose your  $b_t$ 's so that there are many prime factors, then you can always change each prime factor to its complex conjugate, which gives you a huge number of choices. Making sure this works is a little like the ergodic method that people have used. I don't want to go into the argument because it is rather involved, but for this argument to work, we need a version of Cebotarev in this context. What it says is that

- (1) If you pick one of these CM fields  $L_f^\times \rightarrow D_f^\times \rightarrow X(U_0)$ , then this factors through some class group

$$C_0 = L^\times \backslash L_f^\times / U_0^\times \rightarrow X(U_0).$$

(Here  $U_0^\times$  is some level structure in  $L_f^\times$ , corresponding to the class group of an order.) We need this to be surjective and equidistributed. This can be proven using Waldspurger formula (Zhang, Tian, ...). This kind of argument has been used in this context by Zhang in simpler settings, with some of the new refinements of Waldspurger.

- (2) If you have a map from  $L_f^\times$  to two copies of  $X(U_0)$  that is twisted, say

$$L_f^\times \rightarrow D_f^\times \times D_f^\times \rightarrow X(U_0) \times X(U_0)$$

$$\ell_f \mapsto (\ell_f, \ell_f a_f)$$

where  $a_f$  is some element of  $L_f^\times$  corresponding to some ideal  $\mathfrak{a}_f$  of sufficiently large norm, then you want this map to be surjective and equidistributed as the discriminant  $D_{\mathcal{O}}$  tends to  $\infty$ . This is the mixing conjecture (in this context) of Michel–Venkatesh [1].

Assertion (2) is not yet proved. Over  $\mathbb{Q}$ , Khayutin has proved this. Robinson, in his thesis, and generalized some of this to totally real fields. Unfortunately, we actually need something a bit stronger: for the subgroup of squares  $L_f^{\times 2}$ , and even the subgroup of fourth powers  $L_f^{\times 4}$ .

How does Khayutin do this? This proof depends initially upon a classification of measures of Einsiedler–Lindenstrauss. Basically, the idea is quite  $p$ -adic. Rather than doing it with level  $U_0$ , they allow the level at two primes to go to  $\infty$  (as the discriminant goes to  $\infty$ ) and get some invariance. If you assume that the projections are surjective and equidistributed (which we know from the first method (1)), then the joinings theorem of Einsiedler–Lindenstrauss says that the support must be on some joining. Khayutin finds some nice system of neighborhoods to calculate whether there could be any. Want to know that the measure comes from the full space rather than some Hecke operator. Want to know that no Hecke operator captures too much of the image. In other words, you’re trying to bound the intersection of the Hecke operator with the image of this CM orbit. Khayutin does this by reducing it to a counting argument involving counting numbers of points in a certain Diophantine geometric setup. He manages to bound it using sieve theory, a relative trace formula, and arguments from Diophantine geometry. It’s got a strong  $p$ -adic sense to it, by taking level structures of this kind. It does seem a bit reminiscent of the arguments of Cornut–Vatsal proving Mazur’s conjecture. But the part that needs proving is not in measure theory, but instead in number theory and Diophantine geometry.

[applause]

Assuming this, you get modularity of all compatible systems as in Serre’s conjecture.

**Remark 7.** There is another approach to the mixing conjecture by Blomer–Brumley assuming GRH.

**Remark 8.** The reason we used “split at all finite primes” is that we think this should work for the Artin conjecture in simpler cases, because Artin representations are potentially automorphic – go up to somewhere it’s dihedral. You might think I’m using potential automorphy in this analytic argument, and we are, because it’s simpler, but you can always make that analytic argument over a solvable extension of the base field, using Brauer’s theorem. So we think it may be possible to understand the Artin conjecture from this point of view.

### Part 3. Pol van Hoften (VU Amsterdam), *Igusa stacks and exotic Hecke correspondences*

Let’s say something about exotic Hecke correspondences. This phrase was perhaps coined by Zhu, but it’s a kind of theme in the geometry of Shimura varieties, which is that mod  $p$  Shimura varieties for “related groups” have “related geometry”. Certainly in work of Ribet this already plays a role – his proof of the  $\varepsilon$ -conjecture involves passing because modular curves (such as in the “ $p$ - $q$  switch”). This was taken up by Helm, who considered unitary Shimura varieties. There’s then the

work of Tian–Xiao; in the work of these two people, it was all a bit of an art. The work of Xiao–Zhu turned it into more of a science by giving a rigorous formulation of when these sorts of things should exist, explaining where they should come from.

We would like to start by giving some examples so that we have a feeling for what these things are. As a kind of bonus, the result that we’ll state will apply to these examples.

