New frontiers in Langlands reciprocity

Ana Caraiani

In this survey, I discuss some recent developments at the crossroads of arithmetic geometry and the Langlands programme. The emphasis is on recent progress on the Ramanujan–Petersson and Sato–Tate conjectures. This relies on new results about Shimura varieties and torsion in the cohomology of locally symmetric spaces.

The Langlands programme is a "grand unified theory" of mathematics: a vast network of conjectures that connect number theory to other areas of pure mathematics, such as representation theory, algebraic geometry, and harmonic analysis.

One of the fundamental principles underlying the Langlands conjectures is *reciprocity*, which can be thought of as a magical bridge that connects different mathematical worlds. This principle goes back centuries to the foundational work of Euler, Legendre and Gauss on the law of quadratic reciprocity. A celebrated modern instance of reciprocity is the correspondence between modular forms and rational elliptic curves, which played a key role in Wiles's proof of Fermat's Last Theorem [46] and which relied on the famous Taylor–Wiles method for proving modularity [43]. Recently, the search for new reciprocity laws has begun to expand the scope of the Langlands programme.

The *Ramanujan–Petersson conjecture* is an important consequence of the Langlands programme, which goes back to a prediction Ramanujan made a century ago about the size of the Fourier coefficients of a certain modular form Δ, a highly symmetric function on the upper half plane. The *Sato–Tate conjecture* is an equidistribution result about the number of points of a given elliptic curve modulo varying primes, formulated half a century ago. It is also a consequence of the Langlands programme. In Section 1, I survey progress on these conjectures in two fundamentally different settings: one setting in which there is a direct connection to algebraic geometry (*modular curves*) and one setting in which such a connection is missing (arithmetic hyperbolic 3-manifolds, or *Bianchi manifolds*).

Shimura varieties are certain highly symmetric algebraic varieties that generalise modular curves and that provide, in many cases, a geometric realisation of Langlands reciprocity. In Section 2, I explain a new tool for understanding Shimura varieties called the *Hodge–Tate period morphism*. This was introduced by Scholze in [35] and refined in my joint work with Scholze [16]. I then discuss vanishing theorems for the cohomology of Shimura varieties proved using the geometry of the Hodge–Tate period morphism [16, 17].

The *Calegari–Geraghty method* [11] vastly extends the scope of the Taylor–Wiles method, though it is conjectural on an extension of the Langlands programme to incorporate torsion in the cohomology of locally symmetric spaces. In Section 3, I discuss joint work with Allen, Calegari, Gee, Helm, Le Hung, Newton, Scholze, Taylor, and Thorne [1], where we implement the Calegari–Geraghty method unconditionally over *CM fields*, an important class of number fields that contains imaginary quadratic fields as well as cyclotomic fields. This work relies crucially on one of the vanishing theorems mentioned above [17], and has applications to both the Ramanujan–Petersson and the Sato–Tate conjectures over CM fields.

Remark 1. The Langlands programme is a beautiful but technical subject, with roots in many different areas of mathematics. For a general mathematician, Section 1 is the most accessible, as it highlights two concrete consequences of the Langlands conjectures. The later Sections 2 and 3 assume more background in algebraic geometry and number theory.

I have prioritised references to well-written surveys above references to the original papers. I particularly recommend [21] for a historical account of Langlands reciprocity, [41] for more background on the Langlands correspondence, and [36] for a cutting-edge account of the deep connections between arithmetic geometry and the Langlands programme.

1 The Ramanujan and Sato–Tate conjectures

1.1 Modular curves and Bianchi manifolds

The goal of this section is to discuss two fundamental examples of locally symmetric spaces: *modular curves*, which have an algebraic structure, and *Bianchi manifolds*, which do not. This dichotomy underlies the fundamental difference between reciprocity laws over

EMS MAGAZINE 119 (2021) — DOI 10.4171/MAG-3

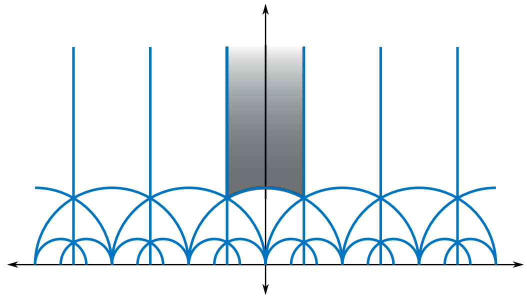


Figure 1. A fundamental domain for $SL_2(\mathbb{Z})$ acting on $\mathbb{H}^2$

the field of rational numbers $\mathbb{Q}$ (and over real quadratic fields such as $\mathbb{Q}(\sqrt{5})$), and reciprocity laws over imaginary quadratic fields such as $\mathbb{Q}(i)$.

Let G be a connected reductive group defined over $\mathbb{Q}$, for example SL_n, GL_n or Sp_{2n}. We can then consider an associated *symmetric space* X, endowed with an action of the real points $G(\mathbb{R})$. This is roughly identified with $G(\mathbb{R})/K_\infty$, where $K_\infty \subset G(\mathbb{R})$ is a maximal compact subgroup. We then want to consider the action of certain arithmetic groups on X: more precisely we want to restrict to finite index subgroups $\Gamma \subset G(\mathbb{Z})$ cut out by congruence conditions. If Γ is sufficiently small, we can form the quotient $\Gamma\backslash X$ and obtain a smooth orientable Riemannian manifold, which is a *locally symmetric space* for G.

