COMPOSITIO MATHEMATICA

Sato–Tate distributions and Galois endomorphism modules in genus 2

Francesc Fité, Kiran S. Kedlaya, Víctor Rotger and Andrew V. Sutherland

Compositio Math. **148** (2012), 1390–1442.

doi:10.1112/S0010437X12000279

FOUNDATION
COMPOSITIO
MATHEMATICA

LONDON
MATHEMATICAL
SOCIETY

Compositio Math. 148 (2012) 1390–1442
doi:10.1112/S0010437X12000279

Sato–Tate distributions and Galois endomorphism modules in genus 2

Francesc Fité, Kiran S. Kedlaya, Víctor Rotger and Andrew V. Sutherland

Abstract

For an abelian surface A over a number field k, we study the limiting distribution of the normalized Euler factors of the L-function of A. This distribution is expected to correspond to taking characteristic polynomials of a uniform random matrix in some closed subgroup of $\mathrm{USp}(4)$; this *Sato–Tate group* may be obtained from the Galois action on any Tate module of A. We show that the Sato–Tate group is limited to a particular list of 55 groups up to conjugacy. We then classify A according to the Galois module structure on the $\mathbb{R}$-algebra generated by endomorphisms of $A_{\overline{\mathbb{Q}}}$ (the *Galois type*), and establish a matching with the classification of Sato–Tate groups; this shows that there are at most 52 groups up to conjugacy which occur as Sato–Tate groups for suitable A and k, of which 34 can occur for $k = \mathbb{Q}$. Finally, we present examples of Jacobians of hyperelliptic curves exhibiting each Galois type (over $\mathbb{Q}$ whenever possible), and observe numerical agreement with the expected Sato–Tate distribution by comparing moment statistics.

1. Introduction

The celebrated *Sato–Tate conjecture* concerns the distribution of Euler factors of an elliptic curve over a number field. It specifically predicts that this distribution always takes one of three forms, one occurring whenever the elliptic curve fails to have complex multiplication (the generic case), and two exceptional cases arising for CM curves (only one of which occurs over $\mathbb{Q}$). Substantial progress has been made on this conjecture only very recently (see § 1.7).

The purpose of this paper is to formulate a precise analogue of the Sato–Tate conjecture for abelian surfaces. We present strong theoretical and experimental evidence that the distribution of Euler factors can take as many as 52 forms, 35 of which occur if we limit to fields with a real place and 34 of which occur if we limit to $\mathbb{Q}$.

In the rest of the introduction, we introduce the Sato–Tate problem for a general abelian variety, describe a conjectural description of the distribution of Euler factors in terms of a certain compact Lie group (the *Sato–Tate group*), and then discuss our analysis of this conjecture from three points of view:

(a) a classification of compact Lie groups that are compatible with restrictions on the Sato–Tate group imposed by existing conjectures;

Received 8 November 2011, accepted in final form 17 January 2012, published online 25 July 2012.
2010 Mathematics Subject Classification 11M50 (primary), 11G10, 11G20, 14G10, 14K15 (secondary).
Keywords: Sato–Tate distributions, abelian surfaces, endomorphism algebras, Galois type.
Financial support was received as follows. Kedlaya: NSF CAREER grant DMS-0545904, NSF grant DMS-1101343, DARPA grant HR0011-09-1-0048, MIT (NEC Fund, Cecil and Ida Green professorship), UC San Diego (Stefan E. Warschawski professorship); Sutherland: NSF grant DMS-1115455; Fité and Rotger: DGICYT Grant MTM2009-13060-C02-01; Fité: Fundació Ferran Sunyer i Balaguer.

(b) a classification of Galois module structures on the $\mathbb{R}$-algebra generated by endomorphisms (which we call the *Galois type*);

(c) numerical computations testing the relationship between these classifications and observed distributions of Euler factors.

We conclude with some speculation about how much of the Sato–Tate conjecture for abelian surfaces may be tractable in the near future.

Throughout the introduction, let k denote a number field, let G_k be an absolute Galois group of k, let $\mathfrak{p}$ be a prime ideal of k, and let $q := \|\mathfrak{p}\|$ denote the absolute norm of $\mathfrak{p}$. When we make statements about averages over prime ideals, we always sort in increasing order by norm; it will not matter how ties are broken. For A an abelian variety over k and k' a field containing k, write $A_{k'}$ for the base extension of A from k to k'.

1.1 The Sato–Tate conjecture and random matrices

Let E be an elliptic curve over k. If E has good reduction at $\mathfrak{p}$ (which excludes only finitely many primes), then Hasse's theorem implies that the integer $a_\mathfrak{p} = q + 1 - \#E(\mathbb{F}_q)$ has absolute value at most $2\sqrt{q}$. One observes in examples[1] that the quantities

$$\theta_\mathfrak{p} := \arccos \frac{a_\mathfrak{p}}{2\sqrt{q}}$$

appear to be equidistributed with respect to a certain measure on $[0, \pi]$.

- If E has complex multiplication defined over k, one takes the uniform measure. Note that this case cannot occur if $k = \mathbb{Q}$, or even if k has a real place.
- If E has complex multiplication not defined over k, one takes half of the uniform measure plus a discrete measure of mass $1/2$ concentrated at $\pi/2$. The discrete measure occurs because $a_\mathfrak{p} = 0$ whenever $\mathfrak{p}$ fails to split in the quadratic extension of k over which the complex multiplication is defined.
- If E fails to have complex multiplication, one takes the measure $(2/\pi)\sin^2\theta\, d\theta$.

In the first two cases, it is easy to prove equidistribution using the explicit description of $a_\mathfrak{p}$ in terms of Grössencharacters. The third case is subtler: it is the *Sato–Tate conjecture*, which has recently been established when k is totally real (see § 1.7).

The measures described above admit the following interpretation. If one chooses a matrix uniformly at random from SU(2) (with respect to the Haar measure), its eigenvalues have the form $e^{i\theta}, e^{-i\theta}$ for θ distributed according to the measure $(2/\pi)\sin^2\theta\, d\theta$ on $[0, \pi]$. If one replaces SU(2) by $\mathrm{U}(1) \simeq \mathrm{SO}(2)$, one obtains the uniform measure; if one instead takes the normalizer of U(1) in SU(2), one gets half of the uniform measure (from the identity connected component) plus a discrete measure of mass $1/2$ concentrated at $\pi/2$ (from the other connected component).

1.2 The Sato–Tate group

The interpretation in terms of random matrices suggests a good formulation of the *Sato–Tate problem*[2] for an abelian variety A of arbitrary dimension $g \geqslant 1$. If A has good reduction at $\mathfrak{p}$, then there exists a polynomial $L_\mathfrak{p}(A, T) = \prod_{i=1}^{2g}(1 - T\alpha_i)$ over $\mathbb{Z}$ such that for each positive

[1] To observe for yourself, see the animations at http://math.mit.edu/~drew.

[2] In the language of [Ser12, § 8], we only consider the *weight* 1 case of the Sato–Tate problem. Some examples of the weight 2 case are discussed in [Ser12, § 8.5.6]. See also Remark 3.3.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

integer n,

$$\#A(\mathbb{F}_{q^n}) = \prod_{i=1}^{2g}(1-\alpha_i^n).$$

For example, if $g=1$, then $L_{\mathfrak{p}}(A,T) = 1 - a_{\mathfrak{p}}T + qT^2$. One can reinterpret $L_{\mathfrak{p}}(A,T)$ as follows. For any prime ℓ, let $V_\ell(A) = \mathbb{Q}\otimes T_\ell(A)$ denote the (rational) ℓ-adic Tate module of A. If $\mathrm{Frob}_{\mathfrak{p}}$ is an arithmetic Frobenius element of G_k for the prime $\mathfrak{p}$, then

$$L_{\mathfrak{p}}(A,T) = \det(1 - T\mathrm{Frob}_{\mathfrak{p}}, V_\ell(A)).$$

This implies that $L_{\mathfrak{p}}(A, q^{-s})$ is the Euler factor at $\mathfrak{p}$ in the L-function of A evaluated at s. Define the *normalized L-polynomial* $\overline{L}_{\mathfrak{p}}(A,T) := L_{\mathfrak{p}}(A, q^{-1/2}T)$; by a theorem of Weil, the roots of $\overline{L}_{\mathfrak{p}}(A,T)$ in $\mathbb{C}$ have norm 1 and are stable (as a multiset) under complex conjugation. The polynomial $\overline{L}_{\mathfrak{p}}(A,T)$ thus corresponds to a unique element of the set $\mathrm{Conj}(\mathrm{USp}(2g))$ of conjugacy classes in the unitary symplectic group $\mathrm{USp}(2g)$.

For generic A, Katz and Sarnak [KS99] predict that the polynomials $\overline{L}_{\mathfrak{p}}(A,T)$ are equidistributed[3] with respect to the image on $\mathrm{Conj}(\mathrm{USp}(2g))$ of the normalized Haar measure on $\mathrm{USp}(2g)$. It is tempting to guess that in general, the $\overline{L}_{\mathfrak{p}}(A,T)$ are equidistributed with respect to the image on $\mathrm{Conj}(\mathrm{USp}(2g))$ of the normalized Haar measure on a suitable closed subgroup G of $\mathrm{USp}(2g)$ (depending on A); however, to formulate a precise[4] conjecture, one needs an explicit definition of the group G.

We will give in § 2 a definition of the *Sato–Tate group* ST_A, using a construction in terms of ℓ-adic monodromy groups described by Serre in [Ser12, § 8.3]. This construction relates strongly to an earlier description given by Serre in terms of motivic Galois groups [Ser94], as well as to the definition of the *Mumford–Tate group* of A; see § 2.2 for more on the relationship between the Mumford–Tate and Sato–Tate groups. From the construction of ST_A, the map $\mathfrak{p} \mapsto \overline{L}_{\mathfrak{p}}(A,T) \in \mathrm{Conj}(\mathrm{USp}(2g))$ will factor through an assignment $\mathfrak{p} \mapsto s(\mathfrak{p}) \in \mathrm{Conj}(\mathrm{ST}_A)$, enabling us to make the following conjecture.

Conjecture 1.1 (Refined Sato–Tate). For ST_A the subgroup of $\mathrm{USp}(2g)$ defined in Definition 2.6 and μ_{ST_A} the image on $\mathrm{Conj}(\mathrm{ST}_A)$ of the normalized Haar measure on ST_A, the classes $s(\mathfrak{p}) \in \mathrm{Conj}(\mathrm{ST}_A)$ are μ_{ST_A}-equidistributed.

This is somewhat stronger than asserting equidistribution of the $\overline{L}_{\mathfrak{p}}(A,T)$ for the image measure on $\mathrm{Conj}(\mathrm{USp}(2g))$, but it is only this last conclusion that we test numerically.

1.3 A group-theoretic classification

For the remainder of this introduction, we assume[5] $g=2$. We can further clarify the Sato–Tate conjecture in this case by classifying groups which can occur as ST_A in Conjecture 1.1.

Theorem 1.2. *Let A be an abelian surface over k. Then ST_A is conjugate to one of 55 particular groups (see Theorem 3.4 for the list).*

[3] Note the analogy with the Chebotarev density theorem: in a finite Galois extension of number fields, the distribution of Frobenius conjugacy classes over the Galois group is also governed by the image of the Haar measure.

[4] It is unclear whether G is even determined up to conjugacy, even though the conjugacy class of an individual element of $\mathrm{USp}(2g)$ is determined by its characteristic polynomial.

[5] It should be possible to make a similar analysis for larger values of g, but even for $g=3$ we have no idea how many groups to expect!

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

The proof of Theorem 1.2 is an exercise with Lie groups, given the limitations placed on ST_A by properties of its construction. Most of the cases arise when the connected part of ST_A is isomorphic to $\mathrm{U}(1)$; these cases are related to the finite subgroups of $\mathrm{SO}(3)$.

The alert reader will notice that Theorem 1.2 mentions a list of 55 groups, whereas we claimed initially that only 52 groups are possible. That is because three cases survive the group-theoretic analysis but are ruled out by the comparison to Galois types (Theorem 1.4).

1.4 Galois types

Our next step towards controlling the Sato–Tate group of an abelian surface is to study the Galois module structure on the endomorphism algebra of A. This data is in general insufficient to control the Sato–Tate group (this is related to Mumford's exotic examples of Hodge groups for abelian fourfolds [Mum69]); however, in dimension 2 such pathologies do not arise, and indeed we gain a complete understanding of the Sato–Tate group this way.

Let $\mathrm{End}(A)$ denote the (not necessarily commutative) ring of endomorphisms of A. Note that these are assumed to be defined over k; if we mean to take endomorphisms of A defined over a larger field k', we will write $\mathrm{End}(A_{k'})$ instead. In fact, we will often write $\mathrm{End}(A_k)$ instead of $\mathrm{End}(A)$ to emphasize rationality over k. For any field L, write $\mathrm{End}(A)_L$ for $\mathrm{End}(A) \otimes_{\mathbb{Z}} L$.

Let K/k denote the minimal extension over which all endomorphisms of $A_{\overline{\mathbb{Q}}}$ are defined; it is a finite Galois extension of k (see Proposition 4.1).

Definition 1.3. Consider pairs $[G, E]$ in which G is a finite group and E is a finite-dimensional $\mathbb{R}$-algebra equipped with an action of G by $\mathbb{R}$-algebra automorphisms. An *isomorphism* between two such pairs $[G, E]$ and $[G', E']$ consists of an isomorphism $G \simeq G'$ of groups and an equivariant isomorphism $E \simeq E'$ of $\mathbb{R}$-algebras. The *Galois type* associated to A is the isomorphism class of the pair $[\mathrm{Gal}(K/k), \mathrm{End}(A_K)_{\mathbb{R}}]$. Note that abelian surfaces defined over different number fields may have the same Galois type.

Theorem 1.4. *Let A be an abelian surface over k. Then the conjugacy class of the Sato–Tate group of A is uniquely determined by the Galois type and vice versa.*

In Theorem 1.4, the passage from the Sato–Tate group to the Galois type is fairly explicit; see Proposition 2.19. The reverse implication can be seen by computing the Galois types associated to the 55 groups named in Theorem 1.2 and seeing that they are pairwise nonisomorphic. This requires significantly less data than the full Galois type; for instance, it is sufficient to keep track of the isomorphism class (as an $\mathbb{R}$-algebra) of the $\mathbb{R}$-subalgebra of E fixed by one subgroup of $\mathrm{Gal}(K/k)$ in each conjugacy class.

We also make a more detailed analysis of the Galois type and its relationship to more apparent arithmetic of the abelian surface of A, such as the shape of the simple factors of A. This leads to the following result; for a more precise statement, see Theorem 4.3.

Theorem 1.5. *There exist exactly 52 Galois types of abelian surfaces over number fields. Of these, exactly 35 can be realized in such a way that k has a real place, and exactly 34 can be realized in such a way that $k = \mathbb{Q}$.*

It is worth pointing out that the definition of the Galois type had to be chosen rather carefully in order to make Theorem 1.4 valid. For example, in [KS09] one finds examples of abelian surfaces $A = \mathrm{Jac}(C)/\mathbb{Q}$ and $A' = \mathrm{Jac}(C')/\mathbb{Q}$ with the same Sato–Tate group such that A is absolutely simple and A' is isogenous to the product of two elliptic curves. This is because $\mathrm{End}(A_K)_{\mathbb{Q}}$ is a

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

division quaternion algebra while $\mathrm{End}(A'_K)_{\mathbb{Q}} \simeq \mathrm{M}_2(\mathbb{Q})$; this shows that in the definition of the Galois type, considering the $\mathbb{Q}$-algebra $\mathrm{End}(A_K)_{\mathbb{Q}}$ yields a classification which is too fine. On the other hand, the $\mathbb{C}$-algebra $\mathrm{End}(A_K)_{\mathbb{C}}$ gives a classification which is too coarse: there are examples of abelian surfaces $A = \mathrm{Jac}(C)/\mathbb{Q}$ and $A' = \mathrm{Jac}(C')/\mathbb{Q}$ with different Sato–Tate groups such that $\mathrm{End}(A_{\mathbb{Q}})_{\mathbb{R}} \simeq \mathbb{R} \times \mathbb{R}$ and $\mathrm{End}(A'_{\mathbb{Q}})_{\mathbb{R}} \simeq \mathbb{C}$, and thus $\mathrm{End}(A_{\mathbb{Q}})_{\mathbb{C}} \simeq \mathrm{End}(A'_{\mathbb{Q}})_{\mathbb{C}} \simeq \mathbb{C} \times \mathbb{C}$.

1.5 Numerical computations

Our final step is to test Conjecture 1.4 numerically for Jacobians of curves of genus 2. For a given curve C, it is a finite problem to identify the Galois type of its Jacobian. This determines the Sato–Tate group $\mathrm{ST}_{\mathrm{Jac}(C)}$, yielding a predicted distribution of normalized L-polynomials. To test convergence to the predicted distribution, we follow the methodology introduced in [KS09], and consider *moment statistics* of the linear and quadratic coefficients a_1 and a_2 of the normalized L-polynomial. We first compute the corresponding moments for all of the Sato–Tate groups that can arise in genus 2; see § 5.1. For any given curve C, we may then use the methods of [KS08] to compute a large quantity of L-polynomial data for C, from which we derive moment statistics for a_1 and a_2 that may be compared to the corresponding moments for the Sato–Tate group. We may also compare histograms of the normalized L-polynomial coefficients with the corresponding density functions for the Sato–Tate group; see § 5.2. Although here we present computational results only after giving the theoretical description of Sato–Tate groups and Galois types, the order of discovery was the reverse: it would have been quite impossible to establish the theoretical results without numerical evidence to lead the way.

It is worth mentioning that one can run the methodology in two different directions. Given an abelian surface in explicit form, one can in principle compute the Galois type and then obtain a prediction for the Sato–Tate distribution. On the other hand, in practice, it is often easier to compute the Sato–Tate distribution numerically and then use this to guess the Galois type! This state of affairs may persist for larger g; for instance, it may be possible to identify examples of Mumford's exotic fourfolds most easily from their Sato–Tate distributions.

1.6 Sato–Tate for abelian surfaces

Combining all of our ingredients, we now have a precise Sato–Tate conjecture for abelian surfaces, including the following features.

(a) The $\overline{L}_{\mathfrak{p}}(A, T)$ in $\mathrm{Conj}(\mathrm{USp}(2g))$ are conjectured to be equidistributed according to the image of the Haar measure for a specific group ST_A (Definition 2.6).

(b) The statement in (a) can be refined to an equidistribution conjecture on $\mathrm{Conj}(\mathrm{ST}_A)$ itself (Conjecture 1.1).

(c) It is known exactly which groups ST_A can occur in general (there are 52 of them), which can occur for $k = \mathbb{Q}$ (there are 34 of them), and how ST_A is related to the endomorphism ring of A (by Theorems 1.4 and 1.5).

By contrast, an analysis of Sato–Tate groups in dimension 2 over $\mathbb{Q}$ had previously been attempted by the second and fourth authors in [KS09] based on results of an exhaustive search, but produced fewer groups than we identify here. One reason is that in [KS09], curves were classified only using the moments of the first coefficient of the L-polynomial, which turns out to be insufficient: over $\mathbb{Q}$, the 34 observed Sato–Tate groups only give rise to 26 distinct trace distributions. Another reason is that in [KS09], the Sato–Tate groups were constructed in a haphazard fashion, without the benefit either of a group-theoretic classification or an

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

interpretation in terms of Galois types. It was the combination of these additional ingredients that made it possible to identify some exceptional cases missed in [KS09]; with the benefit of this hindsight, we then made a larger exhaustive search to identify examples with small coefficients (see § 5.3).

1.7 Tractable cases

Having asserted a precise form of the Sato–Tate conjecture for abelian surfaces, we conclude this introduction by discussing to what extent it may be possible to prove cases of this conjecture in the near future (though not in this paper).

We first recall the paradigm for proving equidistribution described by Serre [Ser68, § I.A.2]. Conjecture 1.1 asserts that for any continuous function $f : \mathrm{Conj}(\mathrm{ST}_A) \to \mathbb{C}$,

$$\mu_{\mathrm{ST}_A}(f) = \lim_{n\to\infty} \frac{\sum_{\|\mathfrak{p}\|\leqslant n} f(s(\mathfrak{p}))}{\#\{\mathfrak{p} : \|\mathfrak{p}\| \leqslant n\}}. \tag{1.1}$$

By the Peter–Weyl theorem, the space of characters of ST_A is dense for the supremum norm on the space of continuous functions on $\mathrm{Conj}(\mathrm{ST}_A)$, so we need only check (1.1) when $f = \chi$ is an irreducible character; we may omit the trivial character since (1.1) is obvious in that case. By emulating the proofs of the prime number theorem by Hadamard and de la Vallée Poussin, one then obtains the following result.

Theorem 1.6 (Serre). *Suppose that for each nontrivial irreducible linear representation*[6] ρ *of* ST_A*, the Dirichlet series*

$$L(A, \rho, s) = \prod_{\mathfrak{p}} \det(1 - \rho(s(\mathfrak{p}))\|\mathfrak{p}\|^{-s})^{-1}$$

(which converges absolutely for $\mathrm{Re}(s) > 1$*) extends to a holomorphic function on* $\mathrm{Re}(s) \geqslant 1$ *which does not vanish anywhere on the line* $\mathrm{Re}(s) = 1$*. Then Conjecture 1.1 holds for A.*

For the most part, the only known method for establishing meromorphic continuation of $L(A, \rho, s)$ is to show that it arises from an automorphic form. This has been achieved recently in many cases where A corresponds to an automorphic form of GL_2-type, thanks to the work of Barnet-Lamb, Clozel, Gee, Geraghty, Harris, Shepherd-Barron, and Taylor. For example, Conjecture 1.1 is known (with $\mathrm{ST}_A = \mathrm{SU}(2)$) whenever A is an elliptic curve over a totally real field without complex multiplication. (See [BGG11] for an even stronger result covering Hilbert modular forms.)

For $g = 2$, it is unclear at present how to approach the Sato–Tate problem in even a single case with $\mathrm{ST}_A = \mathrm{USp}(4)$; however, one can hope to treat all cases with $\mathrm{ST}_A \neq \mathrm{USp}(4)$ using existing technology. For instance, it should be straightforward to verify Conjecture 1.1 when A is isogenous to the product of two elliptic curves with complex multiplication. For k totally real, it should also be possible to use the result of [BGG11] to handle cases where A is isogenous to the product of two elliptic curves, one of which has complex multiplication. The case where A is isogenous to the product of two nonisogenous non-CM elliptic curves is harder, but has been treated by Harris [Har09] conditionally on some results on the stable trace formula which have subsequently been verified; see [BGHT11] for the appropriate references. (Thanks to Toby Gee and Michael Harris for this explanation.)

[6] It is enough to consider the restrictions of representations of $\mathrm{USp}(2g)$, but then one finds a pole at $s = 1$ of order equal to the multiplicity of the trivial representation in $\rho|_{\mathrm{ST}_A}$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

2. Construction of the Sato–Tate group

Let A be an abelian variety over a number field k of dimension $g \geqslant 1$. In this section, we give an explicit definition of the Sato–Tate group ST_A that appears in our refined form of the Sato–Tate conjecture (Conjecture 1.1). This loosely follows the presentation given by Serre in [Ser12, § 8.3]. We then relate this group to the Mumford–Tate group; for abelian varieties of dimension at most 3, this gives a precise description of the Sato–Tate group in terms of the endomorphisms of A and the Galois action on them. This depends crucially on work of the second author with Grzegorz Banaszak [BK11].

2.1 ℓ-adic monodromy and the Sato–Tate group

Choose a polarization on A and embeddings $k \hookrightarrow \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$. Choose a symplectic basis for the singular homology group $H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q})$ and use it to equip this space with an action of $\mathrm{GSp}_{2g}(\mathbb{Q})$, where $\mathrm{GSp}_{2g}/\mathbb{Q}$ denotes the reductive algebraic group over $\mathbb{Q}$ such that for any field F,

$$\mathrm{GSp}_{2g}(F) = \left\{ \gamma \in \mathrm{GL}_{2g}(F) : \gamma^t \begin{pmatrix} 0 & -\mathrm{Id} \\ \mathrm{Id} & 0 \end{pmatrix} \gamma = \lambda_\gamma \cdot \begin{pmatrix} 0 & -\mathrm{Id} \\ \mathrm{Id} & 0 \end{pmatrix},\ \lambda_\gamma \in F^\times \right\}.$$

Now fix a prime ℓ, let $V_\ell(A)$ denote the rational ℓ-adic Tate module of A, and make the identifications

$$V_\ell(A) \simeq H_{1,\mathrm{et}}(A_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \simeq H_{1,\mathrm{et}}(A_{\mathbb{C}}, \mathbb{Q}_\ell) \simeq H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q}_\ell) \simeq H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell.$$

Under these identifications, the Weil pairing on the Tate module is identified with the cup product pairings in étale and singular homology, so our chosen symplectic basis for $H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q})$ maps to a symplectic basis of $V_\ell(A)$. The action of G_k on $V_\ell(A)$ thus defines a continuous homomorphism

$$\varrho_{A,\ell} : G_k \longrightarrow \mathrm{GSp}_{2g}(\mathbb{Q}_\ell). \tag{2.1}$$

Write $G_\ell = G_\ell(A)$ for the image of G_k under $\varrho_{A,\ell}$.

Definition 2.1. Let $G_\ell^{\mathrm{Zar}} = G_\ell^{\mathrm{Zar}}(A)$ denote the Zariski closure of G_ℓ inside $\mathrm{GSp}_{2g}(\mathbb{Q}_\ell)$; this is sometimes called the *ℓ-adic monodromy group* of A.

A result of Bogomolov [Bog80] (plus a bit of p-adic Hodge theory; see [Ser12, § 8.3.2] for references) ensures that G_ℓ is open in G_ℓ^{Zar}, and this construction behaves reasonably under enlargement of the field k. See [BK11, § 3] for similar arguments.

Remark 2.2. Let k' be a finite extension of k. Since $G_\ell(A_{k'})$ is an open subgroup of the compact group G_ℓ by Bogomolov's theorem, $G_\ell^{\mathrm{Zar}}(A_{k'})$ is a subgroup of G_ℓ^{Zar} of finite index; in particular, the two groups have the same connected part. On the other hand, by making k' large enough, we can force $G_\ell(A_{k'})$ to lie within a neighborhood of the identity in $\mathrm{GSp}_{2g}(\mathbb{Q}_\ell)$ that is small enough to miss all of the nonidentity connected components of G_ℓ^{Zar}. Consequently, for any sufficiently large k', $G_\ell^{\mathrm{Zar}}(A_{k'})$ equals the connected part of G_ℓ^{Zar}. This means that, on the one hand, the connected part of G_ℓ^{Zar} is an invariant of $A_{\overline{\mathbb{Q}}}$; on the other hand, the component group $\pi_0(G_\ell^{\mathrm{Zar}})$ receives a surjective continuous homomorphism from G_k, and so may be identified with $\mathrm{Gal}(K/k)$ for some[7] finite extension K of k.

Remark 2.3. By a celebrated theorem of Faltings [Fal83], one has

$$\mathrm{End}(A)_{\mathbb{Q}_\ell} = \mathrm{End}(V_\ell(A))^{G_\ell}. \tag{2.2}$$

[7] It is not clear from this construction that K is independent of ℓ, but this has been shown by Serre [Se81].

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

By (2.2), the ℓ-adic monodromy groups of $A_{k'}$ for all finite extensions k' of k determine the G_k-module $\mathrm{End}(A_{\overline{\mathbb{Q}}})_{\mathbb{Q}_\ell}$. The converse is not true in general: there are examples due to Mumford [Mum69] of simple abelian fourfolds A/k such that $\mathrm{End}(A_{\overline{\mathbb{Q}}}) = \mathbb{Z}$ while $G_\ell^{\mathrm{Zar}} \subsetneq \mathrm{GSp}_{2g}(\mathbb{Q}_\ell)$. Fortunately, since we only consider the $g = 2$ case in this paper, we will avoid such pathologies; see Theorem 2.16.

To get the Sato–Tate group, we must modify the above construction slightly.

DEFINITION 2.4. Let G_k^1 denote the kernel of the cyclotomic character $\chi_\ell : G_k \to \mathbb{Q}_\ell^\times$, and let $G_\ell^{1,\mathrm{Zar}} = G_\ell^{1,\mathrm{Zar}}(A)$ be the Zariski closure of $\varrho_{A,\ell}(G_k^1)$ in $\mathrm{GSp}_{2g}/\mathbb{Q}_\ell$. By Bogomolov's theorem, the natural inclusion of $G_\ell^{1,\mathrm{Zar}}$ into the kernel of the composition $G_\ell^{\mathrm{Zar}} \to \mathrm{GSp}_{2g} \to \mathbb{G}_m$ is an isomorphism.

Remark 2.5. We have an analogue of Remark 2.2: for k' a finite extension of k, $G_\ell^{1,\mathrm{Zar}}(A_{k'})$ is a subgroup of $G_\ell^{1,\mathrm{Zar}}(A)$ with the same connected part, and is itself connected when k' is sufficiently large.

DEFINITION 2.6. Choose an embedding $\iota : \mathbb{Q}_\ell \hookrightarrow \mathbb{C}$ and use it to view $\varrho_{A,\ell}$ as having target $\mathrm{GSp}_{2g}(\mathbb{C})$. Put $G^1 := G_\ell^{1,\mathrm{Zar}} \otimes_\iota \mathbb{C}$. The *Sato–Tate group* ST_A of A (for the prime ℓ and the embedding ι) is a maximal compact Lie subgroup of G^1 contained in $\mathrm{USp}(2g)$ (which exists because the latter is a maximal subgroup of $\mathrm{Sp}_{2g}(\mathbb{C})$).

Remark 2.7. For example, if A is an elliptic curve, then by Serre's open image theorem [Ser72], $\mathrm{ST}_A = \mathrm{SU}(2)$ if and only if A has no complex multiplication.

LEMMA 2.8. *The groups of connected components of the groups*

$$G_\ell^{\mathrm{Zar}},\ G_\ell^{1,\mathrm{Zar}},\ G^1,\ \mathrm{ST}_A$$

are all canonically isomorphic.

As a corollary, it follows that for k' a finite extension of k, $G_\ell^{\mathrm{Zar}}(A_{k'})$ is connected if and only if $G_\ell^{1,\mathrm{Zar}}(A_{k'})$ is connected.

Proof. (This is from [BK11, Theorem 3.4].) Apply Remark 2.2 and its analogue (Remark 2.5) to produce a finite extension k' of k for which $G_\ell^{1,\mathrm{Zar}}(A_{k'})$ and $G_\ell^{\mathrm{Zar}}(A_{k'})$ are both connected. Then the diagram

$$\begin{array}{ccccccccc}
 & & 1 & & 1 & & 1 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & G_\ell^{1,\mathrm{Zar}}(A_{k'}) & \longrightarrow & G_\ell^{\mathrm{Zar}}(A_{k'}) & \longrightarrow & \mathbb{G}_m & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & G_\ell^{1,\mathrm{Zar}}(A_k) & \longrightarrow & G_\ell^{\mathrm{Zar}}(A_k) & \longrightarrow & \mathbb{G}_m & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & \pi_0(G_\ell^{1,\mathrm{Zar}}) & \longrightarrow & \pi_0(G_\ell^{\mathrm{Zar}}) & & 1 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 1 & & 1 & & & &
\end{array}$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

has exact rows and columns, and a diagram chase (as in the proof of the snake lemma) shows that $\pi_0(G_\ell^{1,\mathrm{Zar}}) \to \pi_0(G_\ell^{\mathrm{Zar}})$ is an isomorphism. On the other hand, we have $\pi_0(G_\ell^{1,\mathrm{Zar}}) = \pi_0(G^1)$, because the $\mathbb{Q}_\ell$-rational points of $G_\ell^{1,\mathrm{Zar}}$ are Zariski dense, and $\pi_0(G^1) = \pi_0(\mathrm{ST}_A)$, because any maximal compact subgroup of a connected complex Lie group is a connected real Lie group. □

We now have the group ST_A appearing in the refined Sato–Tate conjecture (Conjecture 1.1), but we still need conjugacy classes corresponding to prime ideals. This construction will imply that $\overline{L}_\mathfrak{p}(A,T)$ belongs to the image of $\mathrm{Conj}(\mathrm{ST}_A) \to \mathrm{Conj}(\mathrm{USp}(2g))$. See [Ser12, § 8.3.3] for further discussion.