**Example 9.** Let  $F$  be a real quadratic field, with  $p$  split. To be precise, maybe we should fix an ordering  $\{\infty_1, \infty_2\}$  of the real place, and  $\{\mathfrak{p}_1, \mathfrak{p}_2\}$  for the  $p$ -adic places. Let  $S$  be the mod  $p$  Hilbert modular surface, choosing the level in such a way that it has good reduction. This is some nice smooth surface over  $\mathbb{F}_p$ . We also need our Shimura curves. For  $i = 1, 2$ , consider  $B_i := B_{\infty_i \mathfrak{p}_i}$ . This is the quaternion algebra over  $F$  ramified precisely at  $\infty_i$  and  $\mathfrak{p}_i$ . We denote the corresponding Shimura curve by  $X_{B_i}$ , over  $\mathbb{F}_p$ .

What we’re being vague about is that these Shimura curves are defined over  $F$ , the reflex field, and I’m not telling you what  $p$ -adic places I’m taking. Secretly, we’re fixing an isomorphism

$$(B_i \otimes \mathbb{A}_f^p)^\times \cong \mathrm{GL}_2(\mathbb{A}_f^p),$$

and everything is canonical relative to this choice.

Tian–Xiao tells us that

$$X_{B_i, \overline{\mathbb{F}_p}} \hookrightarrow S_{\overline{\mathbb{F}_p}}$$

as a closed *Goren–Oort stratum*. (This surface has two Goren–Oort strata.) These curves moreover sit inside here Hecke equivariantly.

**Example 10.** Let  $E$  be imaginary quadratic. Let  $W$  be a Hermitian  $E$ -module of rank  $2n$ . Choose some Hermitian  $E$ -module  $W'$  of rank  $2n$ , with  $W \otimes \mathbb{A}_f \cong W' \otimes \mathbb{A}_f$ . So somehow, as Hermitian  $E$ -modules, they’re isomorphic to each other. Then at an inert prime  $p > 2$  in  $E$ , we can consider the Shimura variety for  $X$  over  $\overline{\mathbb{F}_p}$ , as well as the Shimura variety for  $W'$ , and there exists a correspondence (rather than just a map)

$$\begin{array}{ccc} & \mathrm{Sh}(W|W') & \\ t \swarrow & & \searrow s \\ \mathrm{Sh}(W')_{\overline{\mathbb{F}_p}} & & \mathrm{Sh}(W)_{\overline{\mathbb{F}_p}} \end{array}$$

and  $s, t$  have image a closed nonempty union of Newton strata.

Note that if we were in the definite case for  $W'$ , then the corresponding Shimura variety would be a set; the picture thus depends upon the archimedean situation.

There’s thus some sort of art .

**Remark 11.** If  $\mathrm{Sh}(W)_{\overline{\mathbb{F}_p}}$  is a moduli of certain abelian varieties  $A$  with  $E$ -action, and  $\mathrm{Sh}(W')_{\overline{\mathbb{F}_p}}$  is a moduli of certain abelian varieties  $B$ , then

$$\mathrm{Sh}(W|W')_{\overline{\mathbb{F}_p}} = \{p\text{-power } E\text{-linear isogenies } A \rightarrow B\}.$$

Note that these things don’t lift to characteristic zero, i.e., to  $\overline{\mathbb{Q}}$ , due to issues involving the signature when  $W \not\cong W'$ .

This remark doesn’t apply to the first example, Example 9 but you have to be a bit careful because these Shimura varieties are *not* of PEL type. To construct these, you have to put yourself in a situation where, by passing to auxiliary Shimura

varieties, you can do something related to abelian varieties. This theme will also appear in the rest of the talk.

Given that we've identified  $W$  and  $W'$  over the adeles, given  $\alpha$ , we get in particular that

$$\mathrm{GU}(W)_{\mathbb{Q}_p} \cong \mathrm{GU}(W')_{\mathbb{Q}_p} =: G.$$

Newton strata are indexed by the subsets  $B(G, -\mu)$  and  $B(G, -\mu')$  of  $B(G)$ . Here  $\mu$  and  $\mu'$  are cocharacters coming from the Shimura datum, recording the signature. The Newton strata hit by the correspondence  $(s, t)$  are precisely the intersection  $B(G, -\mu) \cap B(G, -\mu')$ .

#### REFERENCES

- [1] Philippe Michel and Akshay Venkatesh. Equidistribution,  $L$ -functions and ergodic theory: on some problems of Yu. Linnik. In *International Congress of Mathematicians. Vol. II*, pages 421–457. Eur. Math. Soc., Zürich, 2006.