Example 2. If $G = SL_2/\mathbb{Q}$, the corresponding symmetric space is the upper-half plane

$$SL_2(\mathbb{R})/SO_2(\mathbb{R}) \simeq \mathbb{H}^2 := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$$

endowed with the hyperbolic metric. The action of $SL_2(\mathbb{R})$ on $\mathbb{H}^2$ is by Möbius transformations:

$$z \mapsto \frac{az+b}{cz+d} \quad \text{for} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}).$$

For $\Gamma \subset SL_2(\mathbb{Z})$ a finite index congruence subgroup (that is assumed sufficiently small), the quotients $\Gamma\backslash\mathbb{H}^2$ are Riemann surfaces. These Riemann surfaces come from algebraic curves X_Γ defined over $\mathbb{Q}$ (or over finite extensions of $\mathbb{Q}$) called *modular curves*. A fundamental domain for a proper subgroup $\Gamma \subset SL_2(\mathbb{Z})$ acting on $\mathbb{H}^2$ is a finite union of translates of the fundamental domain in Figure 1.

Example 3. If $G = SL_2/F$, where F is an imaginary quadratic field[1], the corresponding symmetric space is 3-dimensional hyperbolic space

$$SL_2(\mathbb{C})/SU_2(\mathbb{R}) \simeq \mathbb{H}^3$$

and the locally symmetric spaces are called *Bianchi manifolds*. These are arithmetic hyperbolic 3-manifolds and, since their real dimension is odd, they do not admit a complex or algebraic structure.

The locally symmetric spaces for a group G are important in what follows because they give a way to access *automorphic representations* of G, the central objects of study in the Langlands programme. This is explained more in Section 2. For example, *modular forms*[2], which are holomorphic functions on $\mathbb{H}^2$ that satisfy a transformation relation under some Γ, contribute to the first Betti cohomology of modular curves (with possibly twisted coefficients).

Some locally symmetric spaces have an algebraic structure. If this happens, they in fact come from smooth, quasi-projective varieties X_Γ defined over number fields, which are called *Shimura varieties*. The geometry of Shimura varieties is a rich and fascinating subject in itself, that we discuss more in Section 2. On the other hand, the Langlands programme is much more mysterious beyond the setting of Shimura varieties, because there is no obvious connection to algebraic geometry or arithmetic. We discuss this more in Section 3.

1.2 The Ramanujan conjecture

A famous example of a modular form is Ramanujan's Δ function. If z is the variable on the upper-half plane $\mathbb{H}^2$ and $q = e^{2\pi iz}$, Δ is given by the Fourier series expansion

$$\Delta(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24} = \sum_{n>0}\tau(n)q^n.$$

[1] This can be viewed as a connected reductive group over $\mathbb{Q}$ using a technical notion called the Weil restriction of scalars.

[2] These give rise to automorphic representations for the group $SL_2/\mathbb{Q}$.

In 1916, Ramanujan made three predictions about the behaviour of the Fourier coefficients $\tau(n)$. The first two of these were immediately proved by Mordell by studying the action on Δ of certain *Hecke operators*, that we return to in Section 2. The *Ramanujan conjecture*, which resisted attempts at proof for much longer, bounds the absolute value of the Fourier coefficients: it states that $|\tau(p)| \leq 2p^{11/2}$ for all primes p.

Deligne finally established this bound in the early 1970's, and this was one of the reasons for which he was awarded a Fields Medal in 1978. While the bound on the Fourier coefficients is purely a statement within harmonic analysis, the proof used the bridge of Langlands reciprocity and was ultimately obtained from a statement in arithmetic geometry. More precisely, Deligne's proof of the Ramanujan conjecture went via the étale cohomology of modular curves, obtaining the desired bound as a consequence of his proof of the Weil conjectures for smooth projective varieties over finite fields.

The generalised *Ramanujan–Petersson conjecture* is a vast extension of the above statement, with numerous applications across mathematics and computer science. See, for example, the survey [31] for its applications to extremal combinatorial objects called Ramanujan graphs. This more general conjecture, which is part of Arthur's conjectures on the automorphic spectrum of GL_n (see also the survey [34]), predicts that the local components at finite places of cuspidal automorphic representations of GL_n are *tempered*.

Temperedness means roughly that the matrix coefficients of the representation are in $L^{2+\epsilon}$ for all $\epsilon > 0$. This singles out the building blocks of the category of irreducible admissible representations of p-adic groups, such as $\mathrm{GL}_n(\mathbb{Q}_p)$, in the sense that everything else can be constructed from tempered representations of smaller groups. Tempered representations also play an important role in the *local Langlands conjecture*, which relates them to arithmetic objects, essentially representations of local Galois groups. For the group GL_n, local Langlands is a theorem, proved by Harris–Taylor and Henniart in the early 2000's, and later reproved by Scholze.

For certain cuspidal automorphic representations of GL_n, which are global objects built from the irreducible admissible representations mentioned above, one can try to follow Deligne's approach to the Ramanujan conjecture using the étale cohomology of higher-dimensional Shimura varieties. When these varieties have singular reduction, the arithmetic counterpart of the Ramanujan–Petersson conjecture is Deligne's *weight-monodromy conjecture*. This goes beyond the Weil conjectures to predict that the étale cohomology of smooth projective varieties over p-adic fields has a remarkably elegant shape, even in the case of singular reduction.

In [12], building on [18, 39, 44] and [25], I follow Deligne's approach and complete the proof of the following result.

Theorem 4. *Let F be a CM field and let π be a regular algebraic, self-dual cuspidal automorphic representation of GL_n/F. Then π satisfies the generalised Ramanujan–Petersson conjecture.*

The global Langlands correspondence relates automorphic representations to global Galois representations. The direction from automorphic to Galois is best understood in the setting of Theorem 4, which is the so-called "self-dual case". This has been a milestone achievement in the field: it required the combined effort of many people over several decades, including Kottwitz, Clozel, Harris, Taylor, Shin, and Chenevier, and was built on fundamental contributions by Arthur, Laumon, Ngô and Waldspurger. In [12, 13], I also complete the proof that the associated Galois representations are compatible with local Langlands[3], by establishing new instances of the weight-monodromy conjecture for Shimura varieties.