Definition 2.9. For $G = G_\ell^{\mathrm{Zar}} \otimes_\iota \mathbb{C}$, we may identify G/G^1 with $\mathbb{C}^\times$ compatibly with the cyclotomic character. The image of $g_\mathfrak{p} := \varrho_{A,\ell}(\mathrm{Frob}_\mathfrak{p}) \in G$ in $\mathbb{C}^\times$ is q, so $g'_\mathfrak{p} := q^{-1/2} g_\mathfrak{p}$ belongs to G^1. The semisimple component of $g'_\mathfrak{p}$ for the Jordan decomposition is an element of G^1 with eigenvalues of norm 1, and therefore belongs to a conjugate of ST_A. We thus associate to $\mathfrak{p}$ a class $s(\mathfrak{p}) \in \mathrm{Conj}(\mathrm{ST}_A)$.

Remark 2.10. It is generally expected that $g_\mathfrak{p}$ is already semisimple, in which case $s(\mathfrak{p})$ would be the class of $g'_\mathfrak{p}$ itself.

2.2 Mumford–Tate groups and Sato–Tate groups

From the above definition of the Sato–Tate group, it is difficult to recover much information relating the Sato–Tate group to the arithmetic of A. To go further, we must control the Sato–Tate group in terms of the endomorphisms of A. This is impossible in general, as shown for instance by Mumford's examples in dimension 4 [Mum69]; however, no such pathologies occur for $g \leqslant 3$.

Definition 2.11. Fix an embedding $k \hookrightarrow \mathbb{C}$.

(i) The Mumford–Tate group MT_A of A is the smallest algebraic subgroup G of $\mathrm{GL}(H_1(A_\mathbb{C}^{\mathrm{top}}, \mathbb{Q}))$ *over* $\mathbb{Q}$ such that $G(\mathbb{R})$ contains $h(\mathbb{C}^\times)$, where

$$h : \mathbb{C} \longrightarrow \mathrm{End}_\mathbb{R}(H_1(A_\mathbb{C}^{\mathrm{top}}, \mathbb{R}))$$

is the complex structure on the $2g$-dimensional real vector space $H_1(A_\mathbb{C}^{\mathrm{top}}, \mathbb{R})$ obtained by identifying it with the tangent space of A at the identity.

(ii) The Hodge group of A is $\mathrm{Hg}_A := (\mathrm{MT}_A \cap \mathrm{SL}_{2g})^0$.

These can be viewed as Archimedean analogues of the groups G_ℓ^{Zar} and $G_\ell^{1,\mathrm{Zar}}$. By a theorem of Deligne [Del82, Proposition 6.2], for $G_\ell^{\mathrm{Zar},0}$ the connected component of the identity of G_ℓ^{Zar}, we have $G_\ell^{\mathrm{Zar},0} \subseteq \mathrm{MT}_A \otimes_\mathbb{Q} \mathbb{Q}_\ell$.

Conjecture 2.12 (Mumford–Tate). The inclusion $G_\ell^{\mathrm{Zar},0} \subseteq \mathrm{MT}_A \otimes_\mathbb{Q} \mathbb{Q}_\ell$ is always an equality. Equivalently, the induced inclusion $G_\ell^{1,\mathrm{Zar},0} \subseteq \mathrm{Hg}_A \otimes_\mathbb{Q} \mathbb{Q}_\ell$ is also an equality.

One cannot hope to describe G_ℓ^{Zar} or the algebraic Sato–Tate group any more closely than this using the Mumford–Tate group, because only the connected parts of these groups are determined by $A_\mathbb{C}$. This suggests the following refinement of the Mumford–Tate conjecture [BK11, Conjecture 2.3].

Conjecture 2.13. There exists an algebraic subgroup AST_A of $\mathrm{GSp}_{2g}/\mathbb{Q}$, called the *algebraic Sato–Tate group* of A, such that for each prime ℓ, $G_\ell^{1,\mathrm{Zar}} = \mathrm{AST}_A \otimes_\mathbb{Q} \mathbb{Q}_\ell$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

When Conjecture 2.13 holds, we may interpret ST_A as a maximal compact subgroup of $\mathrm{AST}_A \otimes_{\mathbb{Q}} \mathbb{C}$. In many cases, including all cases with $g \leqslant 3$, one can resolve Conjecture 2.13 by giving a full description of both the Mumford–Tate group and the algebraic Sato–Tate group in terms of the complex structure on $H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{R})$. This has been carried out by the second author and Grzegorz Banaszak in [BK11], building on much existing literature on Mumford–Tate groups; we summarize the relevant results here.

Definition 2.14. The *Lefschetz group* L_A is defined as

$$\mathrm{L}_A := \{\gamma \in \mathrm{Sp}_{2g} : \gamma^{-1}\alpha\gamma = \alpha \text{ for all } \alpha \in \mathrm{End}(A_{\overline{\mathbb{Q}}})_{\mathbb{Q}}\}^0.$$

Here we view α as an endomorphism of $H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q})$; consequently, L_A is an algebraic subgroup of $\mathrm{GSp}_{2g}/\mathbb{Q}$. There is an obvious inclusion $\mathrm{Hg}_A \subseteq \mathrm{L}_A$.

Definition 2.15. For each $\tau \in G_k$, define

$$\mathrm{L}_A(\tau) := \{\gamma \in \mathrm{Sp}_{2g} : \gamma^{-1}\alpha\gamma = \tau(\alpha) \text{ for all } \alpha \in \mathrm{End}(A_{\overline{\mathbb{Q}}})_{\mathbb{Q}}\}.$$

The *twisted Lefschetz group* TL_A is defined as

$$\mathrm{TL}_A := \bigcup_{\tau \in G_k} \mathrm{L}_A(\tau).$$

It is an algebraic group over $\mathbb{Q}$ with connected part L_A.

Theorem 2.16. (a) *Suppose that* $\mathrm{Hg}_A = \mathrm{L}_A$ *and that Conjecture 2.12 holds for* A. *Then Conjecture 2.13 holds with* $\mathrm{AST}_A = \mathrm{TL}_A$. *Consequently,* AST_A *is reductive and* ST_A *is a maximal compact subgroup of* $\mathrm{TL}_A \otimes_{\mathbb{Q}} \mathbb{C}$.

(b) *The hypotheses of (a) hold when* $g \leqslant 3$.

Proof. See [BK11, Theorems 6.1 and 6.10]. □

2.3 Extracting data from the Sato–Tate group

Using Theorem 2.16, we can recover from ST_A much data about the endomorphisms of A, starting with the minimal field of definition of endomorphisms.

Proposition 2.17. *For* $g \leqslant 3$, *the component groups of* G_ℓ^{Zar}, $G_\ell^{1,\mathrm{Zar}}$, *and* ST_A *may be identified with* $\mathrm{Gal}(K/k)$ *for* K/k *the minimal extension over which all endomorphisms of* $A_{\overline{\mathbb{Q}}}$ *are defined. Moreover, for* k' *a finite extension of* k, $\mathrm{ST}_{A_{k'}}$ *is the inverse image of* $\mathrm{Gal}(Kk'/k')$ *under the map* $\mathrm{ST}_A \to \mathrm{Gal}(K/k)$.

Proof. This is immediate from Theorem 2.16 and Lemma 2.8. □

To extract more information, we use the following construction.

Definition 2.18. Recall that we have fixed a polarization on A, which defines an isogeny ϕ from A to its dual variety $\widehat{A}$. The *Rosati involution* takes $\psi \in \mathrm{End}(A_{\mathbb{C}})_{\mathbb{Q}}$ to $\psi' = \phi^{-1} \circ \widehat{\psi} \circ \phi$. It has the following positivity property: the function

$$\psi \mapsto \mathrm{Trace}(\psi \circ \psi', H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q}))$$

is a positive definite quadratic form on $\mathrm{End}(A_{\mathbb{C}})$ (see [Mum69, § 21, Theorem 1]). We will call this form the *Rosati form*.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

PROPOSITION 2.19. *For $g \leqslant 3$, the subgroup ST_A of $\mathrm{USp}(2g)$ uniquely determines the $\mathbb{R}$-algebra $\mathrm{End}(A_K)_{\mathbb{R}}$ and its action by $\mathrm{Gal}(K/k)$.*

Proof. Note that

$$\mathrm{End}(A_K)_{\mathbb{Q}} = \mathrm{End}(A_{\mathbb{C}})_{\mathbb{Q}} = \mathrm{End}(H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q}))^{\mathrm{Hg}_A} = \mathrm{End}(H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q}))^{\mathrm{MT}_A}$$

(see [Mum69]). By Theorem 2.16, for $g \leqslant 3$ we can recover from ST_A the action of $\mathrm{Gal}(K/k)$ on $\mathrm{End}(A_K)_{\mathbb{C}}$ by taking the action of $\mathrm{TL}_A \,/\, \mathrm{L}_A$ on

$$(\mathrm{End}(H_1(A_{\mathbb{C}}^{\mathrm{top}}, \mathbb{Q})) \otimes_{\mathbb{Q}} \mathbb{C})^{\mathrm{L}_A}.$$

We can then identify $\mathrm{End}(A_K)_{\mathbb{R}}$ as the unique $\mathbb{R}$-subspace of $\mathrm{End}(A_K)_{\mathbb{C}}$ of half the dimension which is positive definite for the real part of the Rosati form. □

It will be important later to have an explicit description of the effect of twists on the Sato–Tate group.

DEFINITION 2.20. Let $f : \mathrm{Gal}(L/k) \to \mathrm{Aut}(A_K)$ be a continuous 1-cocycle, i.e., a function satisfying

$$f(\tau_1\tau_2) = f(\tau_1)\tau_1(f(\tau_2)) \quad \text{for } \tau_1, \tau_2 \in G_k$$

and factoring through $\mathrm{Gal}(L/k)$ for some finite Galois extension L of k containing K. Then there exists an abelian variety A^f over k equipped with an isomorphism $A^f_L \simeq A_L$ such that the action of $\tau \in G_k$ on $A^f(\overline{\mathbb{Q}}) \simeq A^f_L(\overline{\mathbb{Q}})$ corresponds to the action of $f(\tau)\tau$ on $A(\overline{\mathbb{Q}}) \simeq A_L(\overline{\mathbb{Q}})$. We call A^f the *twist* of A by f. In the case where A is the Jacobian of a genus 2 curve C over k and f factors through a 1-cocycle $f_C : \mathrm{Gal}(L/k) \to \mathrm{Aut}(C_K)$, we can identify A^f with the Jacobian of a twist C^f of C.

The isomorphism $A^f_L \simeq A_L$ induces an isomorphism $\mathrm{End}(A^f_L) \simeq \mathrm{End}(A_L)$ in which corresponding $\alpha \in \mathrm{End}(A^f_L)$ and $\beta \in \mathrm{End}(A_L)$ satisfy the relation $\tau(\alpha) = f(\tau)\tau(\beta)f(\tau)^{-1}$. As a consequence, for $\tau \in G_k$ we have

$$\begin{aligned}
\mathrm{L}_{A^f}(\tau) &= \{\gamma \in \mathrm{Sp}_{2g} : \gamma^{-1}\beta\gamma = f(\tau)\tau(\beta)f(\tau)^{-1} \text{ for all } \beta \in \mathrm{End}(A_L)_{\mathbb{Q}}\} \\
&= \{\gamma \in \mathrm{Sp}_{2g} : (\gamma f(\tau))^{-1}\beta\gamma f(\tau) = \tau(\beta) \text{ for all } \beta \in \mathrm{End}(A_L)_{\mathbb{Q}}\} \\
&= \mathrm{L}_A(\tau)f(\tau)^{-1}.
\end{aligned}$$

3. A group-theoretic classification

In this section, we record some necessary properties of the Sato–Tate group ST_A of an abelian surface A over a number field k, and then classify all closed subgroups of $\mathrm{USp}(4)$ exhibiting these properties; there are 55 such groups up to conjugacy. We give explicit representations for each of these groups that will be used to match the groups with Galois types (§ 4) and to compute invariants that characterize the distribution of characteristic polynomials (§ 5.1). Note that the classification includes three spurious groups; these will be ruled out in § 4.4.

3.1 Axioms for Sato–Tate groups

We first record some necessary conditions for a closed subgroup of $\mathrm{USp}(2g)$ to occur as a Sato–Tate group. Although we are only interested in the weight 1 case, for future reference we formulate these conditions in a manner suitable for considering self-dual motives of arbitrary

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

positive weight. (One can get slightly stronger conditions by accounting for the base field k; see Remark 3.5.)

Definition 3.1. Fix a positive integer w and some nonnegative integers $h^{p,q}$ for all $p, q \geqslant 0$ with $p+q=w$, and put $d = \sum_{p,q} h^{p,q}$. If w is odd, assume also that $h^{p,q} = h^{q,p}$ for all p, q. For a group G with identity connected component G^0, the *Sato–Tate axioms* are as follows.

(ST1) The group G is a closed subgroup of $\mathrm{USp}(d)$ (if w is odd) or of $\mathrm{O}(d)$ (if w is even).

(ST2) (Hodge condition) There exists a homomorphism $\theta : \mathrm{U}(1) \to G^0$ such that $\theta(u)$ has eigenvalues u^{p-q} with multiplicity $h^{p,q}$. The image of θ is called a *Hodge circle*.

(ST3) (Rationality condition) For each component H of G and each irreducible character χ of $\mathrm{GL}_d(\mathbb{C})$, the expected value (under the Haar measure) of $\chi(\gamma)$ over $\gamma \in H$ is an integer. In particular, for any positive integers m and n, the quantity $\mathrm{E}[\mathrm{Trace}(\gamma, \wedge^m \mathbb{C}^d)^n : \gamma \in H]$ lies in $\mathbb{Z}$.

The numbers $h^{p,q}$ are meant to be the Hodge numbers of the motive for which one is investigating the Sato–Tate conjecture. In the case of an abelian variety of dimension g, we should thus take $w=1$ and $h^{0,1} = h^{1,0} = g$.

Proposition 3.2. *Let A be an abelian variety over k of dimension g satisfying the Mumford–Tate conjecture (Conjecture 2.12) and the algebraic Sato–Tate conjecture (Conjecture 2.13). Then $G = \mathrm{ST}_A$ satisfies the Sato–Tate axioms for $w=1$ and $h^{0,1} = h^{1,0} = g$.*

Proof. Condition ST1 is clear from the construction. Condition ST2 follows from the definition of the Mumford–Tate group; it can also be derived using p-adic Hodge theory [Ser12, §§ 8.2.3.6 and 8.3.4].

To check ST3, let $\rho : \mathrm{GL}_{2g}(\mathbb{C}) \to V_{\mathbb{C}}$ be the representation corresponding to χ. Since the algebraic Sato–Tate conjecture has been assumed, we may write $V_{\mathbb{C}} = V \otimes_{\mathbb{Q}} \mathbb{C}$ for V a representation of $\mathrm{GL}_{2g}/\mathbb{Q}$. Since AST_A^0 is reductive, we may split $V = V_1 \oplus V_2$ so that AST_A acts on V_1 and V_2, AST_A^0 acts trivially on V_1, and V_2 has no nonzero subquotient on which AST_A^0 acts trivially. Let $H_{\mathbb{Q}}$ be the component of AST_A for which $H \subseteq H_{\mathbb{Q}}(\mathbb{C})$.

Put $t = \mathrm{E}[\mathrm{Trace}(\gamma, V_{\mathbb{C}}) : \gamma \in H]$. Since G^0 is a maximal compact subgroup of AST_A^0, $V_{2,\mathbb{C}}$ also has no G^0-trivial subrepresentations; hence, for $v_2 \in V_{2,\mathbb{C}}$, the G^0-invariant element $\mathrm{E}[\rho(\gamma)(v_2) : \gamma \in H]$ must be zero. It follows that $\mathrm{E}[\mathrm{Trace}(\gamma, V_{2,\mathbb{C}}) : \gamma \in H] = 0$. On the other hand, since V_1 is a trivial AST_A^0-representation, the function $\gamma \mapsto \mathrm{Trace}(\gamma, V_{1,\mathbb{C}})$ is constant on $H_{\mathbb{Q}}(\mathbb{C})$. We deduce that $t = \mathrm{Trace}(\gamma_0, V_{1,\mathbb{C}})$ for any single $\gamma_0 \in H_{\mathbb{Q}}(\mathbb{C})$. Since G/G^0 is finite, t is a sum of roots of unity and therefore an algebraic integer.

Let ℓ be an arbitrary prime and choose an algebraic isomorphism $\overline{\mathbb{Q}}_\ell \simeq \mathbb{C}$. By our hypotheses, $G_\ell^{1,\mathrm{Zar}}$ is a compact open subgroup of $\mathrm{AST}_A \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$; let H_ℓ be the coset of $G_\ell^{1,\mathrm{Zar},0}$ contained in $H_{\mathbb{Q}}(\mathbb{Q}_\ell)$. Put $t_\ell = \mathrm{E}[\mathrm{Trace}(\gamma, V_{\mathbb{Q}_\ell}) : \gamma \in H_\ell]$. Again, $\mathrm{E}[\mathrm{Trace}(\gamma, V_{2,\mathbb{Q}_\ell}) : \gamma \in H_\ell] = 0$, so $t_\ell = \mathrm{Trace}(\gamma_0, V_{1,\mathbb{C}})$ for any single $\gamma_0 \in H_{\mathbb{Q}}(\mathbb{C})$. In other words, $t_\ell = t$.

However, by taking $\gamma_0 \in G_\ell^{1,\mathrm{Zar}}$, we find that $t_\ell \in \mathbb{Q}_\ell$. Since ℓ and the isomorphism $\overline{\mathbb{Q}}_\ell \simeq \mathbb{C}$ were arbitrary and t is an algebraic integer, t must belong to an everywhere unramified number field. By Minkowski's theorem, we have $t \in \mathbb{Z}$. □

Remark 3.3. For any fixed Hodge numbers, the Sato–Tate axioms limit the group G to one of finitely many subgroups of the ambient group $H = \mathrm{USp}(d)$ (if w is odd) or $H = \mathrm{O}(d)$ (if w is even) up to conjugacy, as follows.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 1. Groups satisfying the Sato–Tate axioms with $w = 1$, $h^{0,1} = h^{1,0} = 2$.

G^0	G
U(1)	$\begin{cases} C_1, C_2, C_3, C_4, C_6, D_2, D_3, D_4, D_6, T, O, \\ J(C_1), J(C_2), J(C_3), J(C_4), J(C_6), \\ J(D_2), J(D_3), J(D_4), J(D_6), J(T), J(O), \\ C_{2,1}, C_{4,1}, C_{6,1}, D_{2,1}, D_{3,2}, D_{4,1}, D_{4,2}, D_{6,1}, D_{6,2}, O_1 \end{cases}$
SU(2)	$E_1, E_2, E_3, E_4, E_6, J(E_1), J(E_2), J(E_3), J(E_4), J(E_6)$
U(1) × U(1)	$F, F_a, F_c, F_{a,b}, F_{ab}, F_{ac}, F_{ab,c}, F_{a,b,c}$
U(1) × SU(2)	$\mathrm{U}(1) \times \mathrm{SU}(2)$, $N(\mathrm{U}(1) \times \mathrm{SU}(2))$
SU(2) × SU(2)	$\mathrm{SU}(2) \times \mathrm{SU}(2)$, $N(\mathrm{SU}(2) \times \mathrm{SU}(2))$
USp(4)	USp(4)

Conditions ST1 and ST2 already suffice to limit the connected part G^0 of G to one of finitely many subgroups of H up to conjugacy, as in the proof of Lemma 3.7; we may thus fix a choice of G^0 hereafter. Let N denote the normalizer of G^0 in H; note that N contains G.

Using Tannaka–Krein duality, we may find a representation $\rho : \mathrm{GL}_d(\mathbb{C}) \to V_{\mathbb{C}}$ whose restriction to N contains a factor on which G^0 is trivial and N/G^0 acts faithfully. Condition ST3 then implies that the exponent of G/G^0 may be bounded independently of G. By this fact plus Jordan's theorem on finite linear groups, we may also bound the order of G/G^0 independently of G.

To finish the argument, it is enough to verify that any compact Lie group K only contains finitely many conjugacy classes of subgroups of any given finite order (and then apply this to $K = N/G^0$). This is a result of Weil [Wei64]; see also [Ser92, p. 120, Exercise 1(b)].

3.2 Classification in dimension 2: overview

Our next goal is to establish the following theorem, which will imply Theorem 1.2 thanks to Proposition 3.2 and Theorem 2.16.

Theorem 3.4. *Let G be a group which satisfies the Sato–Tate axioms with $w = 1$ and $h^{0,1} = h^{1,0} = 2$. Let G^0 be the connected part of G. Then G is conjugate to one of the groups listed in Table 1. (The notation in this table is defined within the proof.)*

During the course of the classification, we will obtain explicit presentations[8] for each group; these will be used later to compute Galois types (see § 4) and moments (see § 5.1). These computations will imply that no two of the groups listed in Theorem 3.4 are conjugate to each other.

Remark 3.5. If the number field k has a real place, one can strengthen the Hodge condition (see [Ser12, § 8.3.4]): there must exist $\gamma \in G$ such that $\gamma^2 = -1$, $\mathrm{Trace}(\gamma)$ is equal[9] to 0,

[8] The need for these presentations is the main reason we do not make more use of the exceptional isomorphism $\mathrm{USp}(4)/\{\pm 1\} \simeq \mathrm{SO}(5)$ in our classification; it leads quickly to the list of groups but not to these particular presentations.

[9] This must be modified if one considers weights greater than 1. See [Ser12, § 8.2.3.4].

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

and $\gamma\theta(u)\gamma^{-1} = \theta(u^{-1})$ for all $u \in \mathrm{U}(1)$. This extra condition implies that the groups

$$C_1, C_2, C_3, C_4, C_6, D_2, D_3, D_4, D_6, T, O, F, F_a, F_c, \mathrm{U}(1) \times \mathrm{SU}(2)$$

cannot occur when k has a real place. We will recover and refine this statement in § 4.7.

Remark 3.6. For the purposes of this computation, we will assume that the symplectic form preserved by $\mathrm{USp}(2g)$ is defined by the block matrix

$$S := \begin{pmatrix} 0 & \mathrm{Id}_2 \\ -\mathrm{Id}_2 & 0 \end{pmatrix}$$

unless otherwise specified (as is the case in § 3.6).

3.3 The identity connected component

The remainder of this section will be taken up with the proof of Theorem 3.4; we thus assume until the end of § 3 that $w = 1$ and $g = h^{1,0} = h^{0,1} = 2$. We begin by enumerating the options for the identity component G^0; this classification is well-known in the context of Mumford–Tate groups.

Lemma 3.7. *If G satisfies the Sato–Tate axioms for $w = 1$ and $g = h^{1,0} = h^{0,1} = 2$, then G^0 is conjugate to one of*

$$\mathrm{U}(1),\ \mathrm{SU}(2),\ \mathrm{U}(1) \times \mathrm{U}(1),\ \mathrm{U}(1) \times \mathrm{SU}(2),\ \mathrm{SU}(2) \times \mathrm{SU}(2),\ \mathrm{USp}(4).$$

Proof. We may exploit the exceptional isomorphism $\mathrm{USp}(4)/\{\pm 1\} \simeq \mathrm{SO}(5)$.[10] Let G^0 be a closed connected subgroup of $\mathrm{SO}(5)$. Let T be a maximal torus of G^0; it is contained in a maximal torus of $\mathrm{SO}(5)$, which is 2-dimensional. Let $\mathfrak{h}$ denote the Lie algebra of G^0. By the classification of Dynkin diagrams, the complexification $\mathfrak{h}_{\mathbb{C}}$ of $\mathfrak{h}$ must be isomorphic to one of

$$\begin{array}{ll} \mathfrak{t}_1,\ \mathfrak{sl}_2 = \mathfrak{so}_3 & (\dim(T) = 1), \\ \mathfrak{t}_2,\ \mathfrak{t}_1 \times \mathfrak{sl}_2,\ \mathfrak{sl}_2 \times \mathfrak{sl}_2 = \mathfrak{so}_4,\ \mathfrak{sl}_3,\ \mathfrak{so}_5,\ \mathfrak{g}_2 & (\dim(T) = 2). \end{array}$$

The standard representation of G gives rise to 5-dimensional self-dual orthogonal representations of $\mathfrak{h}$ and $\mathfrak{h}_{\mathbb{C}}$. This immediately rules out $\mathfrak{g}_2$, because the smallest dimension of a nontrivial representation of $\mathfrak{g}_2$ is $7 > 5$. It also rules out $\mathfrak{sl}_3$, because its only nontrivial representation of dimension at most 5 is the standard 3-dimensional representation, which is not self-dual. □

3.4 The case of $G^0 = \mathrm{U}(1)$

We now treat the case where $G^0 = \mathrm{U}(1) = \{u \in \mathbb{C}^\times : |u| = 1\}$. In this case, G^0 must be equal to a Hodge circle, which we may take to be the image of $\mathrm{U}(1)$ under the homomorphism given in block form by

$$u \mapsto \begin{pmatrix} \mathrm{diag}(u) & 0 \\ 0 & \mathrm{diag}(u^{-1}) \end{pmatrix}.$$

Note that the centralizer of G^0 within $\mathrm{GL}(4, \mathbb{C})$ consists of block diagonal matrices $\left(\begin{smallmatrix} A & 0 \\ 0 & D \end{smallmatrix}\right)$. For such a matrix to be unitary, we must have $A, D \in \mathrm{U}(2)$. For such a matrix to also be symplectic, we must have $D = \overline{A}$ (where the bar denotes complex conjugation). We thus conclude that the

[10] Serre kindly pointed out to us that given an element of $\mathrm{USp}(4)$ with characteristic polynomial $T^4 + a_1T^3 + a_2T^2 + a_1T + 1$, if u and v are the angles defining the corresponding element of $\mathrm{SO}(5)$, then $a_1^2 = 4(1 + \cos u)(1 + \cos v)$ and $a_2 = 2(1 + \cos u + \cos v)$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

centralizer Z of G^0 in $\mathrm{USp}(4)$ is isomorphic to $\mathrm{U}(2)$ via the map

$$A \mapsto \begin{pmatrix} A & 0 \\ 0 & \overline{A} \end{pmatrix}. \tag{3.1}$$

The normalizer N of G^0 in $\mathrm{USp}(4)$ has the form

$$N = Z \cup JZ, \quad J := \begin{pmatrix} 0 & J_2 \\ -J_2 & 0 \end{pmatrix}, \quad J_2 := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that J centralizes the copy of $\mathrm{SU}(2)$ inside our embedded $\mathrm{U}(2)$: for any $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SU}(2)$ we have

$$A = (\overline{A}^T)^{-1} = (\overline{A}^{-1})^T = \begin{pmatrix} \overline{d} & -\overline{b} \\ -\overline{c} & \overline{a} \end{pmatrix}^T = \begin{pmatrix} \overline{d} & -\overline{c} \\ -\overline{b} & \overline{a} \end{pmatrix}$$

and therefore

$$J \begin{pmatrix} A & 0 \\ 0 & \overline{A} \end{pmatrix} J^{-1} = \begin{pmatrix} J_2 \overline{A} J_2 & 0 \\ 0 & J_2 A J_2 \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & \overline{A} \end{pmatrix}.$$

Consequently, we have

$$N/G^0 \simeq \mathrm{SU}(2)/(\pm 1) \times \mathbb{Z}/2\mathbb{Z} \simeq \mathrm{SO}(3) \times \mathbb{Z}/2\mathbb{Z}. \tag{3.2}$$

We first enumerate the options for G assuming that $G \subseteq Z$; given (3.2), this constitutes an enumeration over the familiar list of finite subgroups of $\mathrm{SO}(3)$ up to conjugacy. It will be convenient to identify $\mathrm{SU}(2)$ with the group of unit quaternions via the isomorphism

$$a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \mapsto \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix},$$

and to then use (3.1) to view the unit quaternions as a subgroup of $\mathrm{USp}(4)$.

Cyclic groups. A cyclic group of order n within $\mathrm{SO}(3)$ lifts to a cyclic group of order $2n$ in $\mathrm{SU}(2)$, which can be represented as $\langle \zeta_{2n} \rangle$ where $\zeta_{2n} = \cos(\pi/n) + \sin(\pi/n)\mathbf{i}$. By the rationality condition, the average over $r \in [0, 1]$ of the square of the trace of the matrix

$$B = \begin{pmatrix} A & 0 \\ 0 & \overline{A} \end{pmatrix}, \quad A = e^{2\pi i r} \zeta_{2n} = \begin{pmatrix} e^{2\pi i r + \pi i/n} & 0 \\ 0 & e^{2\pi i r - \pi i/n} \end{pmatrix}$$

is an integer. However, this average equals $|\mathrm{Trace}(A)|^2 = (2\cos(\pi/n))^2$, so we must have $2\cos(\pi/n) \in \{0, \pm 1, \pm\sqrt{2}, \pm\sqrt{3}, \pm 2\}$, or $n \in \{1, 2, 3, 4, 6\}$. For these values of n, let C_n denote the resulting subgroup of Z.

Dihedral groups. A dihedral group of order $2n$ within $\mathrm{SO}(3)$ lifts to the group $\langle \zeta_{2n}, \mathbf{j} \rangle$ in $\mathrm{SU}(2)$. For $n = 2, 3, 4, 6$, let D_n denote the resulting subgroup of Z. We omit D_1 since it is conjugate to C_2.

Tetrahedral group. In this case, G/G^0 is the tetrahedral group. Lifting to $\mathrm{SU}(2)$ yields the binary tetrahedral group, which has the standard presentation

$$\{\pm 1, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}, \tfrac{1}{2}(\pm 1 \pm \mathbf{i} \pm \mathbf{j} \pm \mathbf{k})\}.$$

Let T denote the resulting subgroup of Z.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Octahedral group. In this case, G/G^0 is the octahedral group. Lifting to SU(2) yields the binary octahedral group, a standard presentation of which consists of the given presentation of the binary tetrahedral group together with

$$\frac{\sqrt{2}}{2}(\pm 1 \pm \mathbf{i}),\ \frac{\sqrt{2}}{2}(\pm 1 \pm \mathbf{j}),\ \frac{\sqrt{2}}{2}(\pm 1 \pm \mathbf{k}),\ \frac{\sqrt{2}}{2}(\pm \mathbf{i} \pm \mathbf{j}),\ \frac{\sqrt{2}}{2}(\pm \mathbf{i} \pm \mathbf{k}),\ \frac{\sqrt{2}}{2}(\pm \mathbf{j} \pm \mathbf{k}).$$

Let O denote the resulting subgroup of Z.

Icosahedral group. This group cannot occur because it contains a cyclic group of order 5, but C_5 does not satisfy the rationality condition.

We thus obtain the following groups with $G^0 = \mathrm{U}(1)$ and $G \subseteq Z$:

$$C_1, C_2, C_3, C_4, C_6, D_2, D_3, D_4, D_6, T, O. \tag{3.3}$$

We now determine the options for $G \not\subseteq Z$. Since $N/G^0 \simeq \mathrm{SO}(3) \times \mathbb{Z}/2\mathbb{Z}$, the projection of a subgroup of N/G^0 to SO(3) is either two-to-one or one-to-one onto its image H. In the former case, we get a subgroup of the form $H \times \mathbb{Z}/2\mathbb{Z}$, corresponding to one of the groups

$$J(C_1), J(C_2), J(C_3), J(C_4), J(C_6), J(D_2), J(D_3), J(D_4), J(D_6), J(T), J(O)$$

obtained by adjoining J to each group in (3.3). In the latter case, the subgroup forms the graph of a homomorphism $H \to \mathbb{Z}/2\mathbb{Z}$, which must be nontrivial because $G \not\subseteq Z$. We need only consider these homomorphisms up to conjugation within SO(3); this gives some additional groups as follows.

Cyclic groups. For $n = 2, 4, 6$, the nontrivial homomorphism $H \to \mathbb{Z}/2\mathbb{Z}$ gives

$$C_{n,1} := \langle \mathrm{U}(1), J(\cos(\pi/n) + \sin(\pi/n)\mathbf{i}) \rangle.$$

Beware: $C_{6,1}$ contains $C_{2,1}$ but $C_{4,1}$ does not (its subgroup of order 2 is C_2).