More recently, in joint work with Allen, Calegari, Gee, Helm, Le Hung, Newton, Scholze, Taylor, and Thorne, I obtained an application to the Ramanujan–Petersson conjecture beyond the self-dual case. This is the first instance where this conjecture is not deduced from the Weil conjectures, but rather by an approximation of the very different strategy outlined by Langlands in [30].

Theorem 5 ([1]). *Let F be a CM field and π be a cuspidal automorphic representation of GL_2/F of parallel weight 2. Then π satisfies the generalised Ramanujan–Petersson conjecture.*

The condition on the weight means that π contributes to the Betti cohomology with constant coefficients of the relevant locally symmetric space, which is for example a Bianchi manifold. These locally symmetric spaces do not have an algebraic structure, so one cannot appeal directly to arithmetic geometry. We come back to discuss the strategy for the proof of Theorem 5 in Section 3.

1.3 The Sato–Tate conjecture

An *elliptic curve* is a smooth, projective curve of genus one together with a specified point. If F is a number field, an elliptic curve defined over F can be described as a plane curve, given by (the homogenisation of) a cubic equation of the form $y^2 = x^3 + ax + b$ with $a, b \in F$.

Such an elliptic curve E/F, if it does not have complex multiplication, is expected to satisfy the *Sato–Tate conjecture*. When $\mathfrak{p}$ is a prime of F over which E has good reduction, the number

$$\frac{1 + q_{\mathfrak{p}} - \#E\big(k(\mathfrak{p})\big)}{2\sqrt{q_{\mathfrak{p}}}}$$

[3] Local-global compatibility is a crucial property one expects from the Langlands correspondence, which generalises the compatibility between local and global class field theory.

(where $k(\mathfrak{p})$ denotes the residue field at $\mathfrak{p}$, of cardinality $q_\mathfrak{p}$) is contained in the interval $[-1,1]$ by a result of Hasse; this is also a special case of Deligne's result on the Weil conjectures. The Sato–Tate conjecture, formulated in the 1960's, states that, as $\mathfrak{p}$ runs over all the primes of F over which E has good reduction, these numbers become equidistributed in $[-1,1]$ with respect to the semicircle probability measure $\frac{2}{\pi}\sqrt{1-x^2}dx$.

Remark 6. The condition for an elliptic curve to have complex multiplication is very special, and in that case the probability distribution is different and well-understood. See [40] for a survey on Sato–Tate-type conjectures, which explains the expected distributions, and [26] for the more general conceptual framework that underlies this conjecture.

According to the Langlands reciprocity conjecture, any elliptic curve E/F is also expected to come from an automorphic representation of GL_2 over F. If this is the case, we say that E is *automorphic*. The precise relationship between elliptic curves and automorphic representations can be expressed as an equality of the two *L-functions* associated to them. L-functions are complex analytic functions that generalise the Riemann zeta function and that remember deep arithmetic information about the original objects.

For example, the L-functions of all elliptic curves defined over $\mathbb{Q}$ are known to come from modular forms, by work of Breuil–Conrad–Diamond–Taylor [9] building on [46] and [43]. The analogous result for elliptic curves defined over real quadratic fields was later proved by Freitas–Le Hung–Siksek [22]. The L-functions of elliptic curves over imaginary quadratic fields are expected to come from classes in the cohomology of Bianchi manifolds, but this case has historically been much more mysterious.

Soon after the Sato–Tate conjecture was formulated, Serre and Tate discovered that the correct distribution would follow from the expected analytic properties of the symmetric power L-functions of E. In turn, these analytic properties would follow if one knew the automorphy of E and all its symmetric powers. This argument is explained in [37] and uses Tauberian theorems in analytic number theory: the techniques are essentially those that led to the proof of the prime number theorem. In fact, to establish the correct distribution, it suffices to know that E and its symmetric powers are *potentially automorphic*: this means they become automorphic after base change to some Galois field extension F' of F.

The Sato–Tate conjecture for elliptic curves defined over totally real fields was proved in most cases by Clozel, Harris, Shepherd-Barron, and Taylor [19, 24, 42], and completed in work of Barnet-Lamb–Geraghty–Harris–Taylor around 2010 [4]. This relied on the potential automorphy of symmetric powers, which could be established in the self-dual setting using a generalisation of the Taylor–Wiles method. However, the method broke down for elliptic curves defined over imaginary quadratic fields or more general CM fields. In Section 3, we explain how to overcome the barrier to treating elliptic curves defined over CM fields and obtain the following result.

Theorem 7 ([1]). *Let F be a CM field and E/F be an elliptic curve that does not have complex multiplication. Then E is potentially automorphic and satisfies the Sato–Tate conjecture.*

Remark 8. Both Theorems 5 and 7 rely crucially on the vanishing theorem for Shimura varieties proved in [17], which is discussed in Section 2.

Remark 9. The beautiful work of Boxer–Calegari–Gee–Pilloni [7], completed at the same time as [1], proves the potential automorphy of elliptic curves in Theorem 7 independently, and they are even able to show the potential automorphy of abelian surfaces over totally real fields. Moreover, in the recent paper [2], Allen–Khare–Thorne establish actual automorphy of elliptic curves in certain cases (rather than potential automorphy). All of this is hopefully only the beginning of a fascinating story over CM fields!

2 Vanishing theorems for Shimura varieties with torsion coefficients

2.1 Shimura varieties

Recall that, if the locally symmetric spaces for a group $G/\mathbb{Q}$ have an algebraic structure, they in fact come from smooth, quasi-projective varieties X_Γ defined over number fields, which are called Shimura varieties.

The pair (G,X) must satisfy certain axioms in order for the corresponding locally symmetric spaces to come from Shimura varieties. The key point is for the symmetric space X to be a Hermitian symmetric domain (or a finite disjoint union thereof). There is a complete classification of groups G for which this holds. For example, the symplectic group Sp_{2n} and the unitary group $U(n,n)$ give rise to Shimura varieties, which can be described in terms of moduli spaces of abelian varieties equipped with additional structures.

Remark 10. Some locally symmetric spaces that are not Shimura varieties can still be studied by relating them to Shimura varieties. For example, Bianchi manifolds can be realised in the boundary of certain compactifications of Shimura varieties attached to the unitary group $U(2,2)$. We come back to this in Section 3.

Recall also that the locally symmetric spaces for a group G give a way to access automorphic representations of G. More precisely, as the congruence subgroup $\Gamma \subset G(\mathbb{Z})$ varies, we have a tower of locally symmetric spaces. The symmetries of this tower induce correspondences on each individual space $\Gamma\backslash X$ called Hecke operators[4]. Keeping track of the various Hecke operators, we obtain an action of a commutative Hecke algebra $\mathbb{T}$ on the Betti cohomology $H^i(\Gamma\backslash X, \mathbb{C})$. The work of Matsushima, Franke and others shows that the systems of eigenvalues of $\mathbb{T}$ that occur in $H^i(\Gamma\backslash X, \mathbb{C})$ come from certain automorphic representations of G.

In addition to the Hecke symmetry, the cohomology of Shimura varieties also has a Galois symmetry, because Shimura varieties are defined over number fields. Because of these two kinds of symmetries, Shimura varieties give, in many cases, a geometric realisation of the global Langlands correspondence between automorphic and Galois representations.

One can ask a more precise question, about the range of degrees of cohomology to which any particular automorphic representation can contribute. Assume, for simplicity, that $X_\Gamma(\mathbb{C})$ is a compact Shimura variety. Then Borel–Wallach [6] show that, if π is an automorphic representation whose component at ∞ is a tempered representation of $G(\mathbb{R})$, then π can only contribute to $H^i(X_\Gamma(\mathbb{C}), \mathbb{C})$ in the middle degree $i = \dim_{\mathbb{C}} X_\Gamma$. This result, like the Ramanujan–Petersson conjecture, also fits within the framework of Arthur's conjectures [3].

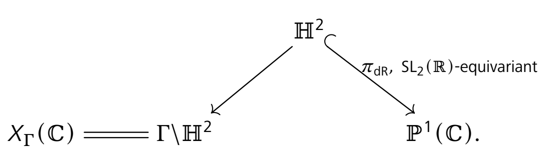
Question 11. The upshot of the Borel–Wallach result is that the cohomology of a Shimura variety X_Γ with $\mathbb{C}$-coefficients is somehow degenerate outside the middle degree. Can we extend this to torsion coefficients, such as $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)$?

More precise versions of this question are formulated as conjectures in [11] and [20]. These are motivated by the Calegari–Geraghty method, which is discussed in Section 3, and by the search for a mod ℓ analogue of Arthur's conjectures. In the next two subsections, we explain a new tool that can be used to compute $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)$ and discuss our results towards Question 11.

2.2 The Hodge–Tate period morphism

This morphism was introduced by Scholze in his breakthrough paper [35] and gives a completely new way to access the geometry and cohomology of Shimura varieties.

In the case of the modular curve, the Hodge–Tate period morphism is a p-adic analogue of the following complex picture, where the map on the right is the standard holomorphic embedding of the upper-half plane $\mathbb{H}^2$ into the Riemann sphere $\mathbb{P}^1(\mathbb{C})$:



This picture has the following moduli interpretation. First, X_Γ is a moduli space of elliptic curves equipped with some additional structures (determined by Γ). The upper-half plane $\mathbb{H}^2$ is the universal cover of $X_\Gamma(\mathbb{C}) = \Gamma\backslash\mathbb{H}^2$; it parametrises (positive) complex structures one can put on a two-dimensional real vector space. This amounts to parameterising Hodge structures of elliptic curves, i.e. direct sum decompositions:

$$\mathbb{C}^2 = H^1(E(\mathbb{C}), \mathbb{C}) \simeq H^0(E, \Omega^1_E) \oplus H^1(E, \mathcal{O}_E)$$

with $H^1(E, \mathcal{O}_E) = \overline{H^0(E, \Omega^1_E)}$. The morphism π_{dR} sends the Hodge decomposition to the associated Hodge filtration

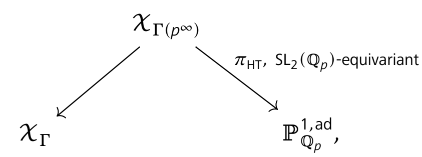
$$H^0(E, \Omega^1_E) \subset H^1(E(\mathbb{C}), \mathbb{C}) = \mathbb{C}^2.$$

This is an example of a *period morphism*. One can construct such a diagram for higher-dimensional Shimura varieties as well, and this has played an important role in studying automorphic forms on Shimura varieties from a geometric point of view.