Dihedral groups. For $n = 2, 4, 6$, there are two nontrivial homomorphisms $H \to \mathbb{Z}/2\mathbb{Z}$ not killing the cyclic subgroup, which are interchanged by an outer automorphism of H. These give rise to the group

$$D_{n,1} := \langle \mathrm{U}(1), J(\cos(\pi/n) + \sin(\pi/n)\mathbf{i}), \mathbf{j} \rangle$$

containing $C_{n,1}$ with index 2. For $n = 3, 4, 6$, we may also use the nontrivial homomorphism killing the cyclic subgroup to obtain

$$D_{n,2} := \langle \mathrm{U}(1), \cos(\pi/n) + \sin(\pi/n)\mathbf{i}, J\mathbf{j} \rangle;$$

this is redundant for $n = 2$ because all three of the nontrivial homomorphisms $H \to \mathbb{Z}/2\mathbb{Z}$ are conjugated transitively by SO(3).

Tetrahedral group. In this case, we have $H \simeq \mathrm{A}_4$, which has no nontrivial homomorphisms to $\mathbb{Z}/2\mathbb{Z}$.

Octahedral group. In this case, we have $H \simeq \mathrm{S}_4$, so there is one nontrivial homomorphism to $\mathbb{Z}/2\mathbb{Z}$ with the tetrahedral group in its kernel. We obtain a new group O_1 by multiplying each of the elements of $O \backslash T$ by J.

This analysis thus adds the additional groups

$$C_{2,1}, C_{4,1}, C_{6,1}, D_{2,1}, D_{3,2}, D_{4,1}, D_{4,2}, D_{6,1}, D_{6,2}, O_1.$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

3.5 The case of $G^0 = \mathrm{SU}(2)$

After the $G^0 = \mathrm{U}(1)$ case, the next most complicated case is that of $G^0 = \mathrm{SU}(2)$. Fortunately, we can reuse some of the analysis from § 3.4 as follows.

Embed G^0 into $\mathrm{USp}(4)$ as in (3.1). Since $\mathrm{SU}(2)$ is centralized by $\mathrm{U}(1)$, the normalizer is again $N = Z \cup JZ$. This time, we see that $Z/G^0 \simeq \mathrm{U}(2)/\mathrm{SU}(2) \simeq \mathrm{U}(1)/(\pm 1)$, and that conjugation by J acts on Z/G^0 by inversion. We thus need only list the finite subgroups of $\mathrm{O}(2)$, which is straightforward: for each positive integer n, the cyclic subgroup of $\mathrm{U}(1)$ of order $2n$ gives rise to the groups

$$E_n := \langle \mathrm{SU}(2), e^{\pi i/n} \rangle,$$
$$J(E_n) := \langle \mathrm{SU}(2), e^{\pi i/n}, J \rangle.$$

It is easy to check that these groups satisfy the rationality condition if and only if $n = 1, 2, 3, 4, 6$. We thus have the groups

$$E_1, E_2, E_3, E_4, E_6, J(E_1), J(E_2), J(E_3), J(E_4), J(E_6).$$

3.6 The remaining cases for G^0

In order to complete the proof of Theorem 3.4, by Lemma 3.7 it remains to consider the cases where $G^0 = \mathrm{U}(1) \times \mathrm{U}(1)$, $\mathrm{U}(1) \times \mathrm{SU}(2)$, $\mathrm{SU}(2) \times \mathrm{SU}(2)$, or $\mathrm{USp}(4)$. The case of $G^0 = \mathrm{USp}(4)$ is trivial because we must have $G = G^0$, so we focus on the other three cases. For these cases, it is convenient to change basis to account for the product structure of G^0, by interchanging the second and third rows and columns; we are thus working with the new symplectic form

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

In these coordinates, we may take G^0 to be embedded into $\mathrm{USp}(4)$ in block form. We take $\mathrm{U}(1)$ to be embedded into $\mathrm{SU}(2)$ via the map

$$u \mapsto \begin{pmatrix} u & 0 \\ 0 & \overline{u} \end{pmatrix}.$$

For $G^0 = \mathrm{U}(1) \times \mathrm{U}(1)$, the normalizer in $\mathrm{USp}(4)$ contains $\mathrm{U}(1) \times \mathrm{U}(1)$ with index 8, with the quotient (isomorphic to a dihedral group) generated by matrices

$$a := \begin{pmatrix} J_2 & 0 \\ 0 & \mathrm{Id}_2 \end{pmatrix}, \quad b := \begin{pmatrix} \mathrm{Id}_2 & 0 \\ 0 & J_2 \end{pmatrix}, \quad c := \begin{pmatrix} 0 & \mathrm{Id}_2 \\ -\mathrm{Id}_2 & 0 \end{pmatrix},$$

each of which defines an involution on the component group. We write F_* for the group generated by G^0 and a list $*$ of matrices generated by a, b, c. In this notation, up to conjugacy, we have the groups

$$F, F_a, F_c, F_{a,b}, F_{ab}, F_{ac}, F_{ab,c}, F_{a,b,c}.$$

For $G^0 = \mathrm{U}(1) \times \mathrm{SU}(2)$, the normalizer in $\mathrm{USp}(4)$ equals $N(\mathrm{U}(1)) \times \mathrm{SU}(2)$. Thus G^0 and its normalizer are the only possible groups.

For $G^0 = \mathrm{SU}(2) \times \mathrm{SU}(2)$, the normalizer in $\mathrm{USp}(4)$ consists of $\mathrm{SU}(2) \times \mathrm{SU}(2)$ plus the coset generated by J. In this case, G^0 and its normalizer are the only possible groups. This completes the proof of Theorem 3.4.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

4. Galois structures of abelian surfaces

In this section, we give the classification of *Galois types* of abelian surfaces (as introduced in Definition 1.3) and the relation of these to Sato–Tate groups. Our main result is Theorem 4.3, which implies both Theorem 1.4 and Theorem 1.5. It gives an alternate description of the Galois type in terms of arithmetic properties of the abelian surface. Strictly speaking, only a small part of this description is needed in order to obtain Theorems 1.4 and 1.5 (namely the analysis of cases corresponding to Sato–Tate groups with connected part $\mathrm{U}(1) \times \mathrm{U}(1)$). However, we have chosen to provide the complete analysis in order to make it easier to recognize Galois types and Sato–Tate groups of abelian surfaces occurring in nature.

Before stating Theorem 4.3, we recall the definition of the Galois type and set some associated notation. For the moment, we take A to be any abelian variety over a number field k.

Proposition 4.1 [Sil92]. *There is a* unique minimal extension K/k *over which all endomorphisms of* $A_{\overline{\mathbb{Q}}}$ *are defined. The extension* K/k *is normal and unramified at the prime ideals of* k *at which* A *has good or semistable reduction.*

Taking K as in Proposition 4.1, $\mathrm{End}(A_K)_{\mathbb{Q}}$ is a semisimple algebra of finite rank over $\mathbb{Q}$ and thus decomposes as a product $\mathrm{End}(A_K)_{\mathbb{Q}} = \prod_i \mathrm{M}_{n_i}(D_i)$ of matrix algebras over division algebras, parallel to the decomposition of A as a product $A \sim \prod_i A_i^{n_i}$ of simple varieties over K.

Note that the Galois group $\mathrm{Gal}(K/k)$ acts in a natural way on the $\mathbb{Q}$-algebra $\mathrm{End}(A_K)_{\mathbb{Q}}$ of endomorphisms of A and induces a Galois representation

$$\rho_A : \mathrm{Gal}(K/k) \hookrightarrow \mathrm{Aut}_{\mathbb{R}\text{-alg}}(\mathrm{End}(A_K)_{\mathbb{R}}),$$

which is faithful precisely because K/k is the minimal extension over which the endomorphisms of $A_{\overline{\mathbb{Q}}}$ are defined.

Definition 4.2. The *Galois type* of A is the equivalence class of the representation ρ_A.

As noted in Definition 1.3, two abelian varieties A/k and A'/k' defined over different number fields may have the same Galois type; the equivalence relation on representations is meant to see $\mathrm{Gal}(K/k)$ only as an abstract group, not as a quotient of G_k.

4.1 Classification of Galois types: overview

We now restrict to the case where A is an abelian surface over k, and formulate the classification theorem for Galois types. In the process, we introduce alternate names for the Galois types corresponding more closely to their arithmetic.

To begin with, recall that Albert's classification of division algebras with involution (see [Mum70]), together with the work of Shimura [Shi63], shows that the $\mathbb{R}$-algebra $\mathrm{End}(A_K)_{\mathbb{R}}$ is isomorphic to one of the following:

(**A**) $\mathbb{R}$, which is the generic case;

(**B**) $\mathbb{R} \times \mathbb{R}$, which occurs when either:
- A_K is isogenous to a product of nonisogenous elliptic curves without CM; or
- A_K is simple and $\mathrm{End}(A_K)$ is an order in a real quadratic field;

(**C**) $\mathbb{C} \times \mathbb{R}$, which occurs when A_K is isogenous to a product of (necessarily nonisogenous) elliptic curves, one with CM and the other without CM;

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

(**D**) $\mathbb{C} \times \mathbb{C}$, which occurs when either:
- A_K is isogenous to a product of nonisogenous elliptic curves with CM; or
- A_K is simple and $\mathrm{End}(A_K)$ is an order in a quartic CM-field;

(**E**) $\mathrm{M}_2(\mathbb{R})$, which occurs when either:
- A_K is isogenous to the square of an elliptic curve without CM; or
- A_K is simple and $\mathrm{End}(A_K)$ is an order in a division quaternion algebra over $\mathbb{Q}$;

(**F**) $\mathrm{M}_2(\mathbb{C})$, which occurs when A_K is isogenous to the square of an elliptic curve with CM.

In case **D**, when $\mathrm{End}(A_K)$ is an order in a quartic CM-field M, we shall assume that a choice of an isomorphism $\iota : M \xrightarrow{\sim} \mathrm{End}(A_K)_{\mathbb{Q}}$ has been made; this singles out a (primitive) CM-type Φ on A, to which we can associate the *reflex field* of the pair (A, Φ), which we denote as usual by M^*. Different choices of Φ give rise to conjugate reflex fields in the Galois closure of M, and our results depend only on the conjugacy class of M^*.

We shall refer to **A**, **B**, **C**, **D**, **E**, or **F** as the *absolute type*, or simply the *type*, of A. Note that the Galois type of A is a much finer invariant.

The six absolute types are in one-to-one correspondence with the six *connected* Lie subgroups of $\mathrm{USp}(4)$ appearing in Lemma 3.7, as indicated below.

A: $\mathrm{USp}(4)$	**B**: $\mathrm{SU}(2) \times \mathrm{SU}(2)$	**C**: $\mathrm{U}(1) \times \mathrm{SU}(2)$
D: $\mathrm{U}(1) \times \mathrm{U}(1)$	**E**: $\mathrm{SU}(2)$	**F**: $\mathrm{U}(1)$

There at least two ways of proving this. One method, which we do not make explicit here but surely follows from existing results in the literature, uses Definition 2.6 to compute $\mathrm{ST}_A^0 = \mathrm{ST}_{A_K}$ for any abelian surface A of given absolute type. Alternatively, we may work in the reverse direction: for each of the six possible connected Sato–Tate groups, use Proposition 2.19 to determine the corresponding Galois type. These computations are made explicit in §§ 4.2–4.5.

Theorem 4.3. *There are exactly* 52 *different Galois types of abelian surfaces, and these correspond to* 52 *of the* 55 *Sato–Tate groups listed in Theorem 3.4, as indicated below (using notation defined in the proof). Of the* 52 *Galois types, exactly* 34 *can (and do) arise from abelian surfaces defined over* $\mathbb{Q}$*; these are decorated with the symbol* $\star$*.*

- $\mathbf{A}[\mathrm{C}_1]^\star$, *matching* $\mathrm{USp}(4)$.
- $\mathbf{B}[\mathrm{C}_1]^\star$ *and* $\mathbf{B}[\mathrm{C}_2]^\star$, *matching* $\mathrm{SU}(2) \times \mathrm{SU}(2)$ *and* $N(\mathrm{SU}(2) \times \mathrm{SU}(2))$.
- $\mathbf{C}[\mathrm{C}_1]$ *and* $\mathbf{C}[\mathrm{C}_2]^\star$, *matching* $\mathrm{U}(1) \times \mathrm{SU}(2)$ *and* $N(\mathrm{U}(1) \times \mathrm{SU}(2))$.
- $\mathbf{D}[\mathrm{C}_1]$, $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{C}]$, $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$, $\mathbf{D}[\mathrm{C}_4]^\star$, *and* $\mathbf{D}[\mathrm{D}_2]^\star$, *matching* F, F_a, F_{ab}, F_{ac}, *and* $F_{a,b}$.
- $\mathbf{E}[\mathrm{C}_1]^\star$, $\mathbf{E}[\mathrm{C}_2, \mathbb{C}]^\star$, $\mathbf{E}[\mathrm{C}_3]^\star$, $\mathbf{E}[\mathrm{C}_4]^\star$, *and* $\mathbf{E}[\mathrm{C}_6]^\star$, *matching* E_1, E_2, E_3, E_4, *and* E_6.
- $\mathbf{E}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]^\star$, $\mathbf{E}[\mathrm{D}_2]^\star$, $\mathbf{E}[\mathrm{D}_3]^\star$, $\mathbf{E}[\mathrm{D}_4]^\star$, *and* $\mathbf{E}[\mathrm{D}_6]^\star$, *matching* $J(E_1)$, $J(E_2)$, $J(E_3)$, $J(E_4)$, *and* $J(E_6)$.
- $\mathbf{F}[\mathrm{C}_1]$, $\mathbf{F}[\mathrm{C}_2]$, $\mathbf{F}[\mathrm{C}_3]$, $\mathbf{F}[\mathrm{C}_4]$, $\mathbf{F}[\mathrm{C}_6]$, $\mathbf{F}[\mathrm{D}_2]$, $\mathbf{F}[\mathrm{D}_3]$, $\mathbf{F}[\mathrm{D}_4]$, $\mathbf{F}[\mathrm{D}_6]$, $\mathbf{F}[\mathrm{A}_4]$, *and* $\mathbf{F}[\mathrm{S}_4]$, *matching* C_1, C_2, C_3, C_4, C_6, D_2, D_3, D_4, D_6, T, *and* O.
- $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$, $\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathbb{H}]^\star$, $\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$, $\mathbf{F}[\mathrm{C}_4 \times \mathrm{C}_2, \mathrm{C}_4]^\star$, $\mathbf{F}[\mathrm{C}_6 \times \mathrm{C}_2, \mathrm{C}_6]^\star$, $\mathbf{F}[\mathrm{D}_2 \times \mathrm{C}_2, \mathrm{D}_2]^\star$, $\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathbb{H}]^\star$, $\mathbf{F}[\mathrm{D}_4 \times \mathrm{C}_2, \mathrm{D}_4]^\star$, $\mathbf{F}[\mathrm{D}_6 \times \mathrm{C}_2, \mathrm{D}_6]^\star$, $\mathbf{F}[\mathrm{A}_4 \times \mathrm{C}_2, \mathrm{A}_4]^\star$, *and* $\mathbf{F}[\mathrm{S}_4 \times \mathrm{C}_2, \mathrm{S}_4]^\star$, *matching* $J(C_1)$, $J(C_2)$, $J(C_3)$, $J(C_4)$, $J(C_6)$, $J(D_2)$, $J(D_3)$, $J(D_4)$, $J(D_6)$, $J(T)$, *and* $J(O)$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

– $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathrm{M}_2(\mathbb{R})]^\star$, $\mathbf{F}[\mathrm{C}_4, \mathrm{C}_2]$, $\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathrm{M}_2(\mathbb{R})]^\star$, $\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathrm{M}_2(\mathbb{R})]^\star$, $\mathbf{F}[\mathrm{D}_4, \mathrm{D}_2]^\star$, $\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathrm{M}_2(\mathbb{R})]^\star$, $\mathbf{F}[\mathrm{D}_3, \mathrm{C}_3]^\star$, $\mathbf{F}[\mathrm{D}_4, \mathrm{C}_4]^\star$, $\mathbf{F}[\mathrm{D}_6, \mathrm{C}_6]^\star$, *and* $\mathbf{F}[\mathrm{S}_4, \mathrm{A}_4]^\star$, *matching* $C_{2,1}$, $C_{4,1}$, $C_{6,1}$, $D_{2,1}$, $D_{4,1}$, $D_{6,1}$, $D_{3,2}$, $D_{4,2}$, $D_{6,2}$, *and* O_1.

Moreover, for any abelian surface A defined over a number field k, each of the following three invariants uniquely determines the other two:

(a) *the conjugacy class of* ST_A *within* $\mathrm{USp}(4)$;

(b) *the Galois type of A;*

(c) *the isomorphism class of* $\mathrm{Gal}(K/k)$ *plus the function on the subgroup lattice of* $\mathrm{Gal}(K/k)$ *taking the subgroup H to the isomorphism class of the $\mathbb{R}$-algebra* $\mathrm{End}(A_K)_{\mathbb{R}}^H$ *fixed by H.*

We devote the remainder of this section to proving Theorem 4.3. In §§ 4.2–4.5, we prove that there exist at most 52 different Galois types over number fields. Along the way, we describe the passage from (a) to (b) in Theorem 4.3; this amounts to making the proof of Proposition 2.19 explicit for each of the 55 groups named in Theorem 3.4. From this computation, we see that the Galois types correspond to 52 of the 55 Lie groups listed in Theorem 3.4. Since the Lie groups F_c, $F_{ab,c}$, and $F_{a,b,c}$ remain unmatched, they cannot occur as the Sato–Tate group of any abelian surface.

As a byproduct of this computation, we explicitly describe the passage from (a) to (c) in Theorem 4.3. To do this, it suffices to compute $\mathrm{End}(A_K)_{\mathbb{R}}^{\mathrm{Gal}(K/k)} = \mathrm{End}(A_k)_{\mathbb{R}}$ for each Sato–Tate group; this data appears in Table 8. With this data, it is also easy to go from (c) back to (a); see § 4.6.

In § 4.7, we verify that the 18 Galois types that are *not* decorated with a $\star$ in Theorem 4.3 *cannot* arise from an abelian surface defined over $\mathbb{Q}$. Among these 18 Galois types, 14 correspond to Sato–Tate groups that were already ruled out over a field with a real place in Remark 3.5. Additional arguments are provided to show that the three Galois types $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$, $\mathbf{F}[\mathrm{C}_4, \mathrm{C}_2]$, and $\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$ cannot occur over a field with a real place (Proposition 4.11). Finally, we show that the Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$, admissible over a field admitting a real place, cannot arise from an abelian surface defined over $\mathbb{Q}$, as a result of the discussion in § 4.4.

In § 4.8, we exhibit one proven example of an abelian surface for each of the 52 Galois types. These examples arise as Jacobians of curves of genus 2 over number fields; for those 34 Galois types decorated with a $\star$, the curve that we exhibit is defined over $\mathbb{Q}$. For the Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$, the curve we present is defined over a totally real field.

We now proceed with the proof of Theorem 4.3 in §§ 4.2–4.6.

4.2 Cases A and B

In case **A**, the group $\mathrm{Aut}_{\mathbb{Q}}(\mathrm{End}(A_K)_{\mathbb{Q}})$ is trivial and $\mathrm{End}(A_K)_{\mathbb{R}} = \mathbb{R}$. Therefore $\mathrm{Gal}(K/k) = \mathrm{C}_1$ and $\rho_A = \chi_1$ is the trivial representation. In case **B**, we have $\mathrm{Aut}_{\mathbb{Q}}(\mathrm{End}(A_K)_{\mathbb{R}}) \simeq \mathrm{C}_2$, $\mathrm{End}(A_K)_{\mathbb{R}} \simeq \mathbb{R} \times \mathbb{R}$, and $\mathrm{End}(A_K)_{\mathbb{R}}^{\mathrm{C}_2} \simeq \mathbb{R}$. This yields three distinct Galois types: $\mathbf{A}[\mathrm{C}_1]$, $\mathbf{B}[\mathrm{C}_1]$, and $\mathbf{B}[\mathrm{C}_2]$.

From the Lie group side, it is clear that if $\mathrm{ST}_A = \mathrm{USp}(4)$, then $\mathrm{End}(A_K)_{\mathbb{R}}$ is $\mathbb{R}$. If $\mathrm{ST}_A^0 = \mathrm{SU}(2) \times \mathrm{SU}(2)$, then $\mathrm{End}(A_K)_{\mathbb{R}} \simeq \mathbb{R} \times \mathbb{R}$ and the normalizer of $\mathrm{SU}(2) \times \mathrm{SU}(2)$ interchanges the two factors, fixing $\mathbb{R}$.

Thus the Galois types of abelian surfaces with Sato–Tate groups $\mathrm{USp}(4)$, $\mathrm{SU}(2) \times \mathrm{SU}(2)$, and $N(\mathrm{SU}(2) \times \mathrm{SU}(2))$ are $\mathbf{A}[\mathrm{C}_1]$, $\mathbf{B}[\mathrm{C}_1]$, and $\mathbf{B}[\mathrm{C}_2]$, respectively.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

4.3 Complex multiplication by a quartic CM-field

Let M be a quartic CM-field, that is, a totally imaginary quadratic extension of a real quadratic field. Assume A has complex multiplication by M and fix an isomorphism $\iota : M \longrightarrow \mathrm{End}(A_K)_{\mathbb{Q}}$, which in turn induces an isomorphism

$$M \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{C} \times \mathbb{C} \simeq \mathrm{End}(A_K)_{\mathbb{R}}.$$

Let $\Phi = \{\phi, \phi'\}$ denote the CM-type of the pair (A, ι), which is necessarily primitive as otherwise ι would only be a (nonsurjective) monomorphism. Let M^* be the reflex field of (M, Φ). The extension $M/\mathbb{Q}$ is either:

(C_4) normal, with $\mathrm{Gal}(M/\mathbb{Q}) \simeq \mathrm{C}_4$ and $M^* = M$; or

(D_4) not normal, with the Galois group of the normal closure $\tilde{M}$ isomorphic to D_4 and M^* a subfield of $\tilde{M}$ of degree 4 over $\mathbb{Q}$, different from M and not normal over $\mathbb{Q}$; see [Shi98, p. 64] or [Str10, ch. I, § 7], for example.

Proposition 4.4. *The field K is the compositum of k and M^*.*

Proof. This is [Shi71, p. 515, Proposition 3]. □

In case (D_4) we must have $|\mathrm{Gal}(K/k)| \leqslant 2$, because otherwise $\mathrm{Aut}_{\mathbb{Q}}(M) \supseteq \mathrm{Gal}(K/k)$ would have order at least 4, implying that $M/\mathbb{Q}$ is Galois. Note that the condition $[kM^* : k] \leqslant 2$ implies that this case cannot occur for $k = \mathbb{Q}$. In case (C_4), we have $\mathrm{Gal}(K/k) = \mathrm{Gal}(kM^*/k) \subseteq \mathrm{Gal}(M^*/\mathbb{Q}) \simeq \mathrm{C}_4$. In any case, $\mathrm{Gal}(K/k) \simeq \mathrm{C}_n$ for $n = 1$, 2, or 4, which gives rise to the following alternatives.

- $\mathrm{Gal}(K/k) = \mathrm{C}_1$ and ρ_A is the trivial representation; this yields Galois type $\mathbf{D}[\mathrm{C}_1]$.
- $\mathrm{Gal}(K/k) = \mathrm{C}_2$ and $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_2} \simeq \mathbb{R} \times \mathbb{R}$, as $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_2} \simeq M^{\mathrm{C}_2} \otimes_{\mathbb{Q}} \mathbb{R}$ is by Artin's lemma an $\mathbb{R}$-vector space of dimension 2 and M contains a single quadratic subfield, which is real; this is Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$.
- $\mathrm{Gal}(K/k) = \mathrm{C}_4$, $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_2} \simeq \mathbb{R} \times \mathbb{R}$, and $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_4} \simeq \mathbb{R}$; this is Galois type $\mathbf{D}[C_4]$, which is the only case that can occur when $k = \mathbb{Q}$.

Thus for case **D** we have found three of the five Galois types listed in Theorem 4.3; we shall find the remaining two in the next section, arising from abelian surfaces that are isogenous to the product of two elliptic curves with CM by distinct imaginary quadratic fields.

Let us now analyze which Galois types correspond to an abelian surface A such that $\mathrm{ST}_A^0 = \mathrm{U}(1) \times \mathrm{U}(1)$. Recall that in this case, we take the symplectic form to be in split form rather than block form. With this in mind, the matrices in $\mathrm{M}_4(\mathbb{C})$ commuting with $\mathrm{U}(1) \times \mathrm{U}(1)$ are

$$\left\{ \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix} : a, b, c, d \in \mathbb{C} \right\},$$

and the Rosati form is a scalar multiple of $2ab + 2cd = ((a+b)^2 - (a-b)^2 + (c+d)^2 - (c-d)^2)/2$. Consequently,

$$\mathrm{End}(A_K)_{\mathbb{R}} = \left\{ \begin{pmatrix} a+bi & 0 & 0 & 0 \\ 0 & a-bi & 0 & 0 \\ 0 & 0 & c+di & 0 \\ 0 & 0 & 0 & c-di \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\} \simeq \mathbb{C} \times \mathbb{C}.$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

The action of a generator of the component group of F_a is

$$\begin{pmatrix} a+bi & 0 & 0 & 0 \\ 0 & a-bi & 0 & 0 \\ 0 & 0 & c+di & 0 \\ 0 & 0 & 0 & c-di \end{pmatrix} \mapsto \begin{pmatrix} a-bi & 0 & 0 & 0 \\ 0 & a+bi & 0 & 0 \\ 0 & 0 & c+di & 0 \\ 0 & 0 & 0 & c-di \end{pmatrix},$$

so the fixed ring has $b=0$ and thus is $\mathbb{R}\times\mathbb{C}$. For F_c, we get

$$\begin{pmatrix} a+bi & 0 & 0 & 0 \\ 0 & a-bi & 0 & 0 \\ 0 & 0 & c+di & 0 \\ 0 & 0 & 0 & c-di \end{pmatrix} \mapsto \begin{pmatrix} c+di & 0 & 0 & 0 \\ 0 & c-di & 0 & 0 \\ 0 & 0 & a+bi & 0 \\ 0 & 0 & 0 & a-bi \end{pmatrix},$$

so the fixed ring has $a=c$ and $b=d$ and thus is $\mathbb{C}$. For both F_{ab} and $F_{a,b}$, the fixed ring has $b=d=0$ and thus is $\mathbb{R}\times\mathbb{R}$. For F_{ac}, the fixed ring has $a=c$ and $b=d=0$ and thus is $\mathbb{R}$, and similarly for the larger group $F_{a,b,c}$. For $F_{ab,c}$, the fixed ring is contained in both $\mathbb{R}\times\mathbb{R}$ (the fixed ring of F_{ab}) and $\mathbb{C}$ (the fixed ring of F_c) and thus is $\mathbb{R}$.

Comparing this analysis with the Galois structure of the above three Galois types, we conclude that if the Sato–Tate group of A is F, F_{ab}, or F_{ac}, then the Galois type of A is $\mathbf{D}[\mathrm{C}_1]$, $\mathbf{D}[\mathrm{C}_2,\mathbb{R}\times\mathbb{R}]$, or $\mathbf{D}[\mathrm{C}_4]$, respectively.

4.4 Products of nonisogenous elliptic curves

Proposition 4.5. *For $i=1,2$, let M_i be either $\mathbb{Q}$ or an imaginary quadratic field. Assume that at least one M_i is quadratic and that if both M_1 and M_2 are quadratic, then $M_1\not\simeq M_2$. Let A/k be an abelian surface such that $\mathrm{End}(A_{\overline{\mathbb{Q}}})_{\mathbb{Q}}\simeq M_1\times M_2$. The following hold:*

(i) *the minimal extension of k over which the endomorphisms of $A_{\overline{\mathbb{Q}}}$ are defined is $K=kM_1M_2$;*

(ii) *there exist elliptic curves $\tilde{E}_1$ and $\tilde{E}_2$ over k for which $A\sim_k \tilde{E}_1\times\tilde{E}_2$.*

Proof. (i) Let us prove the statement under the assumption that both M_1 and M_2 are quadratic; when M_1 is quadratic and $M_2=\mathbb{Q}$, the proof is simpler and we leave the details to the reader.

Let $K\subset\overline{\mathbb{Q}}$ denote the minimal extension of k over which all endomorphisms of $A_{\overline{\mathbb{Q}}}$ are defined, and let Ω/k denote a minimal subextension of K/k over which there exists an isogeny

$$\psi : A \xrightarrow{\sim} E_1\times E_2,$$

defined over Ω onto a product of two elliptic curves E_1/Ω and E_2/Ω. Note that Ω might be properly contained in K, as we do not require that $\mathrm{End}(E_{i,\Omega})_{\mathbb{Q}}\simeq M_i$.

For $i=1,2$, we claim that there exists an elliptic curve E_i' over k such that E_i and E_i' are isogenous over Ω. Indeed, let $k\subseteq k_i\subseteq\Omega$ be a minimal subextension of Ω/k over which such an E_i' exists. The abelian surface $A':=E_1'\times E_2'$ is thus defined over k_1k_2. As an application of [Rib04, Theorem 8.2], it follows from the minimality of k_i that k_1k_2 is also a minimal subextension of Ω/k over which A' admits a model up to isogenies over Ω. Indeed, if there were a proper subextension $k_0\subsetneq k_1k_2$ over which A' admits a model, there would exist a collection of isogenies $\{\varphi_\sigma:(A')^\sigma\longrightarrow A'\}_{\sigma\in\mathrm{Gal}(k_1k_2/k_0)}$ defined over Ω such that ${\varphi_\sigma}^\sigma\varphi_\tau=\varphi_{\sigma\tau}$. Since there are no isogenies between E_1 and any of the Galois conjugates of E_2, we would have $\varphi_\sigma=(\varphi_\sigma^1,\varphi_\sigma^2)$ where $\varphi_\sigma^i:E_i^\sigma\longrightarrow E_i$ are isogenies such that ${\varphi_\sigma^i}^\sigma\varphi_\tau^i=\varphi_{\sigma\tau}^i$. Ribet's theorem would then imply that both E_1 and E_2 admit a model over k_0, contradicting the minimality of k_1 and k_2.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Since A is one such model, we deduce that $k = k_1k_2$ and thus $k = k_1 = k_2$. Hence A and A' are abelian surfaces over k that are isogenous over Ω.

By the theory of complex multiplication on elliptic curves (see, for example, [Sil94, Theorem 2.2]), the minimal extension of k over which all endomorphisms of $A'_{\overline{\mathbb{Q}}}$ are defined is kM_1M_2. Since $\mathrm{End}(A'_{\overline{\mathbb{Q}}})_{\mathbb{Q}} = M_1 \times M_2$ is commutative and A is a twist of A' in the category of abelian varieties up to isogenies, the isogeny class of A over k corresponds to a cocycle $c_A \in H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/k), \mathrm{End}(A'_{\overline{\mathbb{Q}}})^\times_{\mathbb{Q}}) = H^1(\mathrm{Gal}(\overline{\mathbb{Q}}/k), M_1^\times \times M_2^\times)$; it follows that for any number field $k \subseteq F \subseteq \overline{\mathbb{Q}}$,

$$\begin{aligned}\mathrm{End}(A_F)_{\mathbb{Q}} &= \{\alpha \in \mathrm{End}(A'_{\overline{\mathbb{Q}}})_{\mathbb{Q}} : \alpha^\sigma c_A(\sigma) = c_A(\sigma)\alpha \ \ \forall \sigma \in G_F\} \\ &= \mathrm{End}(A'_F)_{\mathbb{Q}} \quad \text{(because } M_1 \times M_2 \text{ is commutative).}\end{aligned}$$

Hence the minimal extension of k over which all endomorphisms of $A_{\overline{\mathbb{Q}}}$ are defined is also $K = kM_1M_2$.

(ii) In the proof of (i), we have seen that $A \sim_K E'_1 \times E'_2$ for elliptic curves E'_1 and E'_2 over k. Since $\mathrm{Gal}(K/k) = \mathrm{C}_1, \mathrm{C}_2, \mathrm{D}_2$, the representation $\mathrm{Hom}(A_K, E'_{1,K})_{\mathbb{Q}}$ decomposes as a sum of characters of order at most 2. Let χ denote any of these characters. Then $\mathrm{Hom}(A_K, (E'_1 \otimes \chi)_K)_{\mathbb{Q}}$ contains the trivial representation and thus $E'_1 \otimes \chi$ is a k-factor of A. This induces a decomposition $A \sim_k (E'_1 \otimes \chi) \times \tilde{E}_2$, for some elliptic curve $\tilde{E}_2/k$. We may then take $\tilde{E}_1 = E'_1 \otimes \chi$. □

In case **C**, let $A \sim_K E_1 \times E_2$, where E_1 and E_2 are elliptic curves defined over k with CM by M and without CM, respectively. The following two cases arise.

(i) $M \subseteq k$. Then $\mathrm{Gal}(K/k) = \mathrm{C}_1$ and ρ_A is the trivial representation. This is Galois type $\mathbf{C}[\mathrm{C}_1]$, and it cannot occur when $k = \mathbb{Q}$.

(ii) M is not contained in k. Then $\mathrm{Gal}(K/k) = \mathrm{C}_2$ and $\mathrm{Trace}\,\rho_A = 2\chi_1 + \chi_2$. This is Galois type $\mathbf{C}[\mathrm{C}_2]$, which can occur when $k = \mathbb{Q}$.

From the Lie group side, it is easy to check that if $\mathrm{ST}^0_A = \mathrm{U}(1) \times \mathrm{SU}(2)$, then $\mathrm{End}(A_K)_{\mathbb{R}} = \mathbb{R} \times \mathbb{C}$ and the normalizer acts nontrivially on $\mathbb{C}$, fixing $\mathbb{R} \times \mathbb{R}$. This shows that the Galois types corresponding to the Lie groups $\mathrm{U}(1) \times \mathrm{SU}(2)$ and $N(\mathrm{U}(1) \times \mathrm{SU}(2))$ are, respectively, $\mathbf{C}[\mathrm{C}_1]$ and $\mathbf{C}[\mathrm{C}_2]$.

In case **D**, let $A \sim_K E_1 \times E_2$, where E_1 and E_2 are nonisogenous elliptic curves defined over k with CM by two different imaginary quadratic fields M_1 and M_2, respectively. Then four cases arise.

(i) $M_1, M_2 \subseteq k$. Then $\mathrm{Gal}(K/k) = \mathrm{C}_1$ and ρ_A is the trivial representation; this is Galois type $\mathbf{D}[\mathrm{C}_1]$, which we already encountered in § 4.3.

(ii) $M_1 \subseteq k$ and M_2 is not contained in k. Then $\mathrm{Gal}(K/k) = \mathrm{C}_2$ and $(\mathrm{End}(A_K)_{\mathbb{Q}})^{\mathrm{C}_2} = M_1 \times \mathbb{Q}$, and thus $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_2} \simeq \mathbb{R} \times \mathbb{C}$. This is Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{C}]$.

(iii) M_1 and M_2 are not contained in k and $kM_1 = kM_2$. Then $\mathrm{Gal}(K/k) = \mathrm{C}_2$ and $(\mathrm{End}(A_K)_{\mathbb{Q}})^{\mathrm{C}_2} = \mathbb{Q} \times \mathbb{Q}$; we thus have $(\mathrm{End}(A_K)_{\mathbb{R}})^{\mathrm{C}_2} \simeq \mathbb{R} \times \mathbb{R}$, yielding the Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$ that we already met in § 4.3.

(iv) M_1 and M_2 are not contained in k and $kM_1 \neq kM_2$. Then $\mathrm{Gal}(K/k) \simeq \mathrm{D}_2$ and the three subalgebras of $\mathrm{End}(A_K)_{\mathbb{Q}}$ fixed by each of the subgroups of order 2 are $M_1 \times \mathbb{Q}$, $M_2 \times \mathbb{Q}$, and $\mathbb{Q} \times \mathbb{Q}$. This is Galois type $\mathbf{D}[\mathrm{D}_2]$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Among the four Galois types listed above, only the last can occur when $k = \mathbb{Q}$. The analysis in § 4.3 implies that the Galois types corresponding to the Lie groups F_a and $F_{a,b}$ are, respectively, $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{C}]$ and $\mathbf{D}[\mathrm{D}_2]$.

As a byproduct, since we have now classified all the possible Galois types of an abelian surface A for which $\mathrm{ST}_A^0 = \mathrm{U}(1) \times \mathrm{U}(1)$, we deduce that the Lie groups F_c, $F_{ab,c}$, and $F_{a,b,c}$ cannot occur as the Sato–Tate group of an abelian surface.

4.5 Products of isogenous elliptic curves

In case **E** or **F**, the endomorphism ring $\mathrm{End}(A_K)_{\mathbb{Q}}$ is a quaternion algebra B over $C = \mathbb{Q}$ or M, respectively. Write $B \to B$, $\alpha \mapsto \alpha'$ for the canonical anti-involution on B, and $n(\alpha) = \alpha\alpha' \in C$.

Proposition 4.6. *Assume that A/k is of type* **F**. *Then K contains M and $\mathrm{Gal}(K/kM)$ acts trivially on the center of $\mathrm{End}(A_K)_{\mathbb{Q}} \simeq \mathrm{M}_2(M)$.*

Proof. Let E/K be an elliptic curve such that $\mathrm{End}(E_K)_{\mathbb{Q}} \simeq M$ and $A \sim_K E^2$. Since any field of definition of the endomorphisms of E contains M (see, e.g., [Sil94, ch. II, Theorem 2.2]), we have $M \subseteq K$. Fix embeddings $\iota : M \subseteq K \subset \mathbb{C}$.

As shown in [Sil94, ch. II, Proposition 1.1], there exists a unique isomorphism $[\ ]_E : M \longrightarrow \mathrm{End}(E_K)_{\mathbb{Q}}$ such that for any invariant differential $\omega \in \Omega^1_E$, we have $[\alpha]^*_E(\omega) = \alpha\omega$ for all $\alpha \in M \subset \mathbb{C}$. There is therefore an unique isomorphism $[\ ]_{E^2} : \mathrm{M}_2(M) \longrightarrow \mathrm{End}(E^2_K)_{\mathbb{Q}}$ such that for any invariant differential $\omega = (\omega_1, \omega_2) \in \Omega^1_{E^2}$, we have $[\alpha]^*_{E^2}(\omega) = \alpha\omega$ for all $\alpha \in \mathrm{M}_2(M) \subset \mathrm{M}_2(\mathbb{C})$. It follows as in the proof of [Sil94, ch. II, Theorem 2.2] that ${}^\sigma[\alpha]_{E^2} = [{}^\sigma\alpha]_{({}^\sigma E)^2}$, hence

$$ {}^\sigma[\alpha]_{E^2} = [\alpha]_{({}^\sigma E)^2} \quad \text{for } \sigma \in \mathrm{Gal}(K/kM). \tag{4.1} $$

Endomorphisms of E^2 induce K-linear endomorphisms of the space of regular differentials $\Omega^1_{E^2}$; any choice of a K-basis of $\Omega^1_{E^2}$ gives rise to a monomorphism $t : \mathrm{M}_2(M) \simeq \mathrm{End}(E^2_K)_{\mathbb{Q}} \hookrightarrow \mathrm{M}_2(K) \hookrightarrow \mathrm{M}_2(\mathbb{C})$, $\alpha \mapsto [\alpha]^*_{E^2}$ whose restriction to the centers is ι.

That E^2 admits the model A over k up to isogenies over K implies, thanks again to Ribet's theorem [Rib04, Theorem 8.2], that there exists a collection of isogenies $\{\Phi_\sigma : ({}^\sigma E)^2 \longrightarrow E^2\}_{\sigma \in \mathrm{Gal}(K/k)}$ such that $\phi_\sigma {}^\sigma\phi_\tau = \phi_{\sigma\tau}$. The $\mathrm{Gal}(K/k)$-module $\mathrm{End}(A_K)_{\mathbb{Q}}$ is then isomorphic to the module $\mathrm{End}(E^2_K)_{\mathbb{Q}}$ equipped with the following action of the group $\mathrm{Gal}(K/k)$: an element $\sigma \in \mathrm{Gal}(K/k)$ acts on an endomorphism $[\alpha] \in \mathrm{End}(E^2_K)_{\mathbb{Q}}$ by the rule $\sigma \cdot [\alpha] = \phi_\sigma [{}^\sigma\alpha] \phi_\sigma^{-1}$.

Similarly, isogenies ϕ_σ induce K-linear isomorphisms $\pi^*_\sigma : \Omega^1_{({}^\sigma E)^2} \simeq \Omega^1_{E^2}$. If $[\alpha]$ lies in the center of $\mathrm{End}(E^2_K)_{\mathbb{Q}}$, then $\sigma \cdot [\alpha]^* = [{}^\sigma\alpha]^*$; if in addition $\sigma \in \mathrm{Gal}(K/kM)$, it follows from (4.1) that $\sigma \cdot [\alpha]^* = [\alpha]^*$. Since t is a monomorphism, we deduce that $\sigma \cdot [\alpha] = [\alpha]$; thus $\mathrm{Gal}(K/kM)$ acts trivially on the center of $\mathrm{End}(A_K)_{\mathbb{Q}} \simeq \mathrm{M}_2(M)$, as claimed. □

By the Skolem–Noether theorem, all automorphisms of $\mathrm{End}(A_K)_{\mathbb{Q}}$ that are the identity on C are inner. Set $\mathbb{P}(\mathrm{End}(A_K)^\times_{\mathbb{Q}}) = \mathrm{End}(A_K)^\times_{\mathbb{Q}} / C^\times$. A homothety class $[\alpha] \in \mathbb{P}(\mathrm{End}(A_K)^\times_{\mathbb{Q}})$ induces the automorphism c_α of $\mathrm{End}(A_K)_{\mathbb{Q}}$ given by the rule $\gamma \mapsto \alpha\gamma\alpha^{-1}$, which is the identity on C. This induces an isomorphism $\mathrm{Aut}_C(\mathrm{End}(A_K)_{\mathbb{Q}}) \simeq \mathbb{P}(\mathrm{End}(A_K)^\times_{\mathbb{Q}})$.

By the previous proposition, $\mathrm{Gal}(K/kC)$ is isomorphic to a subgroup of $\mathrm{Aut}_C(\mathrm{End}(A_K)_{\mathbb{Q}})$. The list of finite subgroups of $\mathbb{P}(\mathrm{End}(A_K)^\times_{\mathbb{Q}})$ is well-known (see [Bea10, CF00], for example), and this allows us to conclude that $\mathrm{Gal}(K/kC)$ is isomorphic to one of C_n, D_n, A_4, or S_4, where $n \in \{1, 2, 3, 4, 6\}$. The groups A_4 or S_4 arise only when -1 can be written as a sum of two squares in C, and thus only occur in case **F**.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

We now make a first analysis of the action of $\mathrm{Gal}(K/kC)$ on $\mathrm{End}(A_K)_{\mathbb{Q}}$, and therefore also on $\mathrm{End}(A_K)_{\mathbb{R}}$. This will address all the Galois types in case **E** and is a first step towards classifying the Galois types in case **F**, which we will conclude in § 4.5.2.

Proposition 4.7. *Let A/k be an abelian surface that is isogenous over K to the square of an elliptic curve.*

(i) *Assume $\mathrm{Gal}(K/kC) \simeq \mathrm{C}_n$ for $n = 1$, 2, 3, 4, or 6. If $n = 1$, then $\mathrm{Gal}(K/kC)$ acts trivially on $\mathrm{End}(A_K)_{\mathbb{Q}}$; if $C = \mathbb{Q}$, we denote the resulting Galois type by $\mathbf{E}[\mathrm{C}_1]$. If $n = 2$, then $\mathrm{End}(A_K)_{\mathbb{Q}}^{\mathrm{Gal}(K/kC)}$ is a quadratic extension of C. If $C = \mathbb{Q}$, this gives rise to two Galois types, according to whether the extension is real or imaginary; we label them $\mathbf{E}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$ and $\mathbf{E}[\mathrm{C}_2, \mathbb{C}]$, respectively. If $n > 2$, then $\mathrm{End}(A_K)_{\mathbb{Q}}^{\mathrm{Gal}(K/kC)} = C$ and for any nontrivial subgroup $H \subseteq \mathrm{Gal}(K/kC)$ we have*

$$\mathrm{End}(A_K)_{\mathbb{Q}}^{H} = C + C \cdot (1 + \zeta_n), \quad \zeta_n^n = 1, \zeta_n^{n_0} \neq 1 \text{ for } n_0 < n.$$

If $C = \mathbb{Q}$, for any $k \subseteq k' \subsetneq K$ we have $\mathrm{End}(A_{k'})_{\mathbb{Q}} = \mathbb{Q}(\zeta_n)$, which is an imaginary quadratic extension of $\mathbb{Q}$. This gives rise to a single Galois type, which we denote by $\mathbf{E}[\mathrm{C}_n]$.

(ii) *Assume that $\mathrm{Gal}(K/kC) \simeq \mathrm{D}_2 = \mathrm{C}_2 \times \mathrm{C}_2$ and $\mathrm{End}(A_K)_{\mathbb{Q}}^{\mathrm{Gal}(K/kC)} = C$, and that the three algebras fixed by each of the subgroups of $\mathrm{Gal}(K/kC)$ of order 2 are quadratic extensions of C. If $C = \mathbb{Q}$, then two of these quadratic extensions are real and one is imaginary; hence a single Galois type arises, which we denote by $\mathbf{E}[\mathrm{D}_2]$.*

(iii) *Assume that $\mathrm{Gal}(K/kC) \simeq \mathrm{D}_n$ for $n = 3$, 4, or 6. Write $\mathrm{Gal}(K/kC) = \langle r, s \rangle$ with $r^n = 1$, $s^2 = 1$, and $srs = r^{-1}$. Then $\mathrm{End}(A_K)_{\mathbb{Q}}^{\langle s \rangle} = C + C \cdot \sqrt{m}$ for some $m \in C$, and for any nontrivial subgroup $H \subseteq \langle r \rangle \subset \mathrm{Gal}(K/kC)$ we have $\mathrm{End}(A_K)_{\mathbb{Q}}^{H} = C + C \cdot (1 + \zeta_n)$. The algebra fixed by any other nontrivial subgroup of $\mathrm{Gal}(K/kC)$ is C. If $C = \mathbb{Q}$, then $\mathbb{Q}(\sqrt{m})$ is real quadratic and $\mathbb{Q}(\zeta_n)$ is imaginary quadratic; we thus obtain the single Galois type $\mathbf{E}[\mathrm{D}_n]$.*

We have relegated the study of the case where $C = M$ and $\mathrm{Gal}(K/kM) \simeq \mathrm{A}_4$ or S_4 to § 4.5.2: see Proposition 4.9 and the discussion following it.

Proof. Write $G = \mathrm{Gal}(K/kC)$. Let us first consider case (i), in which ρ_A induces an isomorphism between G and a cyclic subgroup of $B^\times/C^\times$ of order n. More precisely, $G = \langle c_\alpha \rangle$ for some $\alpha \in B^\times \backslash C^\times$ such that $\alpha^2 = d \in C^\times$ if $n = 2$ or $\alpha = 1 + \zeta_n$ if $n > 2$, where ζ_n is an element in $B^\times \backslash C^\times$ of order n. One checks that $\alpha^{n_0} \notin C$ for $n_0 < n$; thus the subalgebra of B fixed by any of the nontrivial subgroups of G is precisely $C(\alpha)$, which is quadratic over C.

For (ii), we now assume that $G \simeq \mathrm{D}_2$. By [CF00, Lemma 2.3], any subgroup of $B^\times/C^\times$ isomorphic to D_2 is of the form $\langle [\alpha], [\beta] \rangle \subset B^\times/C^\times$ with $\alpha, \beta \in B^\times \backslash C^\times$ satisfying $\alpha^2 = d$, $\beta^2 = m \in C^\times$, and $\alpha\beta = -\beta\alpha$. In particular, $(\alpha\beta)^2 = -dm$. It follows that the subalgebra of B fixed by G (respectively, by each of the three subgroups of G of order 2) is the center C (respectively, the quadratic extension $C(\alpha)$, $C(\beta)$, or $C(\alpha\beta)$ of C). If $C = \mathbb{Q}$, then we know that B is indefinite, that is to say,

$$B \otimes_{\mathbb{Q}} \mathbb{R} = \left(\frac{d, m}{\mathbb{R}} \right) \simeq \mathrm{M}_2(\mathbb{R}),$$

which amounts to saying that at least one of d and m is positive. This implies that exactly two of d, m, and $-dm$ are positive and one is negative.

Similarly, in case (iii), by [CF00, Lemma 2.3] all subgroups of $B^\times/C^\times$ isomorphic to D_n are of the form $\langle [\alpha], [\beta] \rangle$, where $\alpha = 1 + \zeta_n$ and $\beta \in B^\times \backslash C^\times$ satisfy $\beta^2 = m \in C^\times$ and $\zeta_n \beta = \beta \bar{\zeta}_n$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

The subalgebra of B fixed by any of the nontrivial subgroups of $\langle[1+\zeta_n]\rangle$ is $C(\zeta_n)$; the subalgebra fixed by $\langle[\beta]\rangle$ is $C(\beta)$. If $C=\mathbb{Q}$, then $\mathbb{Q}(\zeta_n)$ is either $\mathbb{Q}(\sqrt{-1})$ or $\mathbb{Q}(\sqrt{-3})$, which are both imaginary, and we necessarily have $m>0$ because B is indefinite. All the claims in (iii) follow. □

Remark 4.8. From the proof above, one also deduces that in case **E** the $\mathrm{Gal}(K/k)$-module structure of $\mathrm{End}(A_K)_{\mathbb{R}}$ is given by the rule

$$\mathrm{Trace}\,\rho_A(\sigma)=2+\zeta_r+\overline{\zeta}_r,$$

where r is the order of $\sigma\in\mathrm{Gal}(K/k)$ and ζ_r denotes a primitive rth root of unity. Thus, in case **E**, the $\mathrm{Gal}(K/k)$-module structure of $\mathrm{End}(A_K)_{\mathbb{R}}$ is completely determined by $\mathrm{Gal}(K/k)$ (compare this result with Proposition 4.9).

4.5.1 *Galois types and Sato–Tate groups in case* **E**. For case **E** we have found a total of ten Galois types: $\mathbf{E}[\mathrm{C}_n]$ for $n=1,3,4,6$; $\mathbf{E}[\mathrm{C}_2,\mathbb{R}\times\mathbb{R}]$ and $\mathbf{E}[\mathrm{C}_2,\mathbb{C}]$; and $\mathbf{E}[\mathrm{D}_n]$ for $n=2,3,4,6$.

Let us recover this classification from the Lie group side, matching ten of the groups named in Theorem 3.4 with these Galois types. Assume now that A/k is an abelian surface such that $\mathrm{ST}_A^0=\mathrm{SU}(2)$. In this case, the matrices in $\mathrm{M}_4(\mathbb{C})$ commuting with $\mathrm{SU}(2)$ are

$$\left\{\begin{pmatrix} a\mathrm{Id}_2 & bJ_2\\ cJ_2 & d\mathrm{Id}_2\end{pmatrix} : a,b,c,d\in\mathbb{C}\right\}.$$

The Rosati form is given up to a scalar multiple by

$$\begin{aligned}\psi &\mapsto \mathrm{Trace}(\Psi S^T\Psi^T S)\\ &=\mathrm{Trace}\left(\begin{pmatrix} a\mathrm{Id}_2 & bJ_2\\ cJ_2 & d\mathrm{Id}_2\end{pmatrix}\begin{pmatrix} 0 & -\mathrm{Id}_2\\ \mathrm{Id}_2 & 0\end{pmatrix}\begin{pmatrix} a\mathrm{Id}_2 & -cJ_2\\ -bJ_2 & d\mathrm{Id}_2\end{pmatrix}\begin{pmatrix} 0 & \mathrm{Id}_2\\ -\mathrm{Id}_2 & 0\end{pmatrix}\right)\\ &=\mathrm{Trace}\left(\begin{pmatrix} bJ_2 & -a\mathrm{Id}_2\\ d\mathrm{Id}_2 & -cJ_2\end{pmatrix}\begin{pmatrix} cJ_2 & a\mathrm{Id}_2\\ -d\mathrm{Id}_2 & -bJ_2\end{pmatrix}\right)\\ &=2(ad-bc)=\tfrac{1}{2}((a+d)^2-(a-d)^2-(b+c)^2+(b-c)^2).\end{aligned}$$

The positive definite subspace is defined by the conditions $a+d,b-c\in\mathbb{R}$ and $a-d,b+c\in i\mathbb{R}$, and therefore

$$\mathrm{End}(A_K)_{\mathbb{R}}=\left\{\begin{pmatrix}(a+bi)\mathrm{Id}_2 & (c+di)J_2\\ (-c+di)J_2 & (a-bi)\mathrm{Id}_2\end{pmatrix} : a,b,c,d\in\mathbb{R}\right\}\simeq \mathrm{M}_2(\mathbb{R}).$$

For $n>1$, the action of a generator of the component group of E_n is

$$\begin{pmatrix}(a+bi)\mathrm{Id}_2 & (c+di)J_2\\ (-c+di)J_2 & (a-bi)\mathrm{Id}_2\end{pmatrix}\mapsto\begin{pmatrix}(a+bi)\mathrm{Id}_2 & e^{2\pi i/n}(c+di)J_2\\ e^{-2\pi i/n}(-c+di)J_2 & (a-bi)\mathrm{Id}_2\end{pmatrix},$$

so the fixed ring consists of matrices with $c=d=0$ and is isomorphic to $\mathbb{C}$. The action of J is

$$\begin{pmatrix}(a+bi)\mathrm{Id}_2 & (c+di)J_2\\ (-c+di)J_2 & (a-bi)\mathrm{Id}_2\end{pmatrix}\mapsto\begin{pmatrix}(a-bi)\mathrm{Id}_2 & (c-di)J_2\\ (-c-di)J_2 & (a+bi)\mathrm{Id}_2\end{pmatrix},$$

so the fixed ring of $J(E_1)$ consists of matrices with $b=d=0$ and is isomorphic to $\mathbb{R}\times\mathbb{R}$. For $n>1$, the fixed ring of $J(E_n)$ is isomorphic to $\mathbb{R}$ because we must have both $c=d=0$ and $b=d=0$.

It follows from the previous discussion that the correspondence between Galois types in case **E** and Sato–Tate groups with connected component $\mathrm{SU}(2)$ is as indicated in the statement of Theorem 4.3.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

4.5.2 *Galois types in case* **F**. In this section we assume that A/k is of type **F**, so that[11] $\mathrm{ST}_A^0 \simeq \mathrm{U}(1)$.

Proposition 4.9. *The* $\mathrm{Gal}(K/k)$*-module structure of* $\mathrm{End}(A_K)_{\mathbb{R}}$ *is determined by the pair* $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$. *More precisely, it is given by the following rule: for* $\sigma \in \mathrm{Gal}(K/k)$,

$$\mathrm{Trace}\,\rho_A(\sigma) = \begin{cases} 2(2+\zeta_r+\overline{\zeta}_r) & \text{if } \sigma \in \mathrm{Gal}(K/kM), \\ 0 & \text{otherwise.} \end{cases}$$

Here r *is the order of* σ *and* ζ_r *stands for a primitive* r*th root of unity.*

Proof. Suppose first that $\sigma \in \mathrm{Gal}(K/kM)$. Recall that, except for a set of density zero, a prime $\mathfrak{p}$ of k is supersingular if and only if $\mathfrak{p}$ is inert in kM. Let $\mathfrak{p}$ be a split (i.e., not supersingular) prime in kM of good reduction for A. Let $\mathfrak{P}$ be a prime of kM over $\mathfrak{p}$. We first show that if $\mathrm{Frob}_{\mathfrak{P}}$ lies in the conjugacy class of σ in $\mathrm{Gal}(K/kM)$, then the roots of $L_{\mathfrak{P}}(A_{kM}, T)$ are α, $\overline{\alpha}$, $\zeta_r\alpha$, and $\overline{\zeta}_r\overline{\alpha}$, for a certain $\alpha \in \mathbb{C}$. Indeed, let α be a root of $L_{\mathfrak{P}}(A_{kM}, T)$. Since $\mathfrak{P}$ is not supersingular, $\alpha/\overline{\alpha}$ is a root of unity. Observe that the eigenvalues of $\rho_A(\sigma)$ are quotients of roots of $L_{\mathfrak{P}}(A_{kM}, T)$ (see, for example, [Fit10]). Suppose that σ is not the trivial element (as otherwise the proposition is trivially true). Then $\rho_A(\sigma)$ has an eigenvalue $\omega \neq 1$, and the roots of $L_{\mathfrak{p}}(A/kM, T)$ are α, $\overline{\alpha}$, $\omega\alpha$, and $\overline{\omega\alpha}$. It follows that the set of eigenvalues of $\rho_A(\sigma)$ is $\{1, \omega, \overline{\omega}\}$. Finally, one observes that ω must be a primitive rth root of unity since the order of σ is r.

But again the eigenvalues of $\rho_A(\sigma)$ are quotients of the roots α, $\overline{\alpha}$, $\zeta_r\alpha$, $\overline{\zeta}_r\overline{\alpha}$ of $L_{\mathfrak{P}}(A_{kM}, T)$ and, since $\mathfrak{P}$ is not supersingular, $\alpha/\overline{\alpha}$ is not a root of unity. Thus, among the 16 possible quotients between the roots of $L_{\mathfrak{P}}(A_{kM}, T)$, only the following 8 are roots of unity: 1, 1, 1, 1, ζ_r, ζ_r, $\overline{\zeta}_r$, $\overline{\zeta}_r$. Since ρ_A has dimension 8, we have $\mathrm{Trace}\,\rho_A(\sigma) = 2(2+\zeta_r+\overline{\zeta}_r)$.

Suppose now that $k \neq kM$ and $\sigma \notin \mathrm{Gal}(K/kM)$. Let χ denote the quadratic character of $\mathrm{Gal}(K/k)$ associated to the extension kM/k. Let $\mathfrak{p}$ be any prime of k which does not split completely in kM. Since $\mathfrak{p}$ is supersingular, we must have $\mathrm{Trace}\, V_\ell(A)(\mathrm{Frob}_{\mathfrak{p}}) = 0$. It follows that $V_\ell(A) \simeq V_\ell(A) \otimes \chi$, that is, $A \sim_k A \otimes \chi$. Therefore $\mathrm{End}(A_K)_{\mathbb{R}} = \mathrm{End}(A_K)_{\mathbb{R}} \otimes \chi$, which implies the claim. □

Remark 4.10. It is not true that the pair $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$ determines the Galois type of A: there exist examples of abelian surfaces A/k and A'/k' for which there is an isomorphism of abstract groups

$$(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM)) \overset{\varphi}{\simeq} (\mathrm{Gal}(K'/k'), \mathrm{Gal}(K'/k'M'))$$

but such that for a certain subgroup $H \subseteq \mathrm{Gal}(K/k)$ of order 2, the rings $\mathrm{End}(A_K)_{\mathbb{R}}^{H}$ and $\mathrm{End}(A'_K)_{\mathbb{R}}^{\varphi(H)}$ are not isomorphic as $\mathbb{R}$-algebras. The list of Galois types that are ambiguous in this way can be found in Table 3.

The computations performed in § 3.4 allow us to determine the possible isomorphism classes for the pair of groups $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$. Indeed, observe that ST_A determines not only $\mathrm{Gal}(K/k) \simeq \mathrm{ST}_A/\mathrm{ST}_A^0$ but also the subgroup $\mathrm{Gal}(K/kM)$: since a prime $\mathfrak{p}$ of k is supersingular if and only if it does not split completely in kM, the group $\mathrm{Gal}(K/kM)$ is isomorphic to the component group of ST_A^{ns}, where ST_A^{ns} denotes the index 2 subgroup of ST_A obtained

[11] This was indicated without proof in § 4.1, but it is now a formal consequence of the above: Theorem 2.16, Proposition 3.2, and Lemma 3.7 imply that ST_A^0 is one of the six connected Lie groups listed in Lemma 3.7. If ST_A^0 were not conjugate to U(1), the computations of §§ 4.2–4.5.1 would imply that the type of A is one of **A**, **B**, **C**, **D**, or **E**, rather than **F**.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 2. Pairs $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$.

ST_A		ST_A		ST_A	
C_1	$(\mathrm{C}_1, \mathrm{C}_1)$	$J(C_1)$	$(\mathrm{C}_2, \mathrm{C}_1)$	$C_{2,1}$	$(\mathrm{C}_2, \mathrm{C}_1)$
C_2	$(\mathrm{C}_2, \mathrm{C}_2)$	$J(C_2)$	$(\mathrm{D}_2, \mathrm{C}_2)$	$C_{4,1}$	$(\mathrm{C}_4, \mathrm{C}_2)$
C_3	$(\mathrm{C}_3, \mathrm{C}_3)$	$J(C_3)$	$(\mathrm{C}_6, \mathrm{C}_3)$	$C_{6,1}$	$(\mathrm{C}_6, \mathrm{C}_3)$
C_4	$(\mathrm{C}_4, \mathrm{C}_4)$	$J(C_4)$	$(\mathrm{C}_4 \times \mathrm{C}_2, \mathrm{C}_4)$	$D_{2,1}$	$(\mathrm{D}_2, \mathrm{C}_2)$
C_6	$(\mathrm{C}_6, \mathrm{C}_6)$	$J(C_6)$	$(\mathrm{C}_6 \times \mathrm{C}_2, \mathrm{C}_6)$	$D_{4,1}$	$(\mathrm{D}_4, \mathrm{D}_2)$
D_2	$(\mathrm{D}_2, \mathrm{D}_2)$	$J(D_2)$	$(\mathrm{D}_2 \times \mathrm{C}_2, \mathrm{D}_2)$	$D_{6,1}$	$(\mathrm{D}_6, \mathrm{D}_3)$
D_3	$(\mathrm{D}_3, \mathrm{D}_3)$	$J(D_3)$	$(\mathrm{D}_6, \mathrm{D}_3)$	$D_{3,2}$	$(\mathrm{D}_3, \mathrm{C}_3)$
D_4	$(\mathrm{D}_4, \mathrm{D}_4)$	$J(D_4)$	$(\mathrm{D}_4 \times \mathrm{C}_2, \mathrm{D}_4)$	$D_{4,2}$	$(\mathrm{D}_4, \mathrm{C}_4)$
D_6	$(\mathrm{D}_6, \mathrm{D}_6)$	$J(D_6)$	$(\mathrm{D}_6 \times \mathrm{C}_2, \mathrm{D}_6)$	$D_{6,2}$	$(\mathrm{D}_6, \mathrm{C}_6)$
T	$(\mathrm{A}_4, \mathrm{A}_4)$	O_1	$(\mathrm{S}_4, \mathrm{A}_4)$	$J(T)$	$(\mathrm{A}_4 \times \mathrm{C}_2, \mathrm{A}_4)$
O	$(\mathrm{S}_4, \mathrm{S}_4)$	$J(O)$	$(\mathrm{S}_4 \times \mathrm{C}_2, \mathrm{S}_4)$		

by removing from ST_A those components for which all elements have the same characteristic polynomial. For each Lie group G with $G^0 = \mathrm{U}(1)$ appearing in the list of Theorem 3.4, the pair $(G/G^0, G^{ns}/G^{ns,0})$ is shown in Table 2.

By Propositions 4.6 and 4.9, there are eleven Galois types for the case $\mathbf{F}$ in which $M \subseteq k$, and thus $\mathrm{Gal}(K/k) = \mathrm{Gal}(K/kM)$. Indeed, note first that for any subgroup $H \subseteq \mathrm{Gal}(K/kM)$, the isomorphism class of the $\mathbb{R}$-algebra $\mathrm{End}(A_K)_{\mathbb{R}}^H$ depends only on its dimension: by Proposition 4.6 we know that $\mathrm{End}(A_K)_{\mathbb{Q}}^H$ is either M, a quadratic extension of M, or $\mathrm{M}_2(M)$, and so upon tensoring with $\mathbb{R}$ becomes $\mathbb{C}$, $\mathbb{C} \times \mathbb{C}$, or $\mathrm{M}_2(\mathbb{C})$, respectively. In addition, Proposition 4.9 shows that the $\mathrm{Gal}(K/kM)$-module structure of $\mathrm{End}(A_K)_{\mathbb{Q}}$ is uniquely determined by the isomorphism class of the group $\mathrm{Gal}(K/kM)$. This yields the Galois types $\mathbf{F}[\mathrm{C}_n]$ for $n = 1, 2, 3, 4, 6$; $\mathbf{F}[\mathrm{D}_n]$ for $n = 2, 3, 4, 6$; $\mathbf{F}[\mathrm{A}_4]$; and $\mathbf{F}[\mathrm{S}_4]$. From Table 2, it follows that these correspond to the Sato–Tate groups C_n, D_n, T, and O, respectively.