The Hodge–Tate period morphism is based on the Hodge–Tate filtration on étale cohomology, tracing back to foundational work in p-adic Hodge theory by Tate and Faltings. Let p be a prime and let C be the p-adic completion of an algebraic closure of $\mathbb{Q}_p$, which will play a role analogous to that of $\mathbb{C}$ in what follows. If E/C is an elliptic curve, its étale cohomology admits a Hodge–Tate filtration:

$$0 \to H^1(E, \mathcal{O}_E) \to H^1_{et}(E, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} C \to H^0(E, \Omega^1_E)(-1) \to 0.$$

See Bhatt's article in [5] for an excellent survey on p-adic Hodge theory and more details on the Hodge–Tate filtration. Instead of viewing the curve X_Γ as a Riemann surface, we view it as an adic space $\mathcal{X}_\Gamma$, a kind of p-adic analytic space introduced by Huber. Then there exists a diagram



where $\mathcal{X}_\Gamma(p^\infty)$, which is roughly the inverse limit of modular curves $\mathcal{X}_{\Gamma(p^n)}$ with increasing level at p, is a perfectoid space. Over a point of $\mathcal{X}_{\Gamma(p^\infty)}$ corresponding to an elliptic curve E/C, we have a trivialisation of $H^1_{et}(E, \mathbb{Z}_p) \simeq \mathbb{Z}_p^2$. This point gets sent under π_{HT} to the line

$$H^1(E, \mathcal{O}_E) \subset H^1_{et}(E, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} C \simeq C^2.$$

For higher-dimensional Shimura varieties, the following result describes the geometry of the Hodge–Tate period morphism in detail. While the statement of Theorem 12 involves much non-trivial arithmetic geometry, it has applications to Theorems 16 and 17 below, whose statements are substantially more elementary.

[4] To discuss Hecke operators rigorously, we should use the adelic perspective on locally symmetric spaces and Shimura varieties. The resulting spaces would be disjoint unions of finitely many copies of $\Gamma\backslash X$. We ignore this subtlety here and later on in the text.

Theorem 12 ([35, 16]). *Let X_Γ be a Shimura variety of Hodge type associated to a connected reductive group G. Let μ denote the conjugacy class of Hodge cocharacters and let $\mathcal{F}\ell_{G,\mu} := G/P_\mu$ denote the corresponding flag variety, considered as an adic space over a p-adic completion of the reflex field.*

1. *There exists a unique perfectoid space $\mathcal{X}_{\Gamma(p^\infty)}$ which can be identified with the inverse limit of the adic spaces $(\mathcal{X}_{\Gamma(p^n)})_n$.*
2. *There exists a Hodge–Tate period morphism*
$$\pi_{\mathrm{HT}} : \mathcal{X}_{\Gamma(p^\infty)} \to \mathcal{F}\ell_{G,\mu},$$
which is $G(\mathbb{Q}_p)$-equivariant.
3. *There exists a Newton stratification*
$$\mathcal{F}\ell_{G,\mu} = \bigsqcup_{b \in B(G,\mu)} \mathcal{F}\ell^b_{G,\mu}$$
into locally closed strata.
4. *If $\mathcal{X}_\Gamma$ is compact and of PEL type, and $\bar{x}$ is a geometric point of the Newton stratum $\mathcal{F}\ell^b_{G,\mu}$, we identify the fiber $\pi^{-1}_{\mathrm{HT}}(\bar{x})$ with a "perfectoid" version of an Igusa variety* Ig^b.

Remark 13. The first two parts of Theorem 12 are due to Scholze[5] and play the lead role in his breakthrough construction of Galois representations for torsion in the cohomology of locally symmetric spaces. There are many surveys of this result; see for example [33] or [45]. For more details on the Hodge–Tate period morphism, see also the last article in [5].

Remark 14. Igusa varieties were introduced by Harris–Taylor as part of their proof of local Langlands for GL_n, and generalised by Mantovan. Rapoport–Zink spaces are local analogues of Shimura varieties, which provide a geometric realisation of the local Langlands correspondence. The computation of the fibers of π_{HT} suffices for applications to Theorems 16 and 17 below, but in [16], we go further and prove a conceptually cleaner version of Mantovan's product formula [32], which relates Shimura varieties, Igusa varieties and Rapoport–Zink spaces.

Remark 15. In [17] we extend part (4) of Theorem 12 to $\mathrm{U}(n,n)$-Shimura varieties, which are non-compact. We compute the fibers of π_{HT} for both the minimal and toroidal compactifications of these Shimura varieties, and relate them to partial minimal and toroidal compactifications of Igusa varieties.

2.3 Vanishing theorems

In order to address Question 11, we would like to compute the localisation $H^*(\mathcal{X}_\Gamma, \mathbb{F}_\ell)_\mathfrak{m}$, where the maximal ideal $\mathfrak{m} \subset \mathbb{T}$ is equivalent to a mod ℓ system of Hecke eigenvalues. Using the Hodge–Tate period morphism at an auxiliary prime $p \neq \ell$[6], we obtain an action of $\mathbb{T}$ on the complex of sheaves $R\pi_{\mathrm{HT}*}\mathbb{F}_\ell$ living over $\mathcal{F}\ell_{G,\mu}$, and we are reduced to understanding the localisation $(R\pi_{\mathrm{HT}*}\mathbb{F}_\ell)_\mathfrak{m}$. By the properties of π_{HT}, this behaves similarly to a perverse sheaf, which is the key to controlling the degrees in which $(R\pi_{\mathrm{HT}*}\mathbb{F}_\ell)_\mathfrak{m}$ can have non-zero cohomology. We make these ideas rigorous in [16, 17] for unitary Shimura varieties, under some mild technical assumptions.

Let $F = F^+ \cdot E$ be a CM field, with maximal totally real field $F^+ \neq \mathbb{Q}$ and E imaginary quadratic. Let G be a unitary group preserving a skew-Hermitian form on F^m. Assume that G is quasi-split at all finite places. Let $\mathfrak{m} \subset \mathbb{T}$ be a system of Hecke eigenvalues that occurs in $H^i(X_\Gamma, \mathbb{F}_\ell)$. Assume $\mathfrak{m}$ is *generic* at an auxiliary prime $p \neq \ell$[7]. This condition guarantees that all lifts of $\mathfrak{m}$ to characteristic 0 are as simple as possible at p, from a representation-theoretic point of view: they are generic principal series representations of $G(\mathbb{Q}_p)$.