If $M \not\subseteq k$, then, by Proposition 4.9, the pair $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$ still determines the $\mathrm{Gal}(K/k)$-module structure of $\mathrm{End}(A_K)_{\mathbb{R}}$, but, as we warned in Remark 4.10, more data is needed to determine the Galois type. We now describe these data. A glance at Table 2 shows that each pair $(\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM))$ gives rise to exactly one Galois type, which we denote by $\mathbf{F}[\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM)]$, *except* for the four pairs

$$(\mathrm{C}_2, \mathrm{C}_1),\ (\mathrm{D}_2, \mathrm{C}_2),\ (\mathrm{C}_6, \mathrm{C}_3),\ (\mathrm{D}_6, \mathrm{D}_3). \tag{4.2}$$

In each of these cases we have $\mathrm{Gal}(K/k) \simeq \mathrm{Gal}(K/kM) \times \mathrm{C}_2$. Choose such an isomorphism and write σ for the nontrivial involution of $\mathrm{Gal}(K/k)$ generating that cyclic subgroup of order 2. By the Skolem–Noether theorem, σ acts on $\mathrm{End}(A_K)_{\mathbb{R}} \simeq \mathrm{M}_2(\mathbb{C})$ by

$$x \in \mathrm{M}_2(\mathbb{C}) \mapsto \sigma(x) = \gamma \overline{x} \gamma^{-1}, \tag{4.3}$$

for some $\gamma \in \mathrm{GL}_2(\mathbb{C})$ satisfying $\gamma\overline{\gamma} \in \mathbb{C}^\times \mathrm{Id}$. It is obvious from this description that $\mathrm{End}(A_K)_{\mathbb{R}}^{\langle\sigma\rangle} \cap \mathbb{C} \cdot \mathrm{Id} = \mathbb{R} \cdot \mathrm{Id}$. A further inspection (by writing down the linear equation resulting from (4.3)) shows that $\mathrm{End}(A_K)_{\mathbb{R}}^{\langle\sigma\rangle}$ is an $\mathbb{R}$-algebra of rank 4. It cannot be commutative, because if it were, it would be a maximal commutative subalgebra of $\mathrm{M}_2(\mathbb{C})$ and should thus contain the center. Hence $\mathrm{End}(A_K)_{\mathbb{R}}^{\langle\sigma\rangle}$ is a quaternion algebra over $\mathbb{R}$, which must be isomorphic to either $\mathrm{M}_2(\mathbb{R})$ or Hamilton's division quaternion algebra $\mathbb{H} := (\frac{-1,-1}{\mathbb{R}})$.

For each pair in (4.2), we label these two Galois types as

$$\mathbf{F}[\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM), \mathrm{M}_2(\mathbb{R})] \quad \text{and} \quad \mathbf{F}[\mathrm{Gal}(K/k), \mathrm{Gal}(K/kM), \mathbb{H}].$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 3. Galois types for the exceptional pairs.

ST group	Galois type	ST group	Galois type
$J(C_1)$	$\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$	$C_{2,1}$	$\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathrm{M}_2(\mathbb{R})]$
$J(C_2)$	$\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathbb{H}]$	$D_{2,1}$	$\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathrm{M}_2(\mathbb{R})]$
$J(C_3)$	$\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$	$C_{6,1}$	$\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathrm{M}_2(\mathbb{R})]$
$J(D_3)$	$\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathbb{H}]$	$D_{6,1}$	$\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathrm{M}_2(\mathbb{R})]$

We next apply Proposition 2.19 to determine $\mathrm{End}(A)_{\mathbb{R}}$ for each Sato–Tate group G with $G^0 = \mathrm{U}(1)$. This computation shows, in particular, that for the four preceding pairs, the one-to-one correspondence with Sato–Tate groups is as indicated in Table 3.

Any $\psi \in \mathrm{End}(A_K)_{\mathbb{R}}$ acts via a block diagonal matrix

$$\Psi = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad A, B \in \mathrm{M}_2(\mathbb{C}).$$

Since ψ' acts via the matrix $S^{-1}\Psi^T S$, the Rosati form is given up to scalars by

$$\psi \mapsto \mathrm{Trace}(\Psi S^T \Psi^T S) = 2\,\mathrm{Trace}(AB^T).$$

By positivity of the Rosati form, we must have $B = \overline{A}$, whence

$$\mathrm{End}(A_K)_{\mathbb{R}} = \left\{ \begin{pmatrix} A & 0 \\ 0 & \overline{A} \end{pmatrix} : A \in \mathrm{M}_2(\mathbb{C}) \right\} \simeq \mathrm{M}_2(\mathbb{C}).$$

For $n > 1$, the action of a generator of the component group of C_n is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & e^{2\pi i/n} b \\ e^{-2\pi i/n} c & d \end{pmatrix}.$$

The fixed ring consists of those matrices with $b = c = 0$, and is thus isomorphic to $\mathbb{C} \times \mathbb{C}$. If we reinterpret C_2 as D_1, the action of the generator becomes

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

For $n > 1$, the fixed ring under D_n consists of matrices with $a = d$ and $b = c = 0$, and is isomorphic to $\mathbb{C}$. The same is true for T and O.

The action of J is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} \overline{d} & -\overline{c} \\ -\overline{b} & \overline{a} \end{pmatrix},$$

so the fixed ring under $J(C_1)$ is isomorphic to the Hamilton quaternion ring $\mathbb{H}$. For $n > 1$, the fixed ring under $J(C_n)$ consists of matrices with $b = c = 0$ and $d = \overline{a}$, and is isomorphic to $\mathbb{C}$. The fixed ring under $J(D_n)$ has the additional condition $d = a$, and is isomorphic to $\mathbb{R}$; the same is true for $J(T)$ and $J(O)$.

The action of a generator of $C_{2,1}$ is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} \overline{d} & \overline{c} \\ \overline{b} & \overline{a} \end{pmatrix},$$

so the fixed ring is $\mathrm{M}_2(\mathbb{R})$. The fixed ring under $D_{2,1}$ consists of matrices with $a = d = \overline{d}$ and $b = -c = \overline{c}$, and is isomorphic to $\mathbb{R} \times \mathbb{R}$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

The action of a generator of $C_{4,1}$ is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} \overline{d} & i\overline{c} \\ -i\overline{b} & \overline{a} \end{pmatrix},$$

so the fixed ring consists of matrices with $b = c = 0$ and $d = \overline{a}$, which is isomorphic to $\mathbb{C}$, and similarly for $C_{6,1}$. The fixed ring under $D_{4,1}$ adds the condition $d = a$, hence it is isomorphic to $\mathbb{R}$, and similarly for $D_{6,1}$ and O_1 (which contains $D_{4,1}$).

Since $D_{3,2}$ contains C_3, its fixed ring only contains matrices with $b = c = 0$. Since $D_{3,2}$ also contains $J(D_1)$, we must also have $a = \overline{a}$ and $d = \overline{d}$, so the fixed ring is isomorphic to $\mathbb{R} \times \mathbb{R}$. The same holds for $D_{4,2}$ and $D_{6,2}$.

4.6 Correspondence with Sato–Tate groups

Having completed the description of the 52 Galois types, let us now address the correspondence with Sato–Tate groups included in Theorem 4.3, that is, the equivalence between the three sets of data (a), (b), and (c) named in the theorem. Note that Proposition 2.19 implies that (a) determines (b), and this has been made explicit in the preceding sections. Since it is clear that (b) determines (c), it remains only to show that (c) determines (a).

We first note that the six choices for ST_A^0 give rise to six distinct isomorphism classes **A**, **B**, **C**, **D**, **E**, and **F** for the $\mathbb{R}$-algebra $\mathrm{End}(A_K)_{\mathbb{R}}$, so ST_A^0 and $\mathrm{End}(A_K)_{\mathbb{R}}$ determine each other. Thus, to prove that (c) determines (a), it is sufficient to distinguish Sato–Tate groups with the same connected part.

To finish the argument, let us inspect Table 8 more closely. We find that the data given by $(\mathrm{Gal}(K/k), \mathrm{End}(A_K)_{\mathbb{R}}, \mathrm{End}(A_k)_{\mathbb{R}})$ alone is sufficient to distinguish Sato–Tate groups but for three exceptional pairs of groups where an ambiguity arises. The first ambiguous pair is $J(C_2)$ and D_2; these two may be distinguished by considering $\mathrm{End}(A_L)_{\mathbb{R}}$, where L/k runs over the three quadratic subextensions of K/k. For $J(C_2)$ one obtains $\mathbb{H}$, $\mathbb{H}$, and $\mathbb{C} \times \mathbb{C}$, since the index 2 subgroups of $J(C_2)$ are conjugate to $J(C_1)$, $J(C_1)$, and C_2, whereas for D_2 one obtains $\mathbb{C} \times \mathbb{C}$ in all cases, since all index 2 subgroups of D_2 are conjugate to C_2. The second ambiguous pair is $J(C_3)$ and $C_{6,1}$; these two may be distinguished by passing from the cyclic group $\mathrm{Gal}(K/k)$ of order 6 to its unique subgroup of order 2, thus reducing to the distinction between $J(C_1)$ and $C_{2,1}$. The third ambiguous pair is $J(D_3)$ and $D_{6,1}$; these two may be distinguished by passing from the dihedral group $\mathrm{Gal}(K/k)$ of order 12 to its unique cyclic subgroup of order 6, thus reducing to the distinction between $J(C_3)$ and $C_{6,1}$.

4.7 Realizability over $\mathbb{Q}$

We now show that certain Galois types cannot occur over $\mathbb{Q}$ or, more generally, over a field with a real place.

As determined at the end of § 4.4, the group F_c does not occur as a Sato–Tate group of an abelian surface over any number field. The remaining 14 Galois types corresponding to Sato–Tate groups ruled out by Remark 3.5 over a field with a real place are

$$\mathbf{F}[\mathrm{C}_n] \text{ with } n \in \{1, 2, 3, 4, 6\},$$
$$\mathbf{F}[\mathrm{D}_n] \text{ with } n \in \{2, 3, 4, 6\},$$
$$\mathbf{F}[\mathrm{A}_4],\ \mathbf{F}[\mathrm{S}_4],\ \mathbf{D}[\mathrm{C}_1],\ \mathbf{D}[\mathrm{C}_2, \mathbb{C}],\ \mathbf{C}[\mathrm{C}_1].$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

This list can be recovered immediately from the discussion in §§ 4.2–4.5: for all of these Galois types, k must contain either a quadratic imaginary field or a quartic CM-field.

The Galois type $\mathbf{D}[\mathrm{C}_2, \mathbb{R} \times \mathbb{R}]$, corresponding to F_{ab}, cannot occur over $\mathbb{Q}$ since it corresponds to an abelian surface A over k such that $A \sim_k E_1 \times E_2$, where E_i is an elliptic curve over k with CM by a quadratic imaginary field $M_i = \mathbb{Q}(\sqrt{-d_i})$, $i = 1, 2$, such that $M_1 \not\simeq M_2$ and $kM_1 = kM_2$. Of course, this last condition does not hold if $k = \mathbb{Q}$, but it can hold over a totally real field (e.g., $k = \mathbb{Q}(\sqrt{d_1 d_2})$).

In order to complete the proof of Theorem 4.3, we still need to prove that three other Galois types do not occur over $\mathbb{Q}$. In fact, we show that they cannot arise over a field with a real place.

Proposition 4.11. *The Galois types* $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$, $\mathbf{F}[\mathrm{C}_4, \mathrm{C}_2]$, *and* $\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$ *cannot occur over a field with a real place.*

Proof. Let A be an abelian surface over k with Galois type $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$ (or, equivalently, with $\mathrm{ST}_A = J(C_1)$). Note that $K = kM/k$ is a quadratic extension and write $\mathrm{Gal}(K/k) = \{e, \tau\}$, where e and τ denote the identity and complex conjugation, respectively. Extend scalars from k to $\mathbb{R}$. Thus, $K \otimes \mathbb{R} = \mathbb{C}$ and $A_{\mathbb{C}}$ is isogenous to the square of some CM elliptic curve E over $\mathbb{C}$. Let $\mathcal{O}$ be the endomorphism ring of E, and take the curve E' corresponding to the lattice $\mathcal{O}$ in $\mathbb{C}$; this is defined over $\mathbb{R}$ because the lattice is stable under complex conjugation. Then E and E' are isogenous over an algebraic closure of $\mathbb{C}$, which is again $\mathbb{C}$.

Thus $A_{\mathbb{R}}$ is a twist of $(E')^2$ by some 1-cocycle f. If we normalize $f(e) = 1$, then f sends the complex conjugation τ to some endomorphism α of $(E')^2$ for which $\alpha\tau(\alpha) = 1$. If we translate the action of τ on $\mathrm{End}(A_{\mathbb{C}})$ into an action on $\mathrm{End}((E'_{\mathbb{C}})^2)_{\mathbb{R}} \simeq M_2(\mathbb{C})$, then it is described by conjugating the complex conjugation on $\mathrm{M}_2(\mathbb{C})$ by α. By the Hilbert–Speiser theorem 90 (for $H^1(\mathrm{Gal}(\mathbb{C}/\mathbb{R}), \mathrm{GL}_2(\mathbb{C}))$, we can factor α as $\beta\tau(\beta)^{-1}$ for some β in $\mathrm{GL}_2(\mathbb{C})$. Using this, we can then calculate that the fixed subring of $\mathrm{End}(A_{\mathbb{C}})_{\mathbb{R}}$ under τ is isomorphic to $\mathrm{M}_2(\mathbb{R})$, just as for $(E'_{\mathbb{C}})^2$. However, for $J(C_1)$ this $\mathbb{R}$-algebra must be isomorphic to $\mathbb{H}$.

Suppose now that the Galois type of A is $\mathbf{F}[\mathrm{C}_4, \mathrm{C}_2]$ (equivalently, that $\mathrm{ST}_A = C_{4,1}$). Note that kM/k is the unique quadratic extension of K/k. Again, extend scalars from k to $\mathbb{R}$, and note that $M \otimes \mathbb{R} = \mathbb{C}$. By the argument of the previous paragraph, $\mathrm{End}(A_{\mathbb{R}})_{\mathbb{R}}$ has rank 4 as an $\mathbb{R}$-algebra. However, in case $C_{4,1}$, no extension L of k contained in $\mathbb{R}$ can have endomorphism algebra of rank greater than 2 (namely, if L does not contain M, then $\mathrm{End}(A_L)_{\mathbb{R}} = \mathrm{End}(A_k)_{\mathbb{R}} = \mathbb{C}$).

We can reduce the case of an abelian surface A with Galois type $\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$ to the case of an abelian surface with Galois type $\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$ by considering A_L, where L/k is the only cubic subextension of K/k (note that L preserves the property of having a real place). □

4.8 Examples

In §§ 4.2–4.5 we have shown that there are at most 52 Galois types that can arise for an abelian surface over a number field. In § 4.7 we showed that 18 (respectively, 17) of these cannot occur over $\mathbb{Q}$ (respectively, over a field with a real place). To complete the proof of Theorem 4.3, it remains to show that each of the 52 Galois types is actually realized by an abelian surface A/k, and that for the 34 Galois types admissible over $\mathbb{Q}$, this can be achieved with $k = \mathbb{Q}$.

Here we accomplish this goal by exhibiting explicit examples that are Jacobians of genus 2 curves; these curves have been chosen so that the Sato–Tate groups can be explained entirely using automorphisms of the curves themselves, without having to study endomorphisms of the Jacobians. The 52 curves are listed in Table 11 together with a corresponding Sato–Tate group,

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

which, as shown in § 4.6, uniquely determines a Galois type. In this section we prove that each of these curves has the Sato–Tate group listed in Table 11, and thus realizes the corresponding Galois type. We note that the curves corresponding to Galois types admissible over $\mathbb{Q}$ are all defined over $\mathbb{Q}$, and that the curve given for $\mathbf{D}[\mathrm{C}_2, \mathbb{R}\times\mathbb{R}]$ is defined over a totally real field.

We begin with the Galois types in cases **E** and **F**. The first step is to compute the field K for each curve. Let α and γ be automorphisms of a genus 2 curve C, such that α is a nonhyperelliptic involution and α and γ do not commute. Let L/k denote the minimal field extension over which α and γ are defined. The quotient $E := C/\langle\alpha\rangle$ is an elliptic curve over L. Recall that M denotes $\mathbb{Q}$ if E does not have CM and denotes the CM-field of E otherwise. We claim that $K = LM$. Indeed, there exists an elliptic curve E' over L such that $\mathrm{Jac}(C) \sim_L E \times E'$. Since $\mathrm{Aut}(C_L)$ is nonabelian and injects into $\mathrm{End}(\mathrm{Jac}(C)_L)_{\mathbb{Q}}$, we must have $\mathrm{End}(\mathrm{Jac}(C)_L)_{\mathbb{Q}} \simeq \mathrm{M}_2(\mathrm{End}(E_L)_{\mathbb{Q}})$ and $E \sim_L E'$. It follows that $K = LM$.

In Table 12, we list α, γ, and M for each of the 42 curves in cases **E** and **F**. From this data one can immediately compute the fields K, which are listed in Table 11; note that when writing the coefficients of α and γ in Table 12, there is an implicit reference to the generators of K given in Table 11.

Except for $J(E_1)$ and E_2, the Sato–Tate groups in case **E** are uniquely determined by $\mathrm{Gal}(K/k)$, and in each such case this implies that the claimed Sato–Tate group is correct. One finds that the curve $C\colon y^2 = x^5 + x^3 + x$ (respectively, $C : x^6 + x^5 + 3x^4 + 3x^2 - x + 1$) listed for $J(E_1)$ (respectively, E_2) has $\mathrm{End}(\mathrm{Jac}(C)_{\mathbb{Q}})_{\mathbb{Q}} \simeq \mathbb{Q}\times\mathbb{Q}$ (respectively, $\mathbb{Q}(\sqrt{-2})$), from which it follows that $\mathrm{End}(\mathrm{Jac}(C)_{\mathbb{Q}})_{\mathbb{R}} \simeq \mathbb{R}\times\mathbb{R}$ (respectively, $\mathbb{C}$), as desired.

Except for the pairs $J(C_1)$ and $C_{2,1}$, $J(C_2)$ and $D_{2,1}$, $J(C_3)$ and $C_{6,1}$, and $J(D_3)$ and $D_{6,1}$, the Sato–Tate groups in case **F** are uniquely determined by $\mathrm{Gal}(K/k)$ and $\mathrm{Gal}(K/kM)$; these Galois groups can be directly computed from the data in Tables 11 and 12, and in each case one finds that the claimed Sato–Tate group is correct. We now address the eight ambiguous cases.

(i) $J(C_1)$: The curve $C\colon y^2 = x^5 - x$ over $k = \mathbb{Q}(i)$ has an automorphism $\beta(x, y) = (1/x, iy/x^3)$ with $\beta^2 = -1$. Since we also have $\gamma^2 = -1$ and $\gamma\circ\beta = -\beta\circ\gamma$, we conclude that $\mathrm{End}(\mathrm{Jac}(C)_k)_{\mathbb{R}} \simeq \mathbb{H}$.

(ii) $C_{2,1}$: We observe that the curve $C\colon y^2 = x^6 + 1$ over $k = \mathbb{Q}$ has $\alpha^2 = \gamma^2 = 1$ and $\alpha\circ\gamma = -\gamma\circ\alpha$. It follows that $\mathrm{End}(\mathrm{Jac}(C)_k)_{\mathbb{R}} \simeq \mathbb{M}_2(\mathbb{R})$.

(iii) $J(C_2)$: The curve $C\colon y^2 = x^5 - x$ over $k = \mathbb{Q}$ has an automorphism $\beta(x, y) = (-1/x, y/x^3)$ of order 4, from which we deduce that $\mathrm{End}(\mathrm{Jac}(C)_k)_{\mathbb{Q}} \simeq \mathbb{Q}(i)$ and therefore $\mathrm{End}(\mathrm{Jac}(C)_k)_{\mathbb{R}} \simeq \mathbb{C}$.

(iv) $D_{2,1}$: Since the nonhyperelliptic involution α of $C\colon y^2 = x^5 + x$ over $k = \mathbb{Q}$ is defined over $\mathbb{Q}$, we have $\mathrm{End}(\mathrm{Jac}(C)_{\mathbb{Q}})_{\mathbb{Q}} \simeq \mathbb{Q}\times\mathbb{Q}$ and therefore $\mathrm{End}(\mathrm{Jac}(C)_k)_{\mathbb{R}} \simeq \mathbb{R}\times\mathbb{R}$.

(v) $J(C_3)$: It is enough to show that the curve $C\colon y^2 = x^6 + 10x^3 - 2$ over $k = \mathbb{Q}(\sqrt{-3})$ has $\mathrm{End}(\mathrm{Jac}(C)_L)_{\mathbb{R}} \simeq \mathbb{H}$, where $L = \mathbb{Q}(a, \sqrt{-3})$ is the unique subfield of K with index 2 (here $a = \sqrt[3]{-2}$). It suffices to find automorphisms β and δ defined over L that satisfy $\beta^2 = \delta^2 = -1$ and $\beta\circ\delta = -\delta\circ\beta$. These automorphisms are

$$\beta(x, y) = \left(\frac{(-\sqrt{-3}+1)a^2x + 2(\sqrt{-3}+1)a}{4x + (\sqrt{-3}-1)a^2}, \frac{-3\cdot 2^5\sqrt{-3}y}{(4x + (\sqrt{-3}-1)a^2)^3}\right),$$

$$\delta(x, y) = \left(\frac{-a^2x - 2a}{2x + a^2}, \frac{12\sqrt{-3}y}{(2x + a^2)^3}\right).$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

(vi) $C_{6,1}$: For $C\colon y^2 = x^6 + 6x^5 - 30x^4 + 20x^3 + 15x^2 - 12x + 1$ over $k = \mathbb{Q}$, it suffices to show that $\mathrm{End}(\mathrm{Jac}(C)_{\mathbb{Q}(a)})_{\mathbb{R}} = \mathrm{M}_2(\mathbb{R})$, where $\mathbb{Q}(a)$ is the unique subfield of K with index 2. But this is clear: the nonhyperelliptic involution α is defined over $\mathbb{Q}(a)$ and the noncommuting element γ is also defined over $\mathbb{Q}(a)$ (in fact over $\mathbb{Q}$), and thus $\mathrm{End}(\mathrm{Jac}(C)_{\mathbb{Q}(a)})_{\mathbb{Q}} = \mathrm{M}_2(\mathbb{Q})$.

(vii) $J(D_3)$: The same argument used for $J(C_3)$ shows that the Jacobian of the curve $y^2 = x^6 + 10x^3 - 2$ over $\mathbb{Q}$ has Sato–Tate group $J(D_3)$.

(viii) $D_{6,1}$: For $C\colon y^2 = x^6 + 6x^5 - 30x^4 - 40x^3 + 60x^2 + 24x - 8$ over $k = \mathbb{Q}$, it suffices to show $\mathrm{End}(\mathrm{Jac}(C)_L)_{\mathbb{R}} = \mathrm{M}_2(\mathbb{R})$, where L is determined in the following way: it is the field fixed by the unique subgroup of order 2 in $\mathrm{Gal}(K/k)$ contained in the unique cyclic subgroup of order 6 in $\mathrm{Gal}(K/k)$. More explicitly, $L = \mathbb{Q}(a, \sqrt{6})$. It is enough to observe that

$$\beta(x,y) = \left(\frac{(-\frac{1}{2}a\sqrt{6} + \frac{1}{2}(a^2 + a - 2))x + 2}{x + \frac{1}{2}a\sqrt{6} + \frac{1}{2}(-a^2 - a + 2)}, \frac{((9a^2 + 6a - 8)\sqrt{6} - 9a^2 - 63a + 54)y}{(x + \frac{1}{2}a\sqrt{6} + \frac{1}{2}(-a^2 - a + 2))^3}\right),$$

$$\delta(x,y) = \left(\frac{(-a^2 - a + 8)x + 2}{x + a^2 + a - 8}, \frac{(18a^2 + 12a - 154)\sqrt{6}y}{(x + a^2 + a - 8)^3}\right)$$

do not commute and satisfy $\beta^2 = \delta^2 = 1$.

We now consider the five Sato–Tate groups corresponding to Galois types in case **D**, which are addressed in Table 11 using just two curve equations (over various number fields k). It is enough to observe the following. First, the Jacobian of the curve $C\colon y^2 = x^6 + 3x^4 + x^2 - 1$ splits over $\mathbb{Q}$ as the product of elliptic curves with CM by $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$, respectively. Indeed, using an algorithm of Gaudry and Schost [GS01], one proves that the j-invariants of the elliptic quotients of this curve are 1728 and 2000, which correspond to (nonisogenous) elliptic curves E_1 and E_2 defined over $\mathbb{Q}$ with CM by $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$, respectively. From Proposition 4.5, it follows that $K = \mathbb{Q}(i, \sqrt{2})$ and that $\mathrm{Jac}(C)$ is isogenous over $\mathbb{Q}$ to $E_1 \times E_2$. Second, for $C\colon y^2 = x^5 + 1$, it is well-known that $\mathrm{End}(\mathrm{Jac}(C)_K)_{\mathbb{Q}} \simeq \mathbb{Q}(\zeta_5)$, where $K = \mathbb{Q}(\zeta_5)$.

For the two Sato–Tate groups in case **C**, which are both addressed using the same curve equation, the above procedure shows that the Jacobian of the curve $y^2 = x^6 + 3x^4 - 2$ splits over $\mathbb{Q}$ as the product of an elliptic curve without CM and an elliptic curve with CM by $\mathbb{Q}(i)$ (now the j-invariants of the elliptic quotients are 3456 and 1728, respectively).

For the two Sato–Tate groups in case **B**, it is enough to check that the Jacobian of the curve $C_1\colon y^2 = x^6 + x^2 + 1$ splits over $\mathbb{Q}$ as the product of two nonisogenous curves without CM, while the Jacobian of the curve $C_2\colon y^2 = x^6 + x^5 + x - 1$ splits over $\mathbb{Q}(i)$ as the product of two nonisogenous curves without CM that are Galois conjugates. Indeed, the nonhyperelliptic involution $\alpha(x,y) = (-x, y)$ guarantees that $\mathrm{Jac}(C_1)$ splits over $\mathbb{Q}$ as the product of two elliptic curves E and E' over $\mathbb{Q}$, which do not have CM since their j-invariants are $-256/31$ and $6912/31$. Moreover, by Lemma 4.12 below, E and E' are not $\overline{\mathbb{Q}}$-isogenous, because

$$L_5(\mathrm{Jac}(C_1), T) = (1 - T + 5T^2)(1 + 3T + 5T^2).$$

The curve C_2 has a nonhyperelliptic involution $\alpha(x,y) = (-1/x, -iy/x^3)$, which shows that $\mathrm{Jac}(C_2)$ splits over $\mathbb{Q}(i)$ as the product of two elliptic curves E and E' over $\mathbb{Q}(i)$. These two

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

elliptic curves do not have CM, since their j-invariants are the roots of the polynomial

$$j^2 - \frac{5328000}{107}j + \frac{9826000000000}{11449}.$$

Moreover, they are not $\overline{\mathbb{Q}}$-isogenous, because

$$L_{13}(\mathrm{Jac}(C_2), T) = (1 - T + T^2)(1 + 5T + T^2).$$

Lemma 4.12. *Let A be an abelian surface over k for which there exists a field extension L/k such that $A_L \sim_L E \times E'$, where E and E' are elliptic curves defined over L without complex multiplication. Suppose there exists a prime $\mathfrak{p}$ of k, of good reduction for A and of residue degree 1 in L, such that $L_{\mathfrak{p}}(A, T)$ is not of the form $P(T) \cdot P(\pm T)$ for any degree 2 polynomial $P(T) \in \mathbb{Q}[T]$. Then E and E' are not $\overline{\mathbb{Q}}$-isogenous.*

Proof. Suppose that E and E' are $\overline{\mathbb{Q}}$-isogenous. Then there exists a quadratic extension L'/L such that $E' \sim_L E \otimes \chi$, where χ is a character (either trivial or quadratic) of $\mathrm{Gal}(L'/L)$. Let $\mathfrak{P}$ be a prime of L lying over $\mathfrak{p}$. It follows that

$$L_{\mathfrak{P}}(E', T) = L_{\mathfrak{P}}(E, \pm T).$$

Since $\mathfrak{p}$ has residue degree 1 in L, we have

$$L_{\mathfrak{p}}(A, T) = L_{\mathfrak{P}}(A_L, T) = L_{\mathfrak{P}}(E, T) \cdot L_{\mathfrak{P}}(E, \pm T),$$

which proves the claim. □

For case **A**, we note that the Galois group of $x^5 - x + 1$ is the symmetric group S_5. It follows from a theorem of Zarhin [Zar00] that for the curve C defined by $y^2 = x^5 - x + 1$ over $\mathbb{Q}$ we have $\mathrm{End}(\mathrm{Jac}(C)_K)_{\mathbb{Q}} = \mathbb{Q}$.

4.8.1 *Sato–Tate groups over field extensions.* The curves in Table 11 were chosen to minimize the degree of their fields of definition; this makes it necessary to use 34 distinct curves (some considered over multiple number fields). However, if one relaxes this restriction on the field of definition, one can reduce the number of curves by using the fact that every Sato–Tate group that can arise in genus 2 is conjugate (in $\mathrm{USp}(4)$) to a subgroup of one the groups

$$J(D_6), J(O), J(E_6), J(E_4), F_{ac}, F_{a,b}, N(\mathrm{U}(1) \times \mathrm{SU}(2)), N(\mathrm{SU}(2) \times \mathrm{SU}(2)), \mathrm{USp}(4).$$

One can thus realize all 52 Sato–Tate groups by considering just the 9 curves of Table 11 corresponding to these groups over the appropriate number field. To justify this, we recall that for an abelian surface A defined over k, Proposition 2.17 asserts that there is a bijection between the conjugacy classes of subgroups of $\mathrm{Gal}(K/k)$ and the conjugacy classes of subgroups of ST_A containing ST_A^0. This bijection is given by sending the class of the subgroup H to the conjugacy class of the group ST_{A_L}, where L is the fixed field K^H.

In Table 4, we illustrate how the Jacobian $A = \mathrm{Jac}(C)$ of the curve C of Table 11 corresponding to $J(O)$ realizes 24 other Sato–Tate groups when considering subextensions of K/k. We note that in this example there are 33 conjugacy classes of subgroups of $\mathrm{Gal}(K/k)$, and thus 33 conjugacy classes of subgroups of $J(O)$ containing $\mathrm{U}(1)$. However, considering conjugation in $\mathrm{USp}(4)$ yields exactly the 25 nonconjugate subgroups listed in Table 4. We leave the corresponding exercise for the other groups listed above to the reader.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 4. Sato–Tate groups realized by the Jacobian A of the curve defined by $y^2 = x^6 - 5x^4 + 10x^3 - 5x^2 + 2x - 1$ over extensions of $\mathbb{Q}$, where $a^3 - 4a + 4 = 0$, $b^4 + 22b + 22 = 0$, and $c^2 + a + 4 = 0$.