Theorem 16 ([16]). *If $\mathcal{X}_\Gamma$ is compact and $\mathfrak{m}$ is generic, then $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in the middle degree $i = \dim_\mathbb{C} X_\Gamma$.*

In the non-compact case, genericity, which is a local condition at an auxiliary prime $p \neq \ell$, is not enough. We also need a global condition to control the boundary of the Shimura variety. To formulate the global condition, we consider the semi-simple Galois representation $\bar{\rho}_\mathfrak{m}$ associated to the system of eigenvalues $\mathfrak{m}$ by [35]; the existence of $\bar{\rho}_\mathfrak{m}$ is an instance of the global Langlands correspondence in the torsion setting. We want to assume that $\bar{\rho}_\mathfrak{m}$ is not too degenerate; this amounts to bounding the number of its absolutely irreducible factors.

Theorem 17 ([17]). *If X_Γ is a $\mathrm{U}(n,n)$-Shimura variety (so m is even and G is quasi-split at the infinite places as well), $\mathfrak{m}$ is generic, and $\bar{\rho}_\mathfrak{m}$ has at most two absolutely irreducible factors, then:*

1. *$H^i_c(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in degrees $i \leq \dim_\mathbb{C} X_\Gamma$, and*
2. *$H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in degrees $i \geq \dim_\mathbb{C} X_\Gamma$.*

Remark 18. There are previous results in this direction, due to Dimitrov, Shin, Emerton–Gee, and especially Lan–Suh [28, 29]. Compared to previous work, our result is sharper and better adapted to applications. There is also intriguing ongoing work of Boyer [8], which proves a stronger result in the special case of Harris–Taylor Shimura varieties: he goes beyond genericity and investigates the distribution of non-generic systems of Hecke eigenvalues.

[5] Up to the precise identification of the target of the Hodge–Tate period morphism as the flag variety $\mathcal{F}\ell_{G,\mu}$ in all cases, which is done in [16].
[6] Here, we assume that the Hecke operators in $\mathbb{T}$ are all supported at primes different from p.
[7] See [17, Theorem 1.1] for the precise condition, which is technical, but explicit. This condition should be thought of as a mod ℓ analogue of temperedness.

Remark 19. The idea of the proof in the compact case is the following: start with a top-dimensional Newton stratum $\mathcal{F}\ell_{G,\mu}^b \subset \mathcal{F}\ell_{G,\mu}$ in the support of $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$. Since the complex $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$ behaves like a perverse sheaf, its restriction to $\mathcal{F}\ell_{G,\mu}^b$ is concentrated in one degree. Therefore, $(R\pi_{\mathrm{HT},*}\mathbb{Q}_\ell)_\mathfrak{m}$ is also concentrated in one degree over $\mathcal{F}\ell_{G,\mu}^b$. On the other hand, we can compute the alternating sum of cohomology groups of Ig^b with $\mathbb{Q}_\ell$-coefficients, using the trace formula and work of Shin [38]. In the end, the genericity condition is contradicted unless b corresponds to the zero-dimensional ordinary stratum. The upshot is that $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in one degree over a zero-dimensional stratum!

Remark 20. In parallel to Question 11, one can also study the cohomology of locally symmetric spaces with torsion coefficients and with increasing level at p. The resulting structure is called *completed cohomology* and was introduced by Emerton as a general framework for studying congruences modulo p^k between automorphic forms. Motivated by heuristics coming from the p-adic Langlands programme, Calegari–Emerton [10] formulated a vanishing conjecture for completed cohomology. For most Shimura varieties, the Calegari–Emerton conjecture is now a theorem due to Scholze and Hansen–Johansson.

In [14, 15], we prove a vanishing result for the compactly supported cohomology of Shimura varieties of Hodge type with unipotent level at p. The only assumption is that the group G giving rise to the Shimura variety is split over $\mathbb{Q}_p$. This result is stronger than what Calegari–Emerton conjectured, and it also points towards analogues of Theorems 16 and 17 for $\ell = p$, with *generic* replaced by *ordinary* in the sense of Hida.

3 Potential automorphy over CM fields

Theorem 5 on the Ramanujan–Petersson conjecture and Theorem 7 on the Sato–Tate conjecture would follow if we knew that all the symmetric powers of the associated Galois representations were automorphic, or even just potentially automorphic. The original method developed by Taylor–Wiles is a powerful technique for proving automorphy, but it is restricted to settings where a certain numerical criterion holds: these are roughly the settings where the objects on the automorphic side arise from the middle degree cohomology of a Shimura variety.

When F is a number field, the locally symmetric spaces for GL_n/F, such as the Bianchi manifolds discussed in Example 3, do not have an algebraic structure (outside very special cases). Calegari–Geraghty [11] proposed an extension of the Taylor–Wiles method to general number fields F, conjectural on a precise understanding of the cohomology of locally symmetric spaces for GL_n/F. Part of their insight was to realise the central role played by torsion classes in the cohomology of these locally symmetric spaces, which should be thought of as modulo p^k versions of automorphic forms and treated on equal footing with their characteristic 0 counterparts. Another part of their insight was to reinterpret the failure of the Taylor–Wiles numerical criterion in terms of certain non-negative integers q_0, l_0 seen on the automorphic side.

The Calegari–Geraghty method gives an automorphy lifting result for GL_n/F as long as the following prerequisites are in place:

1. The construction of Galois representations associated to classes in the cohomology with $\mathbb{Z}_p$ coefficients of the locally symmetric spaces for GL_n/F.
2. Local-global compatibility for these Galois representations at all primes of F, including at primes above p.
3. A vanishing conjecture for the cohomology with $\mathbb{Z}_p$ coefficients outside the range of degrees $[q_0, q_0+l_0]$, under an appropriate non-degeneracy condition.