ST_{A_L}	L	ST_{A_L}	L
$J(O)$	$\mathbb{Q}$	$J(C_3)$	$\mathbb{Q}(b, \sqrt{-11})$
O	$\mathbb{Q}(\sqrt{-2})$	$D_{3,2}$	$\mathbb{Q}(b, \sqrt{22})$
$J(T)$	$\mathbb{Q}(\sqrt{-11})$	C_4	$\mathbb{Q}(c, \sqrt{-2})$
O_1	$\mathbb{Q}(\sqrt{22})$	$J(C_2)$	$\mathbb{Q}(c, \sqrt{-11})$
$J(D_4)$	$\mathbb{Q}(a)$	$C_{4,1}$	$\mathbb{Q}(c, \sqrt{22})$
T	$\mathbb{Q}(\sqrt{-2}, \sqrt{-11})$	$D_{2,1}$	$\mathbb{Q}(c\sqrt{-2}, \sqrt{-11})$
$J(D_3)$	$\mathbb{Q}(b)$	D_2	$\mathbb{Q}(c\sqrt{-11}, \sqrt{-2})$
D_4	$\mathbb{Q}(a, \sqrt{-2})$	C_3	$\mathbb{Q}(b, \sqrt{-2}, \sqrt{-11})$
$J(D_2)$	$\mathbb{Q}(a, \sqrt{-11})$	C_2	$\mathbb{Q}(a, b, \sqrt{-2})$
$D_{4,1}$	$\mathbb{Q}(a, \sqrt{22})$	$J(C_1)$	$\mathbb{Q}(a, b, \sqrt{-11})$
$J(C_4)$	$\mathbb{Q}(c)$	$C_{2,1}$	$\mathbb{Q}(a, b, \sqrt{22})$
$D_{4,2}$	$\mathbb{Q}(c\sqrt{-2})$	C_1	$\mathbb{Q}(a, b, \sqrt{-2}, \sqrt{-11})$
D_3	$\mathbb{Q}(b, \sqrt{-2})$		

5. Numerical verification

In this section, we describe some numerical verifications of the refined Sato–Tate conjecture for abelian surfaces.

5.1 Densities and moments

Using numerical computations, one can both provisionally identify the Sato–Tate group associated to a particular abelian surface (which can then be confirmed through analysis of the Galois type) and then test the equidistribution property predicted by the Sato–Tate conjecture. In order to do this, however, we need a way to numerically compare the observed distribution of normalized L-polynomials to the Sato–Tate prediction.

To facilitate this, we compute the distributions of the first and second coefficients of the characteristic polynomial of a random conjugacy class in each of the 55 groups named in Theorem 3.4 (including the three groups excluded by the comparison to Galois types in § 4.3), under the image of the Haar measure. These distributions can be described in two equivalent ways, via their *density functions* or their *moment sequences*. By computing these for the 55 groups, we see that the separate distributions of the first and second coefficients are already sufficient to distinguish the groups. Thus no joint statistics are needed, but as a matter of interest we give joint density functions for the six connected cases.

Remark 5.1. The moment statistics that we associate to a closed subgroup G of $\mathrm{USp}(2g)$ are integer symmetric polynomials in the eigenvalues of a matrix chosen uniformly over G. They are thus forced to be integers [KS09, Proposition 2].

Table 8 lists the real dimension d and the number of connected components $c = |G/G^0|$ for each group G, along with the associated endomorphism algebra $\mathrm{End}(A)_{\mathbb{R}}$. We also list invariants $z_1 = z_{1,0}$ and $z_2 = [z_{2,-2}, z_{2,-1}, z_{2,0}, z_{2,1}, z_{2,2}]$ defined by

$$\Pr[a_i = j] = z_{i,j}/c,$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 5. Joint density functions $c\sqrt{\max\{\rho(a_1,a_2),0\}}$ for the Sato–Tate groups $G_{1,1} = \mathrm{U}(1) \times \mathrm{U}(1)$, $G_{1,3} = \mathrm{U}(1) \times \mathrm{SU}(2)$, $G_{3,3} = \mathrm{SU}(2) \times \mathrm{SU}(2)$, and $\mathrm{USp}(4)$.

G	c	$\rho(a_1,a_2)$
$G_{1,1}$	$\frac{2}{\pi^2}$	$1/((a_1^2-4a_2+8)(a_2-2a_1+2)(a_2+2a_1+2))$
$G_{1,3}$	$\frac{1}{2\pi^2}$	$(4+2a_2-a_1^2)^2/((a_1^2-4a_2+8)(a_2-2a_1+2)(a_2+2a_1+2))$
$G_{3,3}$	$\frac{1}{2\pi^2}$	$(a_2-2a_1+2)(a_2+2a_1+2)/(a_1^2-4a_2+8)$
$\mathrm{USp}(4)$	$\frac{1}{4\pi^2}$	$(a_2-2a_1+2)(a_2+2a_1+2)(a_1^2-4a_2+8)$

where the random variables a_1 and a_2 denote the linear and quadratic coefficients, respectively, of the characteristic polynomial of a random conjugacy class in G. Additionally, we give the first three nontrivial moments $\mathrm{E}[a_1^2]$, $\mathrm{E}[a_1^4]$, $\mathrm{E}[a_1^6]$ and $\mathrm{E}[a_2], \mathrm{E}[a_2^2], \mathrm{E}[a_2^3]$ of a_1 and a_2. We note that the invariants d, c, z_1, z_2, and $\mathrm{E}[a_2]$ already suffice to uniquely distinguish each Sato–Tate group in genus 2.

To save space, we use the symbols G_1, G_3, $G_{1,1}$, $G_{1,3}$, and $G_{3,3}$ to identify the connected subgroups $G^0 = \mathrm{U}(1)$, $\mathrm{SU}(2)$, $\mathrm{U}(1) \times \mathrm{U}(1)$, $\mathrm{U}(1) \times \mathrm{SU}(2)$, and $\mathrm{SU}(2) \times \mathrm{SU}(2)$ of $\mathrm{USp}(4)$, respectively.

5.1.1 *Computing the distributions of a_1 and a_2.* Tables 9 and 10 give explicit formulas for the moments of a_1 and a_2. Here we describe the derivation of these formulas, as well as the computation of probability density functions for a_1 and a_2.

For $G = \mathrm{USp}(4)$, the moments of a_1 and a_2 may be directly computed using the Weyl integration formula, as in [KS08]. A bit of calculus shows that the joint density function of a_1 and a_2 is given by $\sqrt{\max\{\rho(a_1,a_2),0\}}/(4\pi^2)$, where

$$\rho(a_1,a_2) = (a_1^2 - 4a_2 + 8)(a_2 - 2a_1 + 2)(a_2 + 2a_1 + 2). \tag{5.1}$$

The support of the joint density function is the region[12] where ρ is nonnegative:

$$S = \{(a_1,a_2) \in \mathbb{R}^2 : a_2 \geqslant 2a_1 - 2,\ a_2 \geqslant -2a_1 - 2,\ a_2 \leqslant \tfrac{1}{4}a_1^2 + 2\}.$$

One recovers the density function for a_1 by integrating with respect to a_2, and vice versa; the results can be expressed in terms of complete elliptic integrals, because $\rho(a_1,a_2)$ is a polynomial of degree 4 in a_1 and of degree 3 in a_2. A plot of the joint density function for $\mathrm{USp}(4)$ can be found in Figure 2. A similar analysis can be applied to the groups $\mathrm{U}(1) \times \mathrm{U}(1)$, $\mathrm{U} \times \mathrm{SU}(2)$, and $\mathrm{SU}(2) \times \mathrm{SU}(2)$, using products of the appropriate measures; the results are tabulated in Table 5. In each of these cases the support of the joint density function is the two-dimensional region S.

In all other cases the support is one-dimensional, as may be seen in Table 6, which lists all the component density functions for a_1 and a_2 that can arise in genus 2. With the exception of $\mathrm{USp}(4)$, these distributions are all derived from the distributions $a_{1,\mathrm{U}(1)}$ and $a_{1,\mathrm{SU}(2)}$ that arise

[12] Serre points out that the parabolic arc bounding the top of this region corresponds precisely to the Hodge circle, and that the factorization of ρ is a special case of a general formula giving the product of the differentials of fundamental characters [Ste65, Lemma 8.2].

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

for the two connected Sato–Tate groups U(1) and SU(2) in genus 1 (where U(1) is embedded in $\mathrm{SU}(2) = \mathrm{USp}(2)$). We recall that the even moments are given by

$$\mathrm{E}[a_{1,\mathrm{U}(1)}^{2n}] = \binom{2n}{n}, \tag{5.2}$$

$$\mathrm{E}[a_{1,\mathrm{SU}(2)}^{2n}] = \frac{1}{n+1}\binom{2n}{n}, \tag{5.3}$$

while the odd moments are zero, and we have the density functions

$$\mathrm{dens}(a_{1,\mathrm{U}(1)} = t) = \frac{1}{\pi\sqrt{4-t^2}} \quad \text{for } |t| < 2, \tag{5.4}$$

$$\mathrm{dens}(a_{1,\mathrm{SU}(2)} = t) = \frac{\sqrt{4-t^2}}{2\pi} \quad \text{for } |t| < 2. \tag{5.5}$$

For convenience, we define the sequences

$$b_n = [X^n](X^2+1)^n \quad \text{and} \quad c_n = b_n/\left(\tfrac{n}{2}+1\right),$$

where $[X^n](X^2+1)^n$ denotes the coefficient of X^n in the expansion of $(X^2+1)^n$, so that $\mathrm{E}[a_{1,\mathrm{U}(1)}^n] = b_n$ and $\mathrm{E}[a_{1,\mathrm{SU}(2)}^n] = c_n$. These are sequences A126869 and A126120, respectively, in the *On-Line Encyclopedia of Integer Sequences* [OEIS].

Each group $G \subsetneq \mathrm{USp}(4)$ appearing in Theorem 3.4 may be expressed in the form $\langle G^0, H\rangle$, where H is a finite subgroup of $\mathrm{USp}(2g)$ whose intersection with G^0 is $\{\pm 1\}$, so that $G/G^0 \simeq H/\{\pm 1\}$ (in most cases H may be constructed by simply omitting G^0 from the list of generators given for G in § 3). After picking a representative $h \in H$ for each coset of $H/\{\pm 1\}$, the distributions of the coefficients of the characteristic polynomial $\sum a_i T^i$ of a random matrix gh may then be computed in terms of $a_{1,\mathrm{U}(1)}$ and $a_{1,\mathrm{SU}(2)}$, where $g \in G^0$ is distributed according to the Haar measure. This allows the moments and density functions of a_1 and a_2 to be computed for the component hG^0; averaging over the components yields results for G.

The derivation of the component distributions in the split cases, where G^0 is $\mathrm{U}(1)\times\mathrm{U}(1)$, $\mathrm{U}(1)\times\mathrm{SU}(2)$, or $\mathrm{SU}(2)\times\mathrm{SU}(2)$, is straightforward. The results are tabulated in Table 6 (see Table 5 for the joint distribution on the identity components). We now focus on the $G^0 = \mathrm{U}(1)$ and $G^0 = \mathrm{SU}(2)$ cases, which account for 42 of the 55 Sato–Tate groups in genus 2, including the most complicated cases. There are 76 distinct components contained in these groups, but only 20 different component distributions that arise, each of which is determined by a triple (G^0, s, r) that we now define. Let $\varphi : G/G^0 \to \{\pm 1\}$ be the (possibly trivial) homomorphism with kernel $G \cap Z$. For each component hG^0 of G, let $k = k(h) \in \{1, 2, 3, 4, 6\}$ be the order of hG^0 in G/G^0 when $s = s(h) = \varphi(hG^0) = 1$, and let k be the order of JhG^0 in $\langle G, J\rangle/G^0$ when $s = -1$ (see § 3.4 for the definitions of Z and J). Finally, we define $r = r(h) = \zeta_{2k} + \zeta_{2k}^{-1} \in \{-2, 0, 1, \sqrt{2}, \sqrt{3}\}$. Table 6 gives the component distributions of a_1 and a_2 for each triple (G^0, s, r), as well as the 10 component distributions that arise among the 13 remaining groups (30 component distributions in total). The notation $a'_{1,\mathrm{U}(1)}$ denotes an independent random variable with the same distribution as $a_{1,\mathrm{U}(1)}$, and similarly for $a'_{1,\mathrm{SU}(2)}$.

Using Table 6, one can determine the moments and probability density functions of a_1 and a_2 on each component, and then compute moments and density functions for any particular group G. The invariants z_1 and z_2, which simply count components on which a_1 or a_2 has a particular fixed value, are also easily derived from Table 6.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 6. Component distributions of a_1 and a_2. Here $G_1 = \mathrm{U}(1)$, $G_3 = \mathrm{SU}(2)$, $G_{1,1} = \mathrm{U}(1) \times \mathrm{U}(1)$, $G_{1,3} = \mathrm{U}(1) \times \mathrm{SU}(2)$, and $G_{3,3} = \mathrm{SU}(2) \times \mathrm{SU}(2)$. See § 3.6 for definitions of the matrices a, b, c, and J.

Component	a_1	a_2
$(G_1, 1, r)$	$ra_{1,\mathrm{U}(1)}$	$a_{1,\mathrm{U}(1)}^2 + r^2 - 2$
$(G_1, -1, r)$	0	$2 - r^2$
$(G_3, 1, r)$	$ra_{1,\mathrm{SU}(2)}$	$a_{1,\mathrm{SU}(2)} + r^2 - 2$
$(G_3, -1, r)$	0	$2 - a_{1,\mathrm{SU}(2)}^2$
$G_{1,1}$	$a_{1,\mathrm{U}(1)} + a'_{1,\mathrm{U}(1)}$	$a_{1,\mathrm{U}(1)}a'_{1,\mathrm{U}(1)} + 2$
$aG_{1,1}, bG_{1,1}, (ac)^2G_{1,1}$	$a_{1,\mathrm{U}(1)}$	2
$abG_{1,1}$	0	2
$cG_{1,1}$	0	$a_{1,\mathrm{U}(1)}$
$acG_{1,1}, (ac)^3G_{1,1}$	0	0
$G_{1,3}$	$a_{1,\mathrm{U}(1)} + a_{1,\mathrm{SU}(2)}$	$a_{1,\mathrm{U}(1)}a_{1,\mathrm{SU}(2)} + 2$
$aG_{1,3}$	$a_{1,\mathrm{U}(1)}$	2
$G_{3,3}$	$a_{1,\mathrm{SU}(2)} + a'_{1,\mathrm{SU}(2)}$	$a_{1,\mathrm{SU}(2)}a'_{1,\mathrm{SU}(2)} + 2$
$JG_{3,3}$	0	$a_{1,\mathrm{SU}(2)}$
$\mathrm{USp}(4)$	$a_{1,\mathrm{USp}(4)}$	$a_{2,\mathrm{USp}(4)}$

To simplify the moment formulas, for $i = 0, 1, 2, 3, 4$ we define the sequences

$$b_{i,n} = [X^n](X^2 + iX + 1)^n,$$

which for $i = 0, 1, 2, 3, 4$ correspond to sequences A126869, A0002426, A000984, A026375, A081671 in the OEIS, respectively, and we note that $b_n = b_{0,n}$. We also define the sequences

$$d_{i,n} = [X^n](X^2 + iX + 1)^n - [X^{n+1}](X^2 + iX + 1)^n,$$

which for $i = 0, 1, 2, 3, 4$ correspond to sequences A126930, A005043, A000108, A007317, A064613, respectively, and we may write d_n for $d_{0,n}$. Additionally, we let

$$\hat{b}_n = \sum_k \binom{n}{k} 2^{n-k} b_k^2 \quad \text{and} \quad \hat{c}_n = \sum_k \binom{n}{k} 2^{n-k} c_k^2.$$

5.1.2 *Component distributions in the case $G^0 = \mathrm{U}(1)$.* We now consider the component distributions of a_1 and a_2 when $G^0 = \mathrm{U}(1)$. From Table 6 we see that $a_1 = ra_{1,\mathrm{U}(1)}$ when $s = 1$, and $a_1 = 0$ otherwise. On a component hG^0 with $s = 1$, the moments of a_1 are given by $\mathrm{E}_h[a_1^n] = r^n b_n$, where the subscript h identifies the component and thus determines r and s. The density function for a_1 on hG^0 is

$$\mathrm{dens}_h(a_1 = t) = \begin{cases} (\pi\sqrt{4r^2 - t^2})^{-1} & \text{if } t^2 < 4r^2 \text{ and } s = 1, \\ \delta_0 & \text{if } r = 0 \text{ or } s = -1, \\ 0 & \text{otherwise,} \end{cases} \tag{5.6}$$

where δ_k denotes the Dirac delta function centered at k.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

For a_2 we have $a_2 = a_{1,\mathrm{U}(1)}^2 + r^2 - 2$ when $s = 1$ and $a_2 = 2 - r^2$ otherwise. The moments of a_2 on hG^0 are given by

$$\mathrm{E}_h[a_2^n] = \begin{cases} \displaystyle\sum_{k=0}^{n} \binom{n}{k} b_{2k}(r^2-2)^{n-k} = b_{r^2,n} & \text{if } s = 1, \\ (2-r^2)^n & \text{if } s = -1, \end{cases} \tag{5.7}$$

and its density function is

$$\mathrm{dens}_h(a_2 = t) = \begin{cases} (\pi\sqrt{4-(r^2-t)^2})^{-1} & \text{if } (r^2-t)^2 < 4 \text{ and } s = 1, \\ \delta_{2-r^2} & \text{if } s = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

For any particular Sato–Tate group G, the moment sequences and density functions of a_1 and a_2 are then computed by averaging over the components. Taking $G = T$ as an example, averaging over the 12 components of G yields

$$\mathrm{E}[a_2^n] = \tfrac{1}{12}(b_{4,n} + 3b_n + 8b_{1,n}),$$

as listed in Table 10. A plot of the a_2 density function appears in Figure 1.

5.1.3 *Component distributions in the case $G^0 = \mathrm{SU}(2)$.* In this case we have $a_1 = r a_{1,\mathrm{SU}(2)}$ when $s = 1$ and $a_1 = 0$ otherwise. The nth moment of a_1 on hG^0 is $\mathrm{E}_h[a_1^n] = r^n c_n$, and its density function is

$$\mathrm{dens}_h(a_1 = t) = \begin{cases} \dfrac{1}{2\pi r^2}\sqrt{4r^2 - t^2} & \text{if } t^2 < 4r^2 \text{ and } s = 1, \\ \delta_0 & \text{if } r = 0 \text{ or } s = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

For a_2 we have $a_2 = a_{1,\mathrm{SU}(2)}^2 + r^2 - 2$ when $s = 1$ and $a_2 = 2 - a_{1,\mathrm{SU}(2)}^2$ otherwise. The moments of a_2 on hG^0 are given by

$$\mathrm{E}_h[a_2^n] = \begin{cases} \displaystyle\sum_{k=0}^{n} \binom{n}{k} c_{2k}(r^2-2)^{n-k} = d_{r^2,n} & \text{if } s = 1, \\ \displaystyle\sum_{k=0}^{n} \binom{n}{k} (-1)^k c_{2k} 2^{n-k} = (-1)^n d_n & \text{if } s = -1, \end{cases} \tag{5.10}$$

and its density function is

$$\mathrm{dens}_h(a_2 = t) = \begin{cases} \dfrac{1}{2\pi}\sqrt{4/(t-r^2+2) - 1} & \text{if } |t - r^2| < 2 \text{ and } s = 1, \\ \dfrac{1}{2\pi}\sqrt{4/(2-t) - 1} & \text{if } |t| < 2 \text{ and } s = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

5.2 Testing the refined conjecture for genus 2 curves

Let us now see how the densities and moments computed in § 5.1 can be used to numerically test the refined Sato–Tate conjecture; this will also provide some indication of how we assembled

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 7. Moment statistics for $y^2 = x^6 + 6x^5 - 20x^4 + 20x^3 - 20x^2 - 8x + 8$ over $\mathbb{Q}(\sqrt{-2})$.

	a_1				a_2				
N	M_2	M_4	M_6	M_8	M_1	M_2	M_3	M_4	M_5
10	1.79	11.74	134.59	1894.71	0.81	3.55	10.82	51.55	256.29
12	1.90	11.87	127.24	1714.90	0.91	3.81	11.46	52.13	246.15
14	2.01	12.54	130.43	1701.09	0.97	4.02	12.23	54.66	253.10
16	1.94	10.93	102.75	1266.03	0.97	3.86	11.05	46.56	203.16
18	1.96	11.45	111.89	1417.51	0.99	3.92	11.54	49.33	220.74
20	1.99	11.87	118.12	1513.12	1.00	3.98	11.88	51.37	232.49
22	1.99	11.89	118.64	1522.75	1.00	3.98	11.90	51.51	233.49
24	2.00	11.95	119.20	1528.43	1.00	3.99	11.96	51.74	234.53
26	2.00	11.99	119.80	1537.06	1.00	4.00	11.99	51.93	235.62
28	2.00	12.00	119.91	1538.51	1.00	4.00	12.00	51.97	235.83
30	2.00	12.00	120.02	1540.29	1.00	4.00	12.00	52.00	236.03
	2	12	120	1540	1	4	12	52	236

the numerical evidence needed to formulate the conjecture in the first place. We limit ourselves to testing the equidistribution of normalized L-polynomials (rather than the equidistribution of classes in $\mathrm{Conj}(\mathrm{ST}_A)$ described in Conjecture 1.1), and we consider only Jacobians of curves of genus 2.

We first discuss how to test the conjecture for a single genus 2 curve C/k. For a given bound N, we consider the primes $\mathfrak{p}$ of norm $q = \|\mathfrak{p}\| \leqslant N$ at which C has good reduction. For each $\mathfrak{p}$ we compute normalized L-polynomial coefficients

$$\begin{aligned} a_1(\mathfrak{p}) &= q^{-1/2}\big(\#C(\mathbb{F}_q) - q - 1\big), \\ a_2(\mathfrak{p}) &= q^{-1}\big(\#C(\mathbb{F}_{q^2}) + (\#C(\mathbb{F}_q) - q - 1)^2 - q^2 - 1\big)/2, \end{aligned} \tag{5.12}$$

and then calculate *moment statistics* $M_n(a_i)$ as the mean values of $a_i(\mathfrak{p})^n$ over $\|\mathfrak{p}\| \leqslant N$. The refined Sato–Tate conjecture implies that, as $N \to \infty$, each moment statistic $M_n(a_i)$ must converge to the moment $\mathrm{E}[a_i^n]$ for the Sato–Tate group $\mathrm{ST}_{\mathrm{Jac}(C)} \subseteq \mathrm{USp}(4)$. We can test this numerically for the first several values of n by comparing $M_n(a_i)$ to $\mathrm{E}[a_i^n]$ as we increase the bound N. In addition, we may plot histograms of the $a_i(\mathfrak{p})$ and compare the results to the density function for a_i in $\mathrm{ST}_{\mathrm{Jac}(C)}$.

For each of the example curves listed in Table 11, we have prepared animated histograms of the a_1 and a_2 distributions demonstrating the convergence to the conjectured Sato–Tate distribution, at the level of both densities and moments. These can be found online at

http://math.mit.edu/~drew.

We give one representative example here, using the curve $C/\mathbb{Q}(\sqrt{-2})$ with $\mathrm{ST}_{\mathrm{Jac}(C)} = T$ listed in Table 11. We computed moment statistics $M_n(a_i)$ for this curve using $N = 2^k$, with k ranging from 10 to 30, the first several of which are listed in Table 7. For comparison, the last line of the table lists the corresponding moments for the Sato–Tate group T.

The evident convergence of the moment statistics of C to the moments of $\mathrm{ST}_{\mathrm{Jac}(C)}$ extends to moments well beyond the range of the table. With $N = 2^{30}$, the first 20 moment statistics of a_1 and a_2 for C agree with the corresponding moments of $\mathrm{ST}_{\mathrm{Jac}(C)}$ with a relative error of approximately 0.1% (ignoring the even moments $\mathrm{E}[a_1^{2k}] = 0$). By contrast, the best agreement one finds using the moments of any of the other genus 2 Sato–Tate groups is worse than 40%.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

In particular, the corresponding moments of USp(4) are dramatically different: $\mathrm{E}[a_1^8]$ is only 84, rather than 1540, for example.

Histograms of $a_2(\mathfrak{p})$ values for $\|\mathfrak{p}\| \leqslant N = 2^k$ with $k = 12, 14, 16, \ldots, 30$ are shown in Figure 1, together with the a_2 density function for $\mathrm{ST}_{\mathrm{Jac}(C)} = T$. One can see the histogram data steadily converging toward the density function. Indeed, when $N = 2^{30}$, the histogram data matches the density function so closely that it is difficult to distinguish the two.

We may also compare the joint statistics of $a_1(\mathfrak{p})$ and $a_2(\mathfrak{p})$ for a given curve C with the joint density function of a_1 and a_2 for the corresponding Sato–Tate group $\mathrm{ST}_{\mathrm{Jac}(C)}$. Figure 2 shows a plot of these joint statistics for the curve $y^2 = x^5 - x + 1$, using the bound $\|\mathfrak{p}\| \leqslant 2^{30}$, alongside a plot of the joint density function for its Sato–Tate group USp(4), computed using (5.1) and plotted at the same number of points.

With $N = 2^{30}$, we must compute $a_i(\mathfrak{p})$ for more than 54 million values of $\mathfrak{p}$ (for each curve). To make such computations practical, we employ the optimizations described in [KS08], as well as several further improvements recently incorporated in the `smalljac` software library [Sut11a]; the most notable of these is the use of ideas in [GHM08] to efficiently implement the group operation in the Jacobian of curves defined by a sextic equation. For all but very small values of $\|\mathfrak{p}\|$, we do not use (5.12) to compute $a_i(\mathfrak{p})$ but instead apply

$$a_1(\mathfrak{p}) = q^{-1/2}\frac{\#\operatorname{Jac}(C)(\mathbb{F}_q) - \#\operatorname{Jac}(\tilde{C})(\mathbb{F}_q)}{2(q+1)}, \tag{5.13}$$

$$a_2(\mathfrak{p}) = q^{-1}\frac{\#\operatorname{Jac}(C)(\mathbb{F}_q) + \#\operatorname{Jac}(\tilde{C})(\mathbb{F}_q) - 2(q^2+1)}{2}.$$

Here $\tilde{C}$ denotes a nonisomorphic quadratic twist of C over $\mathbb{F}_q$. The computations of the group orders $\#\operatorname{Jac}(C)$ and $\#\operatorname{Jac}(\tilde{C})$ are performed using generic group algorithms described in [Sut07] and [Sut11b]. As discussed in [KS08], the asymptotically faster p-adic and ℓ-adic methods available are not practically faster in genus 2 for the range of N considered here.

In cases where $k \neq \mathbb{Q}$, we may take advantage of the fact that the moment statistics are essentially determined by the degree 1 primes $\mathfrak{p}$, allowing us to work entirely over prime fields. We can also exploit the situation where C/k is actually defined over $\mathbb{Q}$, in which case $a_i(\mathfrak{p})$ depends only on $p = \|\mathfrak{p}\|$. In this situation it suffices to compute $a_i(\mathfrak{p})$ for just one prime of norm p and then weight it with the correct multiplicity, as determined by the number of linear factors of a defining polynomial for $k/\mathbb{Q}$ in $\mathbb{F}_p[x]$.

5.3 An exhaustive search

The general methodology described above allows us to numerically test the Sato–Tate conjecture for individual curves. Using more specialized techniques, we can efficiently analyze L-polynomial distributions for many curves at once. This was originally done to provide an empirical conjecture for the classification of Sato–Tate groups; it now provides a partial check of the completeness of this classification. Of course, one cannot really trust in the completeness without the proofs of the theorems; after all, a similar search described in [KS09] found considerably fewer groups.[13]

To search for curves with exceptional L-polynomial distributions, we considered every nonsingular curve of the form $y^2 = f(x)$ where f is a monic polynomial of degree 5 or 6 whose coefficients lie in the interval $[-128, 128)$. This amounts to more than 2^{48} distinct curve equations,

[13] The search in [KS09] only looked at a_1-distributions of genus 2 curves over $\mathbb{Q}$, finding 23 of the 26 distributions identified here.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

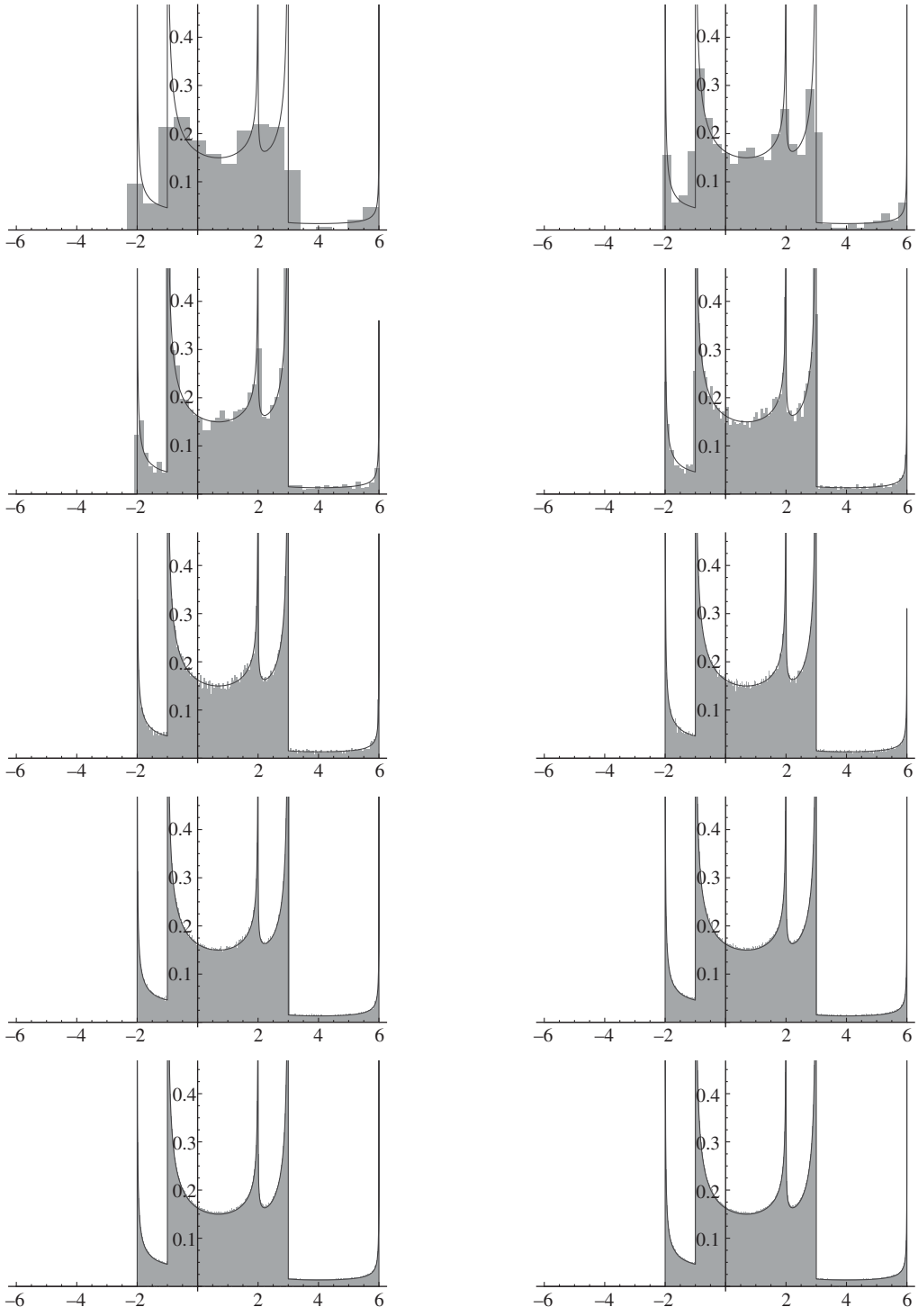

FIGURE 1. The a_2 density function for Sato–Tate group T and $a_2(\mathfrak{p})$ histograms for the curve $y^2 = x^6 + 6x^5 - 20x^4 + 20x^3 - 20x^2 - 8x + 8$ over $\mathbb{Q}(\sqrt{-2})$ for $\|\mathfrak{p}\| \leqslant 2^N$, with $N = 12, 14, \ldots, 30$.