Remark 21. For Shimura varieties, the third problem is closely related to Theorems 16 and 17, since in that case q_0 is the middle degree of cohomology and $l_0 = 0$. For 3-dimensional Bianchi manifolds, the third problem says that the non-degenerate part of cohomology is concentrated in degrees 1 and 2; this can be checked by hand. For general locally symmetric spaces that do not have an algebraic structure, this problem most likely lies deeper than the first two.

When F is a CM field, the first problem was solved by Scholze in [35], strengthening previous results of Harris–Lan–Taylor–Thorne [23] for characteristic 0 coefficients. After completing [16], it became clear to Scholze and me that a non-compact version of Theorem 16 would give a strategy to attack the second (rather than the third!) problem over CM fields. In joint work with Scholze, I set out to prove Theorem 17 and, in November 2016, I co-organised with Taylor an "emerging topics" working group at the IAS, whose goal was to explore this strategy and its consequences. The working group was a resounding success and it led to the paper [1], where we implement the Calegari–Geraghty method in arbitrary dimension for the first time and obtain as consequences Theorems 5 and 7.

The solution to the first problem above, i.e., the construction of Galois representations, is much more subtle than in the self-dual case, because one cannot directly use the étale cohomology of Shimura varieties. Instead, the starting point for both [23] and [35] is to realise the locally symmetric spaces for GL_n/F in the boundary of the *Borel–Serre compactification* of $\mathrm{U}(n,n)$-Shimura varieties. The Borel–Serre compactification is a real manifold with corners, which is homotopy equivalent to the original $\mathrm{U}(n,n)$-Shimura variety. In the torsion setting, Scholze constructs the desired Galois representations by congruences, using the Hodge–Tate period morphism for the $\mathrm{U}(n,n)$-Shimura variety. This increases the level at primes of F dividing p, and makes the second problem, local-global compatibility, particularly tricky at these primes.

In [1], we begin to solve the second problem, by establishing the first instances of local-global compatibility at primes of F dividing p. We need a delicate argument to understand the boundary of the Borel–Serre compactification, which combines algebraic topology and modular representation theory. In addition, Theorem 17 is the crucial new ingredient: in the middle degree, it implies that classes from the boundary lift to the cohomology of a $\mathrm{U}(n,n)$-Shimura variety with $\mathbb{Q}_p$-coefficients, while remembering the level and weight at primes of F dividing p.

The proofs of Theorems 5 and 7 use the Calegari–Geraghty method, together with solutions to the first two problems discussed above. The third problem was not solved with $\mathbb{Z}_p$ coefficients. By an insight of Khare–Thorne [27], this problem could be replaced by its $\mathbb{Q}_p$ coefficient analogue in certain settings. One of the main challenges in [1] was to make this insight compatible with other techniques in automorphy lifting, which rely on reduction modulo p. We resolve this challenge by considering reduction modulo p from a derived perspective. Outside low-dimensional cases, such as Bianchi manifolds, or Shimura varieties, the third problem remains open for $\mathbb{Z}_p$ coefficients.

Acknowledgements. This article was written in relation to my being awarded one of the 2020 Prizes of the European Mathematical Society. I wish to dedicate this article to my father, Cornel Caraiani (1954–2020), who inspired my love of mathematics.

I have been lucky to have many wonderful mentors and collaborators and I am grateful to all of them for the mathematics they have taught me. In addition, I especially want to thank Matthew Emerton, Toby Gee, Sophie Morel, James Newton, Peter Scholze, and Richard Taylor for generously sharing their ideas with me over the years, and for their substantial moral and professional support.

I am also grateful to Toby Gee, James Newton, Steven Sivek, and Matteo Tamiozzo for comments on an earlier version of this article.

References

[1] P. B. Allen, F. Calegari, A. Caraiani, T. Gee, D. Helm, B. V. Le Hung, J. Newton, P. Scholze, R. Taylor, and J. A. Thorne. Potential automorphy over CM fields. arXiv:1812.09999 (December 2018)

[2] P. B. B. Allen, C. Khare, and J. A. Thorne. Modularity of $\mathrm{GL}_2(\mathbb{F}_p)$-representations over CM fields. arXiv:1910.12986 (October 2019)

[3] J. Arthur, Unipotent automorphic representations: Conjectures. 171–172, 13–71 (1989)

[4] T. Barnet-Lamb, D. Geraghty, M. Harris and R. Taylor, A family of Calabi-Yau varieties and potential automorphy II. *Publ. Res. Inst. Math. Sci.* **47**, 29–98 (2011)

[5] B. Bhatt, A. Caraiani, K. S. Kedlaya and J. Weinstein, *Perfectoid spaces*. Mathematical Surveys and Monographs 242, American Mathematical Society, Providence, RI (2019)

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

[7] G. Boxer, F. Calegari, T. Gee, and V. Pilloni. Abelian Surfaces over totally real fields are Potentially Modular. arXiv:1812.09269 (December 2018)

[8] P. Boyer, Sur la torsion dans la cohomologie des variétés de Shimura de Kottwitz–Harris–Taylor. *J. Inst. Math. Jussieu* **18**, 499–517 (2019)

[9] C. Breuil, B. Conrad, F. Diamond and R. Taylor, On the modularity of elliptic curves over $\mathbf{Q}$: Wild 3-adic exercises. *J. Amer. Math. Soc.* **14**, 843–939 (2001)

[10] F. Calegari and M. Emerton, Completed cohomology – A survey. In *Non-abelian fundamental groups and Iwasawa theory*, London Math. Soc. Lecture Note Ser. 393, Cambridge Univ. Press, Cambridge, 239–257 (2012)

[11] F. Calegari and D. Geraghty, Modularity lifting beyond the Taylor–Wiles method. *Invent. Math.* **211**, 297–433 (2018)

[12] A. Caraiani, Local-global compatibility and the action of monodromy on nearby cycles. *Duke Math. J.* **161**, 2311–2413 (2012)

[13] A. Caraiani, Monodromy and local-global compatibility for $l = p$. *Algebra Number Theory* **8**, 1597–1646 (2014)

[14] A. Caraiani, D. R. Gulotta, C.-Y. Hsu, C. Johansson, L. Mocz, E. Reinecke and S.-C. Shih, Shimura varieties at level $\Gamma_1(p^\infty)$ and Galois representations. *Compos. Math.* **156**, 1152–1230 (2020)

[15] A. Caraiani, D. R. Gulotta, and C. Johansson. Vanishing theorems for Shimura varieties at unipotent level. arXiv:1910.09214 (October 2019)

[16] A. Caraiani and P. Scholze, On the generic part of the cohomology of compact unitary Shimura varieties. *Ann. of Math. (2)* **186**, 649–766 (2017)

[17] A. Caraiani and P. Scholze, On the generic part of the cohomology of compact unitary Shimura varieties. arXiv:1909.01898 (September 2019)

[18] L. Clozel, Purity reigns supreme. *Int. Math. Res. Not. IMRN* 328–346 (2013)

[19] L. Clozel, M. Harris and R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations. *Publ. Math. Inst. Hautes Études Sci.* 1–181 (2008)

[20] M. Emerton, Completed cohomology and the p-adic Langlands program. In *Proceedings of the International Congress of Mathematicians – Seoul 2014. Vol. II*, Kyung Moon Sa, Seoul, 319–342 (2014)

[21] M. Emerton. Langlands reciprocity: L-functions, automorphic forms, and Diophantine equations. To appear in *The Genesis of the Langlands program* (2020)

[22] N. Freitas, B. V. Le Hung and S. Siksek, Elliptic curves over real quadratic fields are modular. *Invent. Math.* **201**, 159–206 (2015)

[23] M. Harris, K.-W. Lan, R. Taylor and J. Thorne, On the rigid cohomology of certain Shimura varieties. *Res. Math. Sci.* **3**, Paper No. 37, 308 (2016)

[24] M. Harris, N. Shepherd-Barron and R. Taylor, A family of Calabi-Yau varieties and potential automorphy. *Ann. of Math. (2)* **171**, 779–813 (2010)

[25] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*. Annals of Mathematics Studies 151, Princeton University Press, Princeton, NJ (2001)

[26] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*. American Mathematical Society Colloquium Publications 45, American Mathematical Society, Providence, RI (1999)

[27] C. B. Khare and J. A. Thorne, Potential automorphy and the Leopoldt conjecture. *Amer. J. Math.* **139**, 1205–1273 (2017)

[28] K.-W. Lan and J. Suh, Vanishing theorems for torsion automorphic sheaves on compact PEL-type Shimura varieties. *Duke Math. J.* **161**, 1113–1170 (2012)

[29] K.-W. Lan and J. Suh, Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties. *Adv. Math.* **242**, 228–286 (2013)

[30] R. P. Langlands, Problems in the theory of automorphic forms. In *Lectures in modern analysis and applications, III*, 18–61. Lecture Notes in Math., Vol. 170 (1970)

[31] W.-C. W. Li, The Ramanujan conjecture and its applications. *Philos. Trans. Roy. Soc. A* **378**, 20180441, 14 (2020)

[32] E. Mantovan, On the cohomology of certain PEL-type Shimura varieties. *Duke Math. J.* **129**, 573–610 (2005)

[33] S. Morel, Construction de représentations galoisiennes de torsion [d'après Peter Scholze]. *Astérisque* Exp. No. 1102, 449–473 (2016)

[34] P. Sarnak, Notes on the generalized Ramanujan conjectures. In *Harmonic analysis, the trace formula, and Shimura varieties*, Clay Math. Proc. 4, Amer. Math. Soc., Providence, RI, 659–685 (2005)

[35] P. Scholze, On torsion in the cohomology of locally symmetric varieties. *Ann. of Math. (2)* **182**, 945–1066 (2015)

[36] P. Scholze, *p*-adic geometry. In *Proceedings of the International Congress of Mathematicians – Rio de Janeiro 2018. Vol. I. Plenary lectures*, World Sci. Publ., Hackensack, NJ, 899–933 (2018)

[37] J.-P. Serre, *Abelian l-adic representations and elliptic curves*. McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam (1968)

[38] S. W. Shin, A stable trace formula for Igusa varieties. *J. Inst. Math. Jussieu* **9**, 847–895 (2010)

[39] S. W. Shin, Galois representations arising from some compact Shimura varieties. *Ann. of Math. (2)* **173**, 1645–1741 (2011)

[40] A. V. Sutherland, Sato–Tate distributions. In *Analytic methods in arithmetic geometry*, Contemp. Math. 740, Amer. Math. Soc., Providence, RI, 197–248 (2019)

[41] R. Taylor, Galois representations. *Ann. Fac. Sci. Toulouse Math. (6)* **13**, 73–119 (2004)

[42] R. Taylor, Automorphy for some *l*-adic lifts of automorphic mod *l* Galois representations. II. *Publ. Math. Inst. Hautes Études Sci.* 183–239 (2008)

[43] R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)* **141**, 553–572 (1995)

[44] R. Taylor and T. Yoshida, Compatibility of local and global Langlands correspondences. *J. Amer. Math. Soc.* **20**, 467–493 (2007)

[45] J. Weinstein, Reciprocity laws and Galois representations: Recent breakthroughs. *Bull. Amer. Math. Soc. (N.S.)* **53**, 1–39 (2016)

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

Ana Caraiani is a Royal Society University Research Fellow and Reader in Pure Mathematics at Imperial College London.

a.caraiani@imperial.ac.uk