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

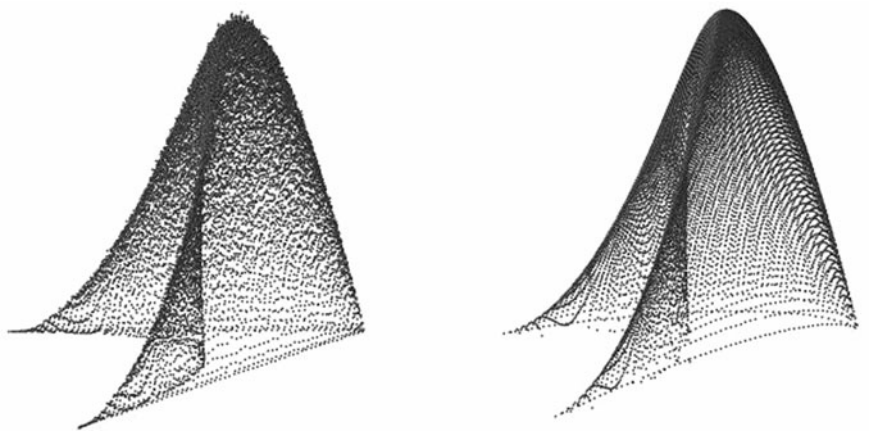

FIGURE 2. Joint $a_1(\mathfrak{p})$ and $a_2(\mathfrak{p})$ statistics for the curve $y^2 = x^5 - x + 1$ (left), and the joint density function for the Sato–Tate group USp(4) (right). The vertical scale is exaggerated; the peak at $a_1 = 0$ and $a_2 = 2/3$ has height $8/(\pi^2\sqrt{27})$.

of which approximately 2^{47} are nonisomorphic. This is much a larger range than was used in [KS09], which examined some 2^{35} curves, yet it actually required less computational effort. Here we summarize some of the optimizations that made this possible.

Of the 34 Sato–Tate groups that can arise for a genus 2 curve over $\mathbb{Q}$, all but eight have a density $z_1/c \geqslant 1/2$ of zero traces. These eight exceptions correspond to the first eight distributions listed in [KS09, Tables 11 and 13], where one can already find representative curves with small coefficients. Thus we chose to focus our search on curves with $z_1/c \geqslant 1/2$, allowing us to quickly discard curves that do not exhibit an abundance of zero traces at small primes.

For a suitably chosen bound B, we imposed the constraint $\pi(B) - 2z(C, B) \leqslant 3$, where $\pi(B)$ counts the primes $p \leqslant B$ and $z(C, B)$ counts the primes $p \leqslant B$ where C has good reduction and $\#C(\mathbb{F}_p) = p + 1$. By initially checking this constraint for a small value of B, we can very quickly discard the vast majority of curves. With this procedure, we ignore some curves with $z_1/c \geqslant 1/2$, but on average we expect to discard no more than half of the exceptional curves that we seek.

To distinguish exceptional curves, and to provisionally identify the Sato–Tate group G for each curve C, we computed various statistics for C up to a larger bound B to obtain a 'signature' $\sigma(C, B)$ that could be compared to signatures $\sigma(G)$ derived from the group invariants defined in § 5.1. Let $\hat{z}_{i,j}$ denote the integer $48z_{i,j}/c$. We define $\sigma(G)$ to be the tuple of integers

$$\sigma(G) = (\hat{z}_{1,0}, \hat{z}_{2,-2}, \hat{z}_{2,-1}, \hat{z}_{2,0}, \hat{z}_{2,1}, \hat{z}_{2,2}, \mathrm{E}[a_1^2], \mathrm{E}[a_2^4], \mathrm{E}[a_2], \mathrm{E}[a_2^2], \mathrm{E}[a_2^3]),$$

which suffices to uniquely distinguish all 55 of the Sato–Tate groups listed in Theorem 3.4. Given a curve C and a bound B, one can compute the analogous tuple $\sigma(C, B)$ by computing the corresponding statistics and rounding to the nearest integer. We note that the number of components c is typically not known *a priori*, but the ratios $z_{i,j}/c$ and the corresponding values of $\hat{z}_{i,j}$ can be computed without knowing c.

We now outline the search algorithm for exceptional curves C of the form $y^2 = f(x)$, where $f(x) = \sum f_i x^i$ is a monic sextic with $f_i \in I = [-R, R)$ and $f_5 \geqslant 0$, using bounds B_1, B_2, and B_3. For each combination of f_2, f_3, f_4, f_5 we perform the following three steps.

(1) For odd primes $p \leqslant B_1$, count points on the curve $C/\mathbb{F}_p$ defined by $y^2 = f(x)$ for every value of f_0, $f_1 \in \mathbb{F}_p$, using the method of [KS08, § 3]. Let $z_p(f_0, f_1) = 1$ if $\#C(\mathbb{F}_p) = p + 1$ and 0

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

otherwise. For each $f_0, f_1 \in I$, compute $z(C, B_1) = \sum_p z_p(f_0, f_1)$ for the curve $C/\mathbb{Q}$ defined by $y^2 = f(x)$. If $2z(C, B_1) < \pi(B_1) - 3$, then reject C.

(2) For each remaining curve C, compute $z(C, B_2)$. If $2z(C, B_2) < \pi(B_2) - 3$, then reject C.

(3) For each remaining curve C, initialize B to B_3 and compute $\sigma(C, B)$ using L-polynomial data for C at primes $p \leqslant B_3$. Then increase B by 50% and repeat until $\sigma(C, B)$ is stable for three consecutive values of B.

In our search, we used $R = 128$ as the coefficient bound, and used the prime bounds $B_1 = 83$, $B_2 = 1229$, and $B_3 = 2741$ in each of the three steps, values that were chosen after some initial performance testing. Using $B_1 = 83$, fewer than 1 in 100 000 curves pass step 1, and the average time spent on each curve is very small: about 100 ns on a 3.0 GHz AMD Phenom II core. With $B_2 = 1229$, fewer than 1 in 100 of the curves that pass step 1 also pass step 2. Thus, out of a total of 2^{48} curves, we only needed to compute signatures for some ten million curves. On average, this takes 1–2 s per curve, although in particularly difficult cases it may take as much as a minute. Overall, we spent an average of less than 200 ns per curve.

The search found curves with matching signatures for all 26 of the 34 genus 2 Sato–Tate groups over $\mathbb{Q}$ that have $z/c \geqslant 1/2$. Indeed, we found at least three curves for each group that are not isomorphic over $\mathbb{Q}$. As can be seen in Table 11, in each case we found a representative curve with integer coefficients of absolute value at most 60. Thus, *a posteriori*, we see that we could have used $R = 64$ rather than $R = 128$, which would have reduced the search time dramatically.

6. Tables

In this section, we give tables listing the Sato–Tate groups identified in § 3 (Table 8), the moments of a_1 and a_2 computed in § 5.1.1 (Tables 9 and 10), the curves analyzed in § 4.8 realizing each Sato–Tate group (Table 11), and the automorphism data needed to verify these Sato–Tate groups (Tables 12 and 13). To make the tables comprehensible, we recall in detail what data is tabulated.

In Table 8, each line corresponds to one of the 55 groups G named in Theorem 1.2. The quantities d and c indicate the dimension of G and the order of the component group G/G^0, whose isomorphism class is also given. To partially determine the Galois type, we list the $\mathbb{R}$-algebra $\mathrm{End}(A)_{\mathbb{R}}$, i.e., the fixed subalgebra of $\mathrm{End}(A_K)_{\mathbb{R}}$ under the action of $\mathrm{Gal}(K/k)$ as determined using Proposition 2.19. (It is not necessary to list the fixed subalgebras under subgroups of $\mathrm{Gal}(K/k)$, as this can be inferred from the rows of the table corresponding to those subgroups.) We also list the label associated to the Galois type by Theorem 4.3 (or * in the three cases that cannot arise from abelian surfaces). The quantities z_1 and z_2 count components of H on which a_1 and a_2 are constant; see § 5.2. The quantities $M[a_1^2]$ and $M[a_2]$ are some initial terms of moment sequences that are described more thoroughly in the tables that follow.

Tables 9 and 10 provide explicit formulas and initial terms for the a_1 and a_2 moment sequences associated to each group in Table 8, as computed using the methods of § 5.1. Note that the 55 groups only give rise to 37 distinct a_1 moment sequences, corresponding to 37 distinct Sato–Tate trace distributions, of which 26 can arise over $\mathbb{Q}$. Each of these distributions has been assigned an identifier of the form $\#N$ consistent with the numbering used in [KS09]; note that only the indices 1–23 correspond to distributions found in [KS09]. By contrast, the a_2 moment sequences in Table 10 are all distinct with one exception: the groups F_a and F_{ab} have identical a_2 distributions (but distinct a_1 distributions).

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 8. Sato–Tate groups in genus 2.

d	c	G	$[G/G^0]$	$\mathrm{End}(A)_{\mathbb{R}}$	Galois type	z_1	z_2	$M[a_1^2]$	$M[a_2]$
1	1	C_1	C_1	$\mathrm{M}_2(\mathbb{C})$	$\mathbf{F}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	8, 96, 1280	4, 18, 88
1	2	C_2	C_2	$\mathbb{C}\times\mathbb{C}$	$\mathbf{F}[\mathrm{C}_2]$	1	0, 0, 0, 0, 0	4, 48, 640	2, 10, 44
1	3	C_3	C_3	$\mathbb{C}\times\mathbb{C}$	$\mathbf{F}[\mathrm{C}_3]$	0	0, 0, 0, 0, 0	4, 36, 440	2, 8, 34
1	4	C_4	C_4	$\mathbb{C}\times\mathbb{C}$	$\mathbf{F}[\mathrm{C}_4]$	1	0, 0, 0, 0, 0	4, 36, 400	2, 8, 32
1	6	C_6	C_6	$\mathbb{C}\times\mathbb{C}$	$\mathbf{F}[\mathrm{C}_6]$	1	0, 0, 0, 0, 0	4, 36, 400	2, 8, 32
1	4	D_2	D_2	$\mathbb{C}$	$\mathbf{F}[\mathrm{D}_2]$	3	0, 0, 0, 0, 0	2, 24, 320	1, 6, 22
1	6	D_3	D_3	$\mathbb{C}$	$\mathbf{F}[\mathrm{D}_3]$	3	0, 0, 0, 0, 0	2, 18, 220	1, 5, 17
1	8	D_4	D_4	$\mathbb{C}$	$\mathbf{F}[\mathrm{D}_4]$	5	0, 0, 0, 0, 0	2, 18, 200	1, 5, 16
1	12	D_6	D_6	$\mathbb{C}$	$\mathbf{F}[\mathrm{D}_6]$	7	0, 0, 0, 0, 0	2, 18, 200	1, 5, 16
1	12	T	A_4	$\mathbb{C}$	$\mathbf{F}[\mathrm{A}_4]$	3	0, 0, 0, 0, 0	2, 12, 120	1, 4, 12
1	24	O	S_4	$\mathbb{C}$	$\mathbf{F}[\mathrm{S}_4]$	9	0, 0, 0, 0, 0	2, 12, 100	1, 4, 11
1	2	$J(C_1)$	C_2	$\mathbb{H}$	$\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathbb{H}]$	1	1, 0, 0, 0, 0	4, 48, 640	1, 11, 40
1	4	$J(C_2)$	D_2	$\mathbb{C}$	$\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathbb{H}]$	3	1, 0, 0, 0, 1	2, 24, 320	1, 7, 22
1	6	$J(C_3)$	C_6	$\mathbb{C}$	$\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathbb{H}]$	3	1, 0, 0, 2, 0	2, 18, 220	1, 5, 16
1	8	$J(C_4)$	$\mathrm{C}_4\times\mathrm{C}_2$	$\mathbb{C}$	$\mathbf{F}[\mathrm{C}_4\times\mathrm{C}_2, \mathrm{C}_4]$	5	1, 0, 2, 0, 1	2, 18, 200	1, 5, 16
1	12	$J(C_6)$	$\mathrm{C}_6\times\mathrm{C}_2$	$\mathbb{C}$	$\mathbf{F}[\mathrm{C}_6\times\mathrm{C}_2, \mathrm{C}_6]$	7	1, 2, 0, 2, 1	2, 18, 200	1, 5, 16
1	8	$J(D_2)$	$\mathrm{D}_2\times\mathrm{C}_2$	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_2\times\mathrm{C}_2, \mathrm{D}_2]$	7	1, 0, 0, 0, 3	1, 12, 160	1, 5, 13
1	12	$J(D_3)$	D_6	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathbb{H}]$	9	1, 0, 0, 2, 3	1, 9, 110	1, 4, 10
1	16	$J(D_4)$	$\mathrm{D}_4\times\mathrm{C}_2$	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_4\times\mathrm{C}_2, \mathrm{D}_4]$	13	1, 0, 2, 0, 5	1, 9, 100	1, 4, 10
1	24	$J(D_6)$	$\mathrm{D}_6\times\mathrm{C}_2$	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_6\times\mathrm{C}_2, \mathrm{D}_6]$	19	1, 2, 0, 2, 7	1, 9, 100	1, 4, 10
1	24	$J(T)$	$\mathrm{A}_4\times\mathrm{C}_2$	$\mathbb{R}$	$\mathbf{F}[\mathrm{A}_4\times\mathrm{C}_2, \mathrm{A}_4]$	15	1, 0, 0, 8, 3	1, 6, 60	1, 3, 7
1	48	$J(O)$	$\mathrm{S}_4\times\mathrm{C}_2$	$\mathbb{R}$	$\mathbf{F}[\mathrm{S}_4\times\mathrm{C}_2, \mathrm{S}_4]$	33	1, 0, 6, 8, 9	1, 6, 50	1, 3, 7
1	2	$C_{2,1}$	C_2	$\mathrm{M}_2(\mathbb{R})$	$\mathbf{F}[\mathrm{C}_2, \mathrm{C}_1, \mathrm{M}_2(\mathbb{R})]$	1	0, 0, 0, 0, 1	4, 48, 640	3, 11, 48
1	4	$C_{4,1}$	C_4	$\mathbb{C}$	$\mathbf{F}[\mathrm{C}_4, \mathrm{C}_2]$	3	0, 0, 2, 0, 0	2, 24, 320	1, 5, 22
1	6	$C_{6,1}$	C_6	$\mathbb{C}$	$\mathbf{F}[\mathrm{C}_6, \mathrm{C}_3, \mathrm{M}_2(\mathbb{R})]$	3	0, 2, 0, 0, 1	2, 18, 220	1, 5, 18
1	4	$D_{2,1}$	D_2	$\mathbb{R}\times\mathbb{R}$	$\mathbf{F}[\mathrm{D}_2, \mathrm{C}_2, \mathrm{M}_2(\mathbb{R})]$	3	0, 0, 0, 0, 2	2, 24, 320	2, 7, 26
1	8	$D_{4,1}$	D_4	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_4, \mathrm{D}_2]$	7	0, 0, 2, 0, 2	1, 12, 160	1, 4, 13
1	12	$D_{6,1}$	D_6	$\mathbb{R}$	$\mathbf{F}[\mathrm{D}_6, \mathrm{D}_3, \mathrm{M}_2(\mathbb{R})]$	9	0, 2, 0, 0, 4	1, 9, 110	1, 4, 11
1	6	$D_{3,2}$	D_3	$\mathbb{R}\times\mathbb{R}$	$\mathbf{F}[\mathrm{D}_3, \mathrm{C}_3]$	3	0, 0, 0, 0, 3	2, 18, 220	2, 6, 21
1	8	$D_{4,2}$	D_4	$\mathbb{R}\times\mathbb{R}$	$\mathbf{F}[\mathrm{D}_4, \mathrm{C}_4]$	5	0, 0, 0, 0, 4	2, 18, 200	2, 6, 20
1	12	$D_{6,2}$	D_6	$\mathbb{R}\times\mathbb{R}$	$\mathbf{F}[\mathrm{D}_6, \mathrm{C}_6]$	7	0, 0, 0, 0, 6	2, 18, 200	2, 6, 20
1	24	O_1	S_4	$\mathbb{R}$	$\mathbf{F}[\mathrm{S}_4, \mathrm{A}_4]$	15	0, 0, 6, 0, 6	1, 6, 60	1, 3, 8
3	1	E_1	C_1	$\mathrm{M}_2(\mathbb{R})$	$\mathbf{E}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	4, 32, 320	3, 10, 37
3	2	E_2	C_2	$\mathbb{C}$	$\mathbf{E}[\mathrm{C}_2, \mathbb{C}]$	1	0, 0, 0, 0, 0	2, 16, 160	1, 6, 17
3	3	E_3	C_3	$\mathbb{C}$	$\mathbf{E}[\mathrm{C}_3]$	0	0, 0, 0, 0, 0	2, 12, 110	1, 4, 13
3	4	E_4	C_4	$\mathbb{C}$	$\mathbf{E}[\mathrm{C}_4]$	1	0, 0, 0, 0, 0	2, 12, 100	1, 4, 11
3	6	E_6	C_6	$\mathbb{C}$	$\mathbf{E}[\mathrm{C}_6]$	1	0, 0, 0, 0, 0	2, 12, 100	1, 4, 11
3	2	$J(E_1)$	C_2	$\mathbb{R}\times\mathbb{R}$	$\mathbf{E}[\mathrm{C}_2, \mathbb{R}\times\mathbb{R}]$	1	0, 0, 0, 0, 0	2, 16, 160	2, 6, 20
3	4	$J(E_2)$	D_2	$\mathbb{R}$	$\mathbf{E}[\mathrm{D}_2]$	3	0, 0, 0, 0, 0	1, 8, 80	1, 4, 10
3	6	$J(E_3)$	D_3	$\mathbb{R}$	$\mathbf{E}[\mathrm{D}_3]$	3	0, 0, 0, 0, 0	1, 6, 55	1, 3, 8
3	8	$J(E_4)$	D_4	$\mathbb{R}$	$\mathbf{E}[\mathrm{D}_4]$	5	0, 0, 0, 0, 0	1, 6, 50	1, 3, 7
3	12	$J(E_6)$	D_6	$\mathbb{R}$	$\mathbf{E}[\mathrm{D}_6]$	7	0, 0, 0, 0, 0	1, 6, 50	1, 3, 7
2	1	F	C_1	$\mathbb{C}\times\mathbb{C}$	$\mathbf{D}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	4, 36, 400	2, 8, 32
2	2	F_a	C_2	$\mathbb{R}\times\mathbb{C}$	$\mathbf{D}[\mathrm{C}_2, \mathbb{R}\times\mathbb{C}]$	0	0, 0, 0, 0, 1	3, 21, 210	2, 6, 20
2	2	F_c	C_2	*	*	1	0, 0, 0, 0, 0	2, 18, 200	1, 5, 16
2	2	F_{ab}	C_2	$\mathbb{R}\times\mathbb{R}$	$\mathbf{D}[\mathrm{C}_2, \mathbb{R}\times\mathbb{R}]$	1	0, 0, 0, 0, 1	2, 18, 200	2, 6, 20
2	4	F_{ac}	C_4	$\mathbb{R}$	$\mathbf{D}[\mathrm{C}_4]$	3	0, 0, 2, 0, 1	1, 9, 100	1, 3, 10
2	4	$F_{a,b}$	D_2	$\mathbb{R}\times\mathbb{R}$	$\mathbf{D}[\mathrm{D}_2]$	1	0, 0, 0, 0, 3	2, 12, 110	2, 5, 14
2	4	$F_{ab,c}$	D_2	*	*	3	0, 0, 0, 0, 1	1, 9, 100	1, 4, 10
2	8	$F_{a,b,c}$	D_4	*	*	5	0, 0, 2, 0, 3	1, 6, 55	1, 3, 7
4	1	$G_{1,3}$	C_1	$\mathbb{R}\times\mathbb{C}$	$\mathbf{C}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	3, 20, 175	2, 6, 20
4	2	$N(G_{1,3})$	C_2	$\mathbb{R}\times\mathbb{R}$	$\mathbf{C}[\mathrm{C}_2]$	0	0, 0, 0, 0, 1	2, 11, 90	2, 5, 14
6	1	$G_{3,3}$	C_1	$\mathbb{R}\times\mathbb{R}$	$\mathbf{B}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	2, 10, 70	2, 5, 14
6	2	$N(G_{3,3})$	C_2	$\mathbb{R}$	$\mathbf{B}[\mathrm{C}_2]$	1	0, 0, 0, 0, 0	1, 5, 35	1, 3, 7
10	1	$\mathrm{USp}(4)$	C_1	$\mathbb{R}$	$\mathbf{A}[\mathrm{C}_1]$	0	0, 0, 0, 0, 0	1, 3, 14	1, 2, 4

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 9. Moments of a_1 for Sato–Tate groups in genus 2.

G	$M_n = \mathrm{E}[a_1^n]$	M_2	M_4	M_6	M_8	M_{10}	Type [KS]
C_1	$2^n b_n$	8	96	1280	17920	258048	#27
C_2	$1/2(2^n + 0^n)b_n$	4	48	640	8960	129024	#13
C_3	$1/3(2^n + 2)b_n$	4	36	440	6020	86184	#28
C_4	$1/4(2^n + 0^n + 2 \cdot 2^{n/2})b_n$	4	36	400	5040	68544	#29
C_6	$1/6(2^n + 0^n + 2 + 2 \cdot 3^{n/2})b_n$	4	36	400	4900	63504	#30
D_2	$1/4(2^n + 3 \cdot 0^n)b_n$	2	24	320	4480	64512	#21
D_3	$1/6(2^n + 3 \cdot 0^n + 2)b_n$	2	18	220	3010	43092	#12
D_4	$1/8(2^n + 5 \cdot 0^n + 2 \cdot 2^{n/2})b_n$	2	18	200	2520	34272	#17
D_6	$1/12(2^n + 7 \cdot 0^n + 2 + 2 \cdot 3^{n/2})b_n$	2	18	200	2450	31752	#15
T	$1/12(2^n + 3 \cdot 0^n + 8)b_n$	2	12	120	1540	21672	#31
O	$1/24(2^n + 9 \cdot 0^n + 8 + 6 \cdot 2^{n/2})b_n$	2	12	100	1050	12852	#32
$J(C_1)$	$1/2(2^n + 0^n)b_n$	4	48	640	8960	129024	#13
$J(C_2)$	$1/4(2^n + 3 \cdot 0^n)b_n$	2	24	320	4480	64512	#21
$J(C_3)$	$1/6(2^n + 3 \cdot 0^n + 2)b_n$	2	18	220	3010	43092	#12
$J(C_4)$	$1/8(2^n + 5 \cdot 0^n + 2 \cdot 2^{n/2})b_n$	2	18	200	2520	34272	#17
$J(C_6)$	$1/12(2^n + 7 \cdot 0^n + 2 + 2 \cdot 3^{n/2})b_n$	2	18	200	2450	31752	#15
$J(D_2)$	$1/8(2^n + 7 \cdot 0^n)b_n$	1	12	160	2240	32256	#23
$J(D_3)$	$1/12(2^n + 9 \cdot 0^n + 2)b_n$	1	9	110	1505	21546	#20
$J(D_4)$	$1/16(2^n + 13 \cdot 0^n + 2 \cdot 2^{n/2})b_n$	1	9	100	1260	17136	#22
$J(D_6)$	$1/24(2^n + 19 \cdot 0^n + 2 + 2 \cdot 3^{n/2})b_n$	1	9	100	1225	15876	#24
$J(T)$	$1/24(2^n + 15 \cdot 0^n + 8)b_n$	1	6	60	770	10836	#25
$J(O)$	$1/48(2^n + 33 \cdot 0^n + 8 + 6 \cdot 2^{n/2})b_n$	1	6	50	525	6426	#26
$C_{2,1}$	$1/2(2^n + 0^n)b_n$	4	48	640	8960	129024	#13
$C_{4,1}$	$1/4(2^n + 3 \cdot 0^n)b_n$	2	24	320	4480	64512	#21
$C_{6,1}$	$1/6(2^n + 3 \cdot 0^n + 2)b_n$	2	18	220	3010	43092	#12
$D_{2,1}$	$1/4(2^n + 3 \cdot 0^n)b_n$	2	24	320	4480	64512	#21
$D_{4,1}$	$1/8(2^n + 7 \cdot 0^n)b_n$	1	12	160	2240	32256	#23
$D_{6,1}$	$1/12(2^n + 9 \cdot 0^n + 2)b_n$	1	9	110	1505	21546	#20
$D_{3,2}$	$1/6(2^n + 3 \cdot 0^n + 2)b_n$	2	18	220	3010	43092	#12
$D_{4,2}$	$1/8(2^n + 5 \cdot 0^n + 2 \cdot 2^{n/2})b_n$	2	18	200	2520	34272	#17
$D_{6,2}$	$1/12(2^n + 7 \cdot 0^n + 2 + 2 \cdot 3^{n/2})b_n$	2	18	200	2450	31752	#15
O_1	$1/24(2^n + 15 \cdot 0^n + 8)b_n$	1	6	60	770	10836	#25
E_1	$2^n c_n$	4	32	320	3584	43008	#5
E_2	$1/2(2^n + 0^n)c_n$	2	16	160	1792	21504	#11
E_3	$1/3(2^n + 2)c_n$	2	12	110	1204	14364	#4
E_4	$1/4(2^n + 0^n + 2 \cdot 2^{n/2})c_n$	2	12	100	1008	11424	#7
E_6	$1/6(2^n + 0^n + 2 + 2 \cdot 3^{n/2})c_n$	2	12	100	980	10584	#6
$J(E_1)$	$1/2(2^n + 0^n)c_n$	2	16	160	1792	21504	#11
$J(E_2)$	$1/4(2^n + 3 \cdot 0^n)c_n$	1	8	80	896	10752	#18
$J(E_3)$	$1/6(2^n + 3 \cdot 0^n + 2)c_n$	1	6	55	602	7182	#10
$J(E_4)$	$1/8(2^n + 5 \cdot 0^n + 2 \cdot 2^{n/2})c_n$	1	6	50	504	5712	#16
$J(E_6)$	$1/12(2^n + 7 \cdot 0^n + 2 + 2 \cdot 3^{n/2})c_n$	1	6	50	490	5292	#14
F	b_n^2	4	36	400	4900	63504	#33
F_a	$1/2(b_n + b_n^2)$	3	21	210	2485	31878	#34
F_c	$1/2(b_n^2 + 0^n)$	2	18	200	2450	31752	#35
F_{ab}	$1/2(b_n^2 + 0^n)$	2	18	200	2450	31752	#35
F_{ac}	$1/4(b_n^2 + 3 \cdot 0^n)$	1	9	100	1225	15876	#19
$F_{a,b}$	$1/4(b_n^2 + 2b_n + 0^n)$	2	12	110	1260	16002	#8
$F_{ab,c}$	$1/4(b_n^2 + 3 \cdot 0^n)$	1	9	100	1225	15876	#19
$F_{a,b,c}$	$1/8(b_n^2 + 2b_n + 5 \cdot 0^n)$	1	6	55	630	8001	#37
$G_{1,3}$	$c_n b_{n+2}/2$	3	20	175	1764	19404	#36
$N(G_{1,3})$	$1/2(c_n b_{n+2}/2 + c_n)$	2	11	90	889	9723	#3
$G_{3,3}$	$c_n c_{n+2}$	2	10	70	588	5544	#2
$N(G_{3,3})$	$1/2(c_n c_{n+2} + 0^n)$	1	5	35	294	2772	#9
USp(4)	$c_n c_{n+4} - c_{n+2}^2$	1	3	14	84	594	#1

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 10. Moments of a_2 for Sato–Tate groups in genus 2.

G	$M_n = \mathrm{E}[a_2^n]$	M_1	M_2	M_3	M_4	M_5
C_1	$b_{4,n}$	4	18	88	454	2424
C_2	$1/2(b_{4,n} + b_n)$	2	10	44	230	1212
C_3	$1/3(b_{4,n} + 2b_{1,n})$	2	8	34	164	842
C_4	$1/4(b_{4,n} + b_n + 2b_{2,n})$	2	8	32	150	732
C_6	$1/6(b_{4,n} + b_n + 2b_{1,n} + 2b_{3,n})$	2	8	32	148	712
D_2	$1/4(b_{4,n} + 3b_n)$	1	6	22	118	606
D_3	$1/6(b_{4,n} + 3b_n + 2b_{1,n})$	1	5	17	85	421
D_4	$1/8(b_{4,n} + 5b_n + 2b_{2,n})$	1	5	16	78	366
D_6	$1/12(b_{4,n} + 7b_n + 2b_{1,n} + 2b_{3,n})$	1	5	16	77	356
T	$1/12(b_{4,n} + 3b_n + 8b_{1,n})$	1	4	12	52	236
O	$1/24(b_{4,n} + 9b_n + 8b_{1,n} + 6b_{2,n})$	1	4	11	45	181
$J(C_1)$	$1/2(b_{4,n} + (-2)^n)$	1	11	40	235	1196
$J(C_2)$	$1/4(b_{4,n} + b_n + 2^n + (-2)^n)$	1	7	22	123	606
$J(C_3)$	$1/6(b_{4,n} + 2(b_{1,n} + 1) + (-2)^n)$	1	5	16	85	416
$J(C_4)$	$1/8(b_{4,n} + b_n + 2(b_{2,n} + 0^n) + 2^n + (-2)^n)$	1	5	16	79	366
$J(C_6)$	$1/12(b_{4,n} + b_n + 2(b_{1,n} + b_{3,n} + 1 + (-1)^n) + 2^n + (-2)^n)$	1	5	16	77	356
$J(D_2)$	$1/8(b_{4,n} + 3(b_n + 2^n) + (-2)^n)$	1	5	13	67	311
$J(D_3)$	$1/12(b_{4,n} + 3(b_n + 2^n) + 2(b_{1,n} + 1) + (-2)^n)$	1	4	10	48	216
$J(D_4)$	$1/16(b_{4,n} + 5(b_n + 2^n) + 2(b_{2,n} + 0^n) + (-2)^n)$	1	4	10	45	191
$J(D_6)$	$1/24(b_{4,n} + 2(b_{1,n} + b_{3,n} + 1 + (-1)^n) + 7(b_n + 2^n) + (-2)^n)$	1	4	10	44	186
$J(T)$	$1/24(b_{4,n} + 3(b_n + 2^n) + 8(b_{1,n} + 1) + (-2)^n)$	1	3	7	29	121
$J(O)$	$1/48(b_{4,n} + 9(b_n + 2^n) + 8(b_{1,n} + 1) + 6(b_{2,n} + 0^n) + (-2)^n)$	1	3	7	26	96
$C_{2,1}$	$1/2(b_{4,n} + 2^n)$	3	11	48	235	1228
$C_{4,1}$	$1/4(b_{4,n} + b_n + 2 \cdot 0^n)$	1	5	22	115	606
$C_{6,1}$	$1/6(b_{4,n} + 2(b_{1,n} + (-1)^n) + 2^n)$	1	5	18	85	426
$D_{2,1}$	$1/4(b_{4,n} + b_n + 2^{n+1})$	2	7	26	123	622
$D_{4,1}$	$1/8(b_{4,n} + 3b_n + 2(2^n + 0^n))$	1	4	13	63	311
$D_{6,1}$	$1/12(b_{4,n} + 3b_n + 2(b_{1,n} + (-1)^n) + 2^{n+2})$	1	4	11	48	221
$D_{3,2}$	$1/6(b_{4,n} + 2b_{1,n} + 3 \cdot 2^n)$	2	6	21	90	437
$D_{4,2}$	$1/8(b_{4,n} + b_n + 2b_{2,n} + 2^{n+2})$	2	6	20	83	382
$D_{6,2}$	$1/12(b_{4,n} + b_n + 2(b_{1,n} + b_{3,n}) + 6 \cdot 2^n)$	2	6	20	82	372
O_1	$1/24(b_{4,n} + 3b_n + 8b_{1,n} + 6(2^n + 0^n))$	1	3	8	30	126
E_1	$d_{4,n}$	3	10	37	150	654
E_2	$1/2(d_{4,n} + d_n)$	1	6	17	78	322
E_3	$1/3(d_{4,n} + 2d_{1,n})$	1	4	13	52	222
E_4	$1/4(d_{4,n} + d_n + 2d_{2,n})$	1	4	11	46	182
E_6	$1/6(d_{4,n} + d_n + 2(d_{1,n} + d_{3,n}))$	1	4	11	44	172
$J(E_1)$	$1/2(d_{4,n} + (-1)^n d_n)$	2	6	20	78	332
$J(E_2)$	$1/4(d_{4,n} + d_n + 2(-1)^n d_n)$	1	4	10	42	166
$J(E_3)$	$1/6(d_{4,n} + 2d_{1,n} + 3(-1)^n d_n)$	1	3	8	29	116
$J(E_4)$	$1/8(d_{4,n} + d_n + 2d_{2,n} + 4(-1)^n d_n)$	1	3	7	26	96
$J(E_6)$	$1/12(d_{4,n} + d_n + 2(d_{1,n} + d_{3,n}) + 6(-1)^n d_n)$	1	3	7	25	91
F	$\hat{b}_n$	2	8	32	148	712
F_a	$1/2(\hat{b}_n + 2^n)$	2	6	20	82	372
F_c	$1/2(\hat{b}_n + b_n)$	1	5	16	77	356
F_{ab}	$1/2(\hat{b}_n + 2^n)$	2	6	20	82	372
F_{ac}	$1/4(\hat{b}_n + 2 \cdot 0^n + 2^n)$	1	3	10	41	186
$F_{a,b}$	$1/4(\hat{b}_n + 3 \cdot 2^n)$	2	5	14	49	202
$F_{ab,c}$	$1/4(\hat{b}_n + 2b_n + 2^n)$	1	4	10	44	186
$F_{a,b,c}$	$1/8(\hat{b}_n + 2(b_n + 0^n) + 3 \cdot 2^n)$	1	3	7	26	101
$G_{1,3}$	$\sum_k \binom{n}{k} 2^{n-k} b_k c_k$	2	6	20	76	312
$N(G_{1,3})$	$1/2(\sum_k \binom{n}{k} 2^{n-k} b_k c_k + 2^n)$	2	5	14	46	172
$G_{3,3}$	$\hat{c}_n$	2	5	14	44	152
$N(G_{3,3})$	$1/2(\hat{c}_n + c_n)$	1	3	7	23	76
USp(4)	$\sum_k \binom{n}{k} 2^{n-k} (c_k c_{k+2} - c_{k+1}^2)$	1	2	4	10	27

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 11. Genus 2 curves realizing Sato–Tate groups.

G	Curve $y^2 = f(x)$	k	K
C_1	$x^6 + 1$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(\sqrt{-3})$
C_2	$x^5 - x$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(i, \sqrt{2})$
C_3	$x^6 + 4$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(\sqrt{-3}, \sqrt[3]{2})$
C_4	$x^6 + x^5 - 5x^4 - 5x^2 - x + 1$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(\sqrt{-2}, a)$; $a^4 + 17a^2 + 68 = 0$
C_6	$x^6 + 2$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{2})$
D_2	$x^5 + 9x$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(i, \sqrt{2}, \sqrt{3})$
D_3	$x^6 + 10x^3 - 2$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{-2})$
D_4	$x^5 + 3x$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(i, \sqrt{2}, \sqrt[4]{3})$
D_6	$x^6 + 3x^5 + 10x^3 - 15x^2 + 15x - 6$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(i, \sqrt{2}, \sqrt{3}, a)$; $a^3 + 3a - 2 = 0$
T	$x^6 + 6x^5 - 20x^4 + 20x^3 - 20x^2 - 8x + 8$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(\sqrt{-2}, a, b)$; $a^3 - 7a + 7 = b^4 + 4b^2 + 8b + 8 = 0$
O	$x^6 - 5x^4 + 10x^3 - 5x^2 + 2x - 1$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(\sqrt{-2}, \sqrt{-11}, a, b)$; $a^3 - 4a + 4 = b^4 + 22b + 22 = 0$
$J(C_1)$	$x^5 - x$	$\mathbb{Q}(i)$	$\mathbb{Q}(i, \sqrt{2})$
$J(C_2)$	$x^5 - x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2})$
$J(C_3)$	$x^6 + 10x^3 - 2$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{-2})$
$J(C_4)$	$x^6 + x^5 - 5x^4 - 5x^2 - x + 1$	$\mathbb{Q}$	see entry for C_4
$J(C_6)$	$x^6 - 15x^4 - 20x^3 + 6x + 1$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{3}, a)$; $a^3 + 3a^2 - 1 = 0$
$J(D_2)$	$x^5 + 9x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2}, \sqrt{3})$
$J(D_3)$	$x^6 + 10x^3 - 2$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{-2})$
$J(D_4)$	$x^5 + 3x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2}, \sqrt[4]{3})$
$J(D_6)$	$x^6 + 3x^5 + 10x^3 - 15x^2 + 15x - 6$	$\mathbb{Q}$	see entry for D_6
$J(T)$	$x^6 + 6x^5 - 20x^4 + 20x^3 - 20x^2 - 8x + 8$	$\mathbb{Q}$	see entry for T
$J(O)$	$x^6 - 5x^4 + 10x^3 - 5x^2 + 2x - 1$	$\mathbb{Q}$	see entry for O
$C_{2,1}$	$x^6 + 1$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3})$
$C_{4,1}$	$x^5 + 2x$	$\mathbb{Q}(i)$	$\mathbb{Q}(i, \sqrt[4]{2})$
$C_{6,1}$	$x^6 + 6x^5 - 30x^4 + 20x^3 + 15x^2 - 12x + 1$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, a)$; $a^3 - 3a + 1 = 0$
$D_{2,1}$	$x^5 + x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2})$
$D_{4,1}$	$x^5 + 2x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt[4]{2})$
$D_{6,1}$	$x^6 + 6x^5 - 30x^4 - 40x^3 + 60x^2 + 24x - 8$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-2}, \sqrt{-3}, a)$; $a^3 - 9a + 6 = 0$
$D_{3,2}$	$x^6 + 4$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, \sqrt[3]{2})$
$D_{4,2}$	$x^6 + x^5 + 10x^3 + 5x^2 + x - 2$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-2}, a)$; $a^4 - 14a^2 + 28a - 14 = 0$
$D_{6,2}$	$x^6 + 2$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{2})$
O_1	$x^6 + 7x^5 + 10x^4 + 10x^3 + 15x^2 + 17x + 4$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-2}, a, b)$; $a^3 + 5a + 10 = b^4 + 4b^2 + 8b + 2 = 0$
F	$x^6 + 3x^4 + x^2 - 1$	$\mathbb{Q}(i, \sqrt{2})$	$\mathbb{Q}(i, \sqrt{2})$
F_a	$x^6 + 3x^4 + x^2 - 1$	$\mathbb{Q}(i)$	$\mathbb{Q}(i, \sqrt{2})$
F_{ab}	$x^6 + 3x^4 + x^2 - 1$	$\mathbb{Q}(\sqrt{2})$	$\mathbb{Q}(i, \sqrt{2})$
F_{ac}	$x^5 + 1$	$\mathbb{Q}$	$\mathbb{Q}(a)$; $a^4 + 5a^2 + 5 = 0$
$F_{a,b}$	$x^6 + 3x^4 + x^2 - 1$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2})$
E_1	$x^6 + x^4 + x^2 + 1$	$\mathbb{Q}$	$\mathbb{Q}$
E_2	$x^6 + x^5 + 3x^4 + 3x^2 - x + 1$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{2})$
E_3	$x^5 + x^4 - 3x^3 - 4x^2 - x$	$\mathbb{Q}$	$\mathbb{Q}(a)$; $a^3 - 3a + 1 = 0$
E_4	$x^5 + x^4 + x^2 - x$	$\mathbb{Q}$	$\mathbb{Q}(a)$; $a^4 - 5a^2 + 5 = 0$
E_6	$x^5 + 2x^4 - x^3 - 3x^2 - x$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{7}, a)$; $a^3 - 7a - 7 = 0$
$J(E_1)$	$x^5 + x^3 + x$	$\mathbb{Q}$	$\mathbb{Q}(i)$
$J(E_2)$	$x^5 + x^3 - x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt{2})$
$J(E_3)$	$x^6 + x^3 + 4$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, \sqrt[3]{2})$
$J(E_4)$	$x^5 + x^3 + 2x$	$\mathbb{Q}$	$\mathbb{Q}(i, \sqrt[4]{2})$
$J(E_6)$	$x^6 + x^3 - 2$	$\mathbb{Q}$	$\mathbb{Q}(\sqrt{-3}, \sqrt[6]{-2})$
$G_{1,3}$	$x^6 + 3x^4 - 2$	$\mathbb{Q}(i)$	$\mathbb{Q}(i)$
$N(G_{1,3})$	$x^6 + 3x^4 - 2$	$\mathbb{Q}$	$\mathbb{Q}(i)$
$G_{3,3}$	$x^6 + x^2 + 1$	$\mathbb{Q}$	$\mathbb{Q}$
$N(G_{3,3})$	$x^6 + x^5 + x - 1$	$\mathbb{Q}$	$\mathbb{Q}(i)$
USp(4)	$x^5 - x + 1$	$\mathbb{Q}$	$\mathbb{Q}$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 12. Some automorphisms of the curves in Table 11.

G	α	γ	M
$C_1, C_{2,1}$	$(-x, y)$	$\left(\frac{1}{x}, \frac{1}{x^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$C_2, J(C_1), J(C_2)$	$\left(\frac{i}{x}, \frac{\zeta_8^3}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}(\sqrt{-2})$
$C_3, D_{3,2}$	$(-x, y)$	$\left(\frac{4^{1/3}}{x}, \frac{2}{x^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$C_4, J(C_4)$	$\left(\frac{-(2a^2+13)x+1}{x+2a^2+13}, \frac{Q_{\alpha,C_4}}{(x+2a^2+13)^3}y\right)$	$\left(\frac{-x-1}{x-1}, \frac{-2\sqrt{-2}}{(x-1)^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
$C_6, D_{6,2}$	$(-x, y)$	$\left(\frac{2^{1/3}}{x}, \frac{2^{1/2}}{x^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$D_2, J(D_2)$	$\left(\frac{3}{x}, \frac{3^{3/2}}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}(\sqrt{-2})$
$D_3, J(D_3), J(C_3)$	$\left(\frac{-2^{1/3}}{x}, \frac{\sqrt{-2}}{x^3}y\right)$	$\left(\frac{(1-\sqrt{-3})2^{1/3}}{2x}, \frac{\sqrt{-2}}{x^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
$D_4, J(D_4)$	$\left(\frac{\sqrt{3}}{x}, \frac{3^{3/4}}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}(\sqrt{-2})$
$D_6, J(D_6)$	$\left(\frac{x+1}{x-1}, \frac{2\sqrt{2}}{(x-1)^3}y\right)$	$\left(\frac{P_{\gamma,D_6}}{2x+(a^2+3)}, \frac{Q_{\gamma,D_6}}{(2x+(a^2+3))^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$T, J(T)$	$\left(\frac{P_{\alpha,T}}{R_{\alpha,T}}, \frac{Q_{\alpha,T}}{R_{\alpha,T}^3}y\right)$	$\left(\frac{P_{\gamma,T}}{x+1-a}, \frac{Q_{\gamma,T}}{(x+1-a)^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
$O, J(O)$	$\left(\frac{P_{\alpha,O}}{R_{\alpha,O}}, \frac{Q_{\alpha,O}}{R_{\alpha,O}^3}y\right)$	$\left(\frac{ax+a^2-2}{2x-a}, \frac{Q_{\gamma,O}}{(2x-a)^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
$J(C_6)$	$\left(\frac{-x-2}{2x+1}, \frac{3\sqrt{-3}}{(2x+1)^3}y\right)$	$\left(\frac{-(a+1)x-a}{x+a+1}, \frac{-3ia^2-3ia}{(x+a+1)^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$C_{4,1}, D_{4,1}$	$\left(\frac{\sqrt{2}}{x}, \frac{2^{3/4}}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}(\sqrt{-2})$
$C_{6,1}$	$\left(\frac{(1-a)x+a}{x+(a-1)}, \frac{-3a^2+3a}{(x+(a-1))^3}y\right)$	$\left(\frac{x-1}{x}, \frac{-1}{x^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$D_{2,1}$	$\left(\frac{1}{x}, \frac{1}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}(\sqrt{-2})$
$D_{6,1}$	$\left(\frac{-2}{x}, \frac{-2\sqrt{-2}}{x^3}y\right)$	$\left(\frac{(a-1)x+2}{x+1-a}, \frac{Q_{\gamma,D_{6,1}}}{(x+1-a)^3}y\right)$	$\mathbb{Q}(\sqrt{-3})$
$D_{4,2}$	$\left(\frac{P_{\alpha,D_{4,2}}}{R_{\alpha,D_{4,2}}}, \frac{Q_{\alpha,D_{4,2}}}{R_{\alpha,D_{4,2}}^3}y\right)$	$\left(\frac{P_{\gamma,D_{4,2}}}{x-\sqrt{-2}+1}, \frac{-8}{(x-\sqrt{-2}+1)^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
O_1	$\left(\frac{P_{\alpha,O_1}}{R_{\alpha,O_1}}, \frac{Q_{\alpha,O_1}}{R_{\alpha,O_1}^3}y\right)$	$\left(\frac{P_{\gamma,O_1}}{R_{\gamma,O_1}}, \frac{Q_{\gamma,O_1}}{R_{\gamma,O_1}^3}y\right)$	$\mathbb{Q}(\sqrt{-2})$
E_1	$(-x, y)$	$\left(\frac{-1}{x}, \frac{1}{x^3}y\right)$	$\mathbb{Q}$
E_2	$\left(\frac{x+1}{x-1}, \frac{-2\sqrt{2}}{(x-1)^3}y\right)$	$\left(\frac{-1}{x}, \frac{1}{x^3}y\right)$	$\mathbb{Q}$
E_3	$\left(\frac{-ax-a+1}{x+a}, \frac{3a^2-3a}{(x+a)^3}y\right)$	$\left(\frac{-x-1}{x}, \frac{1}{x^3}y\right)$	$\mathbb{Q}$
E_4	$\left(\frac{(a^2-3)x+1}{x-a^2+3}, \frac{-3a^3+10a}{(x-a^2+3)^3}y\right)$	$\left(\frac{-1}{x}, \frac{1}{x^3}y\right)$	$\mathbb{Q}$
E_6	$\left(\frac{P_{\alpha,E_6}}{R_{\alpha,E_6}}, \frac{Q_{\alpha,E_6}}{R_{\alpha,E_6}^3}y\right)$	$\left(\frac{-1}{x+1}, \frac{1}{(x+1)^3}y\right)$	$\mathbb{Q}$
$J(E_1)$	$\left(\frac{1}{x}, \frac{1}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}$
$J(E_2)$	$\left(\frac{-i}{x}, \frac{i-1}{\sqrt{2}x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}$
$J(E_3)$	$\left(\frac{4^{1/3}}{x}, \frac{2}{x^3}y\right)$	$(\zeta_3 x, y)$	$\mathbb{Q}$
$J(E_4)$	$\left(\frac{\sqrt{2}}{x}, \frac{2^{3/4}}{x^3}y\right)$	$(-x, iy)$	$\mathbb{Q}$
$J(E_6)$	$\left(\frac{(-2)^{1/3}}{x}, \frac{\sqrt{2}}{x^3}\right)$	$(\zeta_3 x, y)$	$\mathbb{Q}$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

TABLE 13. Polynomials describing the automorphisms of Table 12.

$$
\begin{aligned}
Q_{\alpha,C_4} &= \sqrt{-2}(29a^3+187a)\\
P_{\gamma,D_6} &= -(a^2+3)x+2(a^2+2)\\
Q_{\gamma,D_6} &= -21ia^2-6ia-83i\\
P_{\alpha,T} &= \left(\left(\tfrac{-a^2}{7}-\tfrac{a}{6}+\tfrac{2}{3}\right)b^3+\left(\tfrac{3a^2}{14}+\tfrac{a}{3}-\tfrac{7}{6}\right)b^2+\left(\tfrac{-10}{21}a^2-\tfrac{2}{3}a+\tfrac{7}{3}\right)b+\left(\tfrac{-2}{21}a^2+\tfrac{2}{3}\right)\right)x\\
&\quad+\left(\tfrac{1}{21}a^2-\tfrac{1}{3}\right)b^3+\left(\tfrac{-5}{21}a^2-\tfrac{1}{3}a+\tfrac{4}{3}\right)b^2+\left(\tfrac{-2}{7}a^2-\tfrac{2}{3}a+\tfrac{2}{3}\right)b+\left(\tfrac{10}{21}a^2-\tfrac{8}{3}\right)\\
Q_{\alpha,T} &= \sqrt{-2}\left(\left(\tfrac{10}{27}a^2+\tfrac{11}{27}a-\tfrac{49}{27}\right)b^3+\left(\tfrac{8}{27}a^2+\tfrac{22}{27}a-\tfrac{49}{27}\right)b^2\right.\\
&\quad\left.+\left(\tfrac{32}{27}a^2+\tfrac{46}{27}a-\tfrac{182}{27}\right)b+\left(\tfrac{76}{27}a^2+\tfrac{110}{27}a-\tfrac{392}{27}\right)\right)\\
R_{\alpha,T} &= x+\left(\tfrac{a^2}{7}+\tfrac{a}{6}-\tfrac{2}{3}\right)b^3+\left(\tfrac{-3}{14}a^2-\tfrac{a}{3}+\tfrac{7}{6}\right)b^2+\left(\tfrac{10}{21}a^2+\tfrac{2}{3}a-\tfrac{7}{3}\right)b+\left(\tfrac{2}{21}a^2-\tfrac{2}{3}\right)\\
P_{\gamma,T} &= (a-1)x+2a^2+2a-8\\
Q_{\gamma,T} &= \left(-a^2+\tfrac{3}{2}a\right)b^3+\left(\tfrac{a^2}{2}-\tfrac{a}{2}\right)b^2+(-5a^2+12a-7)b-5a^2+8a\\
P_{\alpha,O} &= \left(\left(\tfrac{19}{429}a^2+\tfrac{23}{429}a-\tfrac{58}{429}\right)b^3+\left(\tfrac{-5}{78}a^2-\tfrac{2}{39}a+\tfrac{3}{13}\right)b^2+\left(\tfrac{11}{78}a^2+\tfrac{7}{39}a-\tfrac{4}{13}\right)b\right.\\
&\quad\left.+\left(\tfrac{31}{78}a^2+\tfrac{5}{13}a-\tfrac{35}{39}\right)\right)x+\left(\tfrac{-23}{429}a^2-\tfrac{6}{143}a+\tfrac{76}{429}\right)b^3+\left(\tfrac{2}{39}a^2+\tfrac{1}{39}a-\tfrac{10}{39}\right)b^2\\
&\quad+\left(\tfrac{-7}{39}a^2-\tfrac{10}{39}a+\tfrac{22}{39}\right)b+\left(\tfrac{-23}{26}a^2-\tfrac{9}{13}a+\tfrac{101}{39}\right)\\
Q_{\alpha,O} &= \sqrt{-2}\left(\left(\tfrac{a^2}{39}+\tfrac{20}{351}a-\tfrac{22}{351}\right)b^3+\left(\tfrac{-11}{702}a^2-\tfrac{44}{351}a-\tfrac{22}{117}\right)b^2\right.\\
&\quad\left.+\left(\tfrac{107}{351}a^2+\tfrac{232}{351}a-\tfrac{146}{351}\right)b+\left(\tfrac{11}{26}a^2+\tfrac{110}{117}a-\tfrac{121}{117}\right)\right)\\
R_{\alpha,O} &= x+\left(\tfrac{-19}{429}a^2-\tfrac{23}{429}a+\tfrac{58}{429}\right)b^3+\left(\tfrac{5}{78}a^2+\tfrac{2}{39}a-\tfrac{3}{13}\right)b^2\\
&\quad+\left(\tfrac{-11}{78}a^2-\tfrac{7}{39}a+\tfrac{4}{13}\right)b+\left(\tfrac{-31}{78}a^2-\tfrac{5}{13}a+\tfrac{35}{39}\right)\\
Q_{\gamma,O} &= 8\sqrt{-11}\left(\left(\tfrac{1}{286}a^2+\tfrac{3}{143}a-\tfrac{7}{143}\right)b^3+\left(\tfrac{-7}{572}a^2+\tfrac{5}{286}a+\tfrac{5}{143}\right)b^2\right.\\
&\quad\left.+\left(\tfrac{1}{13}a^2-\tfrac{1}{26}a-\tfrac{1}{13}\right)b+\left(\tfrac{3}{52}a^2+\tfrac{9}{26}a-\tfrac{21}{26}\right)\right)\\
Q_{\gamma,D_{6,1}} &= \sqrt{-3}(3a^2-6a+1)\\
P_{\alpha,D_{4,2}} &= (a^3+a^2-14a+17)x+(2a^3+2a^2-28a+29)\\
Q_{\alpha,D_{4,2}} &= -160a^3-168a^2+1904a-2184\\
R_{\alpha,D_{4,2}} &= 5x-(a^3+a^2-14a+17)\\
P_{\gamma,D_{4,2}} &= -(\sqrt{-2}+1)x+1\\
P_{\alpha,O_1} &= \left(\left(\tfrac{-3}{580}a^2-\tfrac{13}{116}a-\tfrac{7}{58}\right)b^3+\left(\tfrac{-1}{1160}a^2+\tfrac{15}{232}a+\tfrac{17}{116}\right)b^2+\left(\tfrac{-17}{232}a^2-\tfrac{117}{232}a-\tfrac{63}{116}\right)b\right.\\
&\quad\left.+\left(\tfrac{13}{290}a^2-\tfrac{21}{58}a-\tfrac{18}{29}\right)\right)x+\left(\tfrac{-7}{580}a^2-\tfrac{11}{116}a+\tfrac{3}{58}\right)b^3+\left(\tfrac{23}{580}a^2+\tfrac{3}{116}a+\tfrac{15}{58}\right)b^2\\
&\quad+\left(\tfrac{-15}{116}a^2-\tfrac{35}{116}a-\tfrac{1}{58}\right)b+\left(\tfrac{109}{580}a^2-\tfrac{11}{116}a+\tfrac{61}{58}\right)\\
Q_{\alpha,O_1} &= \left(\tfrac{-2277}{24389}a^2+\tfrac{7186}{24389}a-\tfrac{495}{24389}\right)b^3+\left(\tfrac{1287}{48778}a^2-\tfrac{34813}{48778}a-\tfrac{13115}{24389}\right)b^2\\
&\quad+\left(\tfrac{-26733}{48778}a^2+\tfrac{83435}{48778}a+\tfrac{26255}{24389}\right)b+\left(\tfrac{-12375}{24389}a^2+\tfrac{8303}{24389}a-\tfrac{29200}{24389}\right)\\
R_{\alpha,O_1} &= x+\left(\tfrac{3}{580}a^2+\tfrac{13}{116}a+\tfrac{7}{58}\right)b^3+\left(\tfrac{1}{1160}a^2-\tfrac{15}{232}a-\tfrac{17}{116}\right)b^2\\
&\quad+\left(\tfrac{17}{232}a^2+\tfrac{117}{232}a+\tfrac{63}{116}\right)b+\left(\tfrac{-13}{290}a^2+\tfrac{21}{58}a+\tfrac{18}{29}\right)\\
P_{\gamma,O_1} &= (-a^2+a-8)x-a^2-a-6\\
Q_{\gamma,O_1} &= 4\sqrt{-2}((-2a^2+4a-10)b^3+(-3a^2+4a-25)b^2\\
&\quad+(-2a^2+10a)b-18a^2+32a-110)\\
R_{\gamma,O_1} &= 4x+a^2-a+8\\
P_{\alpha,E_6} &= (a^2-a-5)x+a^2-a-4\\
Q_{\alpha,E_6} &= \sqrt{7}(-a^2+2a+6)\\
R_{\alpha,E_6} &= x-a^2+a+5
\end{aligned}
$$

https://doi.org/10.1112/S0010437X12000279 Published online by Cambridge University Press

Table 11 lists the example curves used in § 4.8 to prove that there do exist 52 distinct Galois types arising from abelian surfaces over general number fields, of which 34 do occur over $\mathbb{Q}$. For each curve, we indicate the field of definition k and the minimal extension K/k over which all endomorphisms of its Jacobian are defined. It is proved in § 4.8 that the field K and the Galois type agree with the claimed values; in most cases, the proof makes use of certain noncommuting automorphisms α and γ of the curve. These automorphisms are listed in Table 12, together with the value of the intermediate field M used in the alternate description of the Galois type in § 4. To make things more readable, some rather complicated polynomials appearing in the definitions of the automorphisms have been moved to Table 13. Note that in each formula in Tables 12 and 13, the symbols a and b represent elements of K as presented in the corresponding line of Table 11; consequently, the meaning of these symbols varies from line to line.

Acknowledgements

Thanks to Jean-Pierre Serre for numerous helpful discussions about the general Sato–Tate conjecture, for providing preliminary drafts of [Ser12], and for triggering the collaboration among the four authors by alerting the first and third authors to the existence of [KS08, KS09]. Thanks to Joan-Carles Lario for his help at the first stage of the project. Thanks to Grzegorz Banaszak for helpful discussions with Kedlaya about the algebraic Sato–Tate group and the Mumford–Tate group, leading to the paper [BK11]. Thanks to Kevin Buzzard and Marco Streng for helpful conversations with Rotger about the formulation of Conjecture 1.1 (respectively, on complex multiplication). Thanks to Amanda Clemm for the code used to generate Figure 1 in Sage [Ste11]. Thanks to Xavier Guitart, Joan-C. Lario, and Marc Masdeu for organizing the workshop 'Sato–Tate in higher dimension' at the Centro de Ciencias de Benasque Pedro Pascual (CCBPP) in July–August 2011, which provided the first opportunity for the four authors to collaborate in person. Thanks also to the CCBPP for its hospitality, and to Josep González for his lecture at the workshop providing examples of modular abelian surfaces leading to the correct definition of the Galois type.

References

BK11 G. Banaszak and K. S. Kedlaya, *An algebraic Sato–Tate group and Sato–Tate conjecture*, Preprint (2011), arXiv:1109.4449v1.

BGG11 T. Barnet-Lamb, D. Geraghty and T. Gee, *The Sato–Tate conjecture for Hilbert modular forms*, J. Amer. Math. Soc. **24** (2011), 411–469.

BGHT11 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** (2011), 29–98.

Bea10 A. Beauville, *Finite subgroups of* $\mathrm{PGL}_2(K)$, Contemporary Mathematics, vol. 522 (American Mathematical Society, Providence, RI, 2010).

Bog80 F. A. Bogomolov, *Sur l'algébricité des représentations ℓ-adiques*, C. R. Acad. Sci. Paris **290** (1980), 701–703.

CF00 T. Chinburg and E. Friedman, *The finite subgroups of maximal arithmetic Kleinian groups*, Ann. Inst. Fourier Grenoble **50** (2000), 1765–1798.

Del82 P. Deligne, *Hodge cycles on abelian varieties*, in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Mathematics, vol. 900 (Springer, Berlin, 1982), 9–100.

Fal83 G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), 349–366.

Fit10 F. Fité, *Artin representations attached to pairs of isogenous abelian varieties*, Preprint (2010), arXiv:1012.3390v1.

GHM08 S. D. Galbraith, M. Harrison and D. J. Mireles Morales, *Efficient hyperelliptic arithmetic using balanced representation for divisors*, in *Algorithmic number theory: 8th international symposium, ANTS-VIII (Banff, Canada, May 2008) proceedings*, Lecture Notes in Computer Science, vol. 5011 (Springer, Berlin, 2008), 342–356.

GS01 P. Gaudry and É. Schost, *On the invariants of the quotients of the Jacobian of a curve of genus 2*, in *Applied algebra, algebraic algorithms and error-correcting codes: 14th international symposium, AAECC-14 (Melbourne, Australia, November 2001) proceedings*, Lecture Notes in Computer Science, vol. 2227 (Springer, Berlin, 2001), 373–386.

Har09 M. Harris, *Potential automorphy of odd-dimensional symmetric powers of elliptic curves, and applications*, in *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Volume II*, Progress in Mathematics, vol. 270 (Birkhäuser, Boston, MA, 2009), 1–21.

KS99 N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, vol. 45 (American Mathematical Society, Providence, RI, 1999).

KS08 K. S. Kedlaya and A. V. Sutherland, *Computing L-series of hyperelliptic curves*, in *Algorithmic number theory: 8th international symposium, ANTS-VIII (Banff, Canada, May 2008) proceedings*, Lecture Notes in Computer Science, vol. 5011 (Springer, Berlin, 2008), 312–326.

KS09 K. S. Kedlaya and A. V. Sutherland, *Hyperelliptic curves, L-polynomials, and random matrices*, in *Arithmetic, geometry, cryptography, and coding theory: international conference, November 5–9, 2007, CIRM, Marseilles, France*, Contemporary Mathematics, vol. 487 (American Mathematical Society, Providence, RI, 2009), 119–162.

Mum69 D. Mumford, *A note of Shimura's paper "Discontinuous subgroups and abelian varieties"*, Math. Ann. **181** (1969), 345–351.

Mum70 D. Mumford, *Abelian varieties* (Oxford University Press, Oxford, for Tata Institute of Fundamental Research, Bombay, 1970).

OEIS The OEIS Foundation Inc., *On-Line Encyclopedia of Integer Sequences (OEIS)*, http://oeis.org.

Rib04 K. A. Ribet, *Abelian varieties over $\mathbb{Q}$ and modular forms*, in *Modular curves and abelian varieties*, Progress in Mathematics, vol. 224, eds J. Cremona, J.-C. Lario, J. Quer and K. Ribet (Birkhäuser, Basel, 2004), 241–261.

Ser68 J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves* (W.A. Benjamin, New York, 1968).

Ser72 J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), 259–331.

Se81 J.-P. Serre, *Lettres à Ken Ribet du 1/1/1981 et du 29/1/1981*, in *Œuvres. Collected papers. Volume IV: 1985–1998* (Springer, Berlin, 2000).

Ser92 J.-P. Serre, *Lie algebras and Lie groups*, Lecture Notes in Mathematics, vol. 1500 (Springer, Berlin, 1992).

Ser94 J.-P. Serre, *Propriétés conjecturales des groupes de Galois motiviques et des représentations l-adiques*, in *Motives (Seattle, WA, 1991)*, Proceedings of Symposia in Pure Mathematics, vol. 55 (American Mathematical Society, Providence, RI, 1994), 377–400.

Ser12 J.-P. Serre, *Lectures on $N_X(p)$* (CRC Press, Boca Raton, FL, 2012).

Shi63 G. Shimura, *On analytic families of polarized abelian varieties and automorphic functions*, Ann. of Math. (2) **78** (1963), 149–192.

Shi71 G. Shimura, *On the zeta-function of an abelian variety with complex multiplication*, Ann. of Math. (2) **94** (1971), 504–533.

Shi98 G. Shimura, *Abelian varieties with complex multiplication and modular forms* (Princeton University Press, Princeton, NJ, 1998).

ST61 G. Shimura and Y. Taniyama, *Complex multiplication of abelian varieties and its application to number theory*, Publications of the Mathematical Society of Japan, vol. 6 (Mathematical Society of Japan, Tokyo, 1961).

Sil92 A. Silverberg, *Fields of definition for homomorphisms of abelian varieties*, J. Pure Appl. Algebra **77** (1992), 253–262.

Sil94 J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 151 (Springer, New York, 1994).

Ste11 W. A. Stein *et al.*, *Sage mathematics software (version 4.7.1)*, The Sage Development Team (2011), http://www.sagemath.org.

Ste65 R. Steinberg, *Regular elements of semisimple algebraic groups*, Publ. Math. Inst. Hautes Études Sci. **25** (1965), 49–80.

Str10 M. Streng, *Complex multiplication of abelian surfaces*, PhD thesis, Universiteit Leiden (2010).

Sut07 A. V. Sutherland, *Order computations in generic groups*, PhD thesis, Massachusetts Institute of Technology (2007).

Sut11a A. V. Sutherland, `smalljac` software library, version 4.0 (2011).

Sut11b A. V. Sutherland, *Structure computation and discrete logarithms in finite abelian p-groups*, Math. Comp. **80** (2011), 477–500.

Wei64 A. Weil, *Remarks on the cohomology of groups*, Ann. of Math. (2) **80** (1964), 149–157.

Yos73 H. Yoshida, *On an analogue of the Sato conjecture*, Invent. Math. **19** (1973), 261–277.

Zar00 Yu. G. Zarhin, *Hyperelliptic Jacobians without complex multiplication*, Math. Res. Lett. **7** (2000), 123–132.

Francesc Fité francesc.fite@gmail.com
Department of Mathematics, Universität Bielefeld, Postfach 100131,
D-33501 Bielefeld, Germany

Kiran S. Kedlaya kedlaya@mit.edu, kedlaya@ucsd.edu
Department of Mathematics, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA

Current address: Department of Mathematics, University of California, San Diego,
9500 Gilman Drive #0112, La Jolla, CA 92093-0112, USA

Víctor Rotger victor.rotger@upc.edu
Departament Matemàtica Aplicada II, Universitat Politècnica de Catalunya,
Campus Nord, Edifici Omega, Despatx 413, Jordi Girona, 1-3,
08034 Barcelona, Spain

Andrew V. Sutherland drew@math.mit.edu
Department of Mathematics, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA