ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Number Theory 107 (2004) 8–24

JOURNAL OF Number Theory

http://www.elsevier.com/locate/jnt

Modular units and the surjectivity of a Galois representation

David E. Rohrlich

Department of Mathematics and Statistics, Boston University, Boston, MA 02215, USA

Received 1 November 2002; revised 19 January 2004

Communicated by J. Tate

Abstract

For a prime $p \geqslant 7$ the pth roots of certain modular units are shown to generate the second layer of the extension of function fields cut out by the universal Galois deformation of the representation on p-division points of a universal elliptic curve. It follows that certain Galois representations obtained by specializing the modular invariant to a rational number have large image.

The Galois representation at issue in this note arises from the interplay between two universal constructions: universal Galois deformations on the one hand and universal elliptic curves on the other. Fix a prime $p \geqslant 7$, let j be transcendental over $\mathbb{Q}$, and consider an elliptic curve E over $\mathbb{Q}(j)$ with invariant j. The natural action of $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\overline{\mathbb{Q}}(j))$ on the group $E[p]$ of p-division points of E affords a representation $\bar{\pi}_E : \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\overline{\mathbb{Q}}(j)) \to \mathrm{SL}(2, \mathbb{F}_p)$, and the universal deformation of $\bar{\pi}_E$ is an epimorphism $\pi_E : \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\overline{\mathbb{Q}}(j)) \to \mathrm{SL}(2, \mathbb{Z}_p[[X]])$. More precisely, π_E is universal for deformations of $\bar{\pi}_E$ which are trivial on the inertia subgroups of $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\overline{\mathbb{Q}}(j, E[p]))$ at all places of $\overline{\mathbb{Q}}(j, E[p])$ not lying over the place $j = \infty$ of $\overline{\mathbb{Q}}(j)$. In any case, let $\widetilde{\mathbb{Q}}$ be the cyclotomic extension of $\mathbb{Q}$ generated by all roots of unity of p-power order. Then π_E descends to an epimorphism

$$\rho_E : \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j)) \to \mathrm{SL}(2, \mathbb{Z}_p[[X]]),$$

E-mail address: rohrlich@math.bu.edu.

0022-314X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jnt.2004.03.005

and by restricting ρ_E to decomposition subgroups of $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j))$ we can associate a representation

$$\rho_A : \mathrm{Gal}(\overline{\mathbb{Q}}/\widetilde{\mathbb{Q}}) \to \mathrm{SL}(2, \mathbb{Z}_p[[X]])$$

to any elliptic curve A over $\mathbb{Q}$ with $j(A) \neq 0, 1728$. The recipe is as follows (cf. [5,6]): Given A, choose E to have good reduction at $j = j(A)$ and to specialize to A at this place. Also choose a place v of $\overline{\mathbb{Q}(j)}$ over $j = j(A)$. The restriction of ρ_E to the decomposition subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j))$ at v is trivial on the inertia subgroup and so may be viewed as a representation of the quotient group $\mathrm{Gal}(\overline{\mathbb{Q}}/\widetilde{\mathbb{Q}})$; this is ρ_A.

We are concerned with the possible surjectivity of ρ_A. It is easy to see that the question of surjectivity depends only on the j-invariant of A, and it is also known that there are infinitely many numbers $j_0 \in \mathbb{Q}\backslash\{0, 1728\}$ such that ρ_A is surjective if $j(A) = j_0$. This follows from an application of the Hilbert irreducibility theorem along the lines of Serre [10, pp. 148–149] (a reference which should have been included in [5]). But as remarked in [5], an appeal to the Hilbert irreducibility theorem does not provide even one explicit value of j_0 for which surjectivity holds. The present note is intended to fill the gap. Let

$$\bar{\rho}_A : \mathrm{Gal}(\overline{\mathbb{Q}}/\widetilde{\mathbb{Q}}) \to \mathrm{SL}(2, \mathbb{F}_p)$$

be the reduction of ρ_A modulo the maximal ideal (p, X) of $\mathrm{SL}(2, \mathbb{Z}_p[[X]])$, and write v_p for the standard p-adic valuation on $\mathbb{Q}$, so that $v_p(p) = 1$.

Theorem 1. *If $\bar{\rho}_A$ is surjective and $v_p(j(A)) = -1$ then ρ_A is surjective.*

The proof of Theorem 1 rests on the work of Kubert–Lang [3]. To explain the connection we return to ρ_E but pass immediately to the associated projective representation

$$\rho : \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j)) \to \mathrm{PSL}(2, \mathbb{Z}_p[[X]]),$$

which is independent of E. For integers $v \geqslant 1$, let L_v denote the fixed field of the kernel of the reduction of ρ modulo $(p, X)^v$. If we identify j with the usual elliptic modular function then L_1 is naturally identified with the modular function field over $\widetilde{\mathbb{Q}}$ of congruence level p, while L_2 is identified with the modular function field of a certain noncongruence subgroup of $\mathrm{SL}(2, \mathbb{Z})$. Furthermore L_2/L_1 is an abelian extension of exponent p and hence a Kummer extension, thus generated by pth roots of elements $f \in L_1^\times$. In fact since L_2/L_1 is unramified outside the cusps we see that the divisor of any such f has the form $(f) = D_\infty + pD$, where D is an arbitrary divisor and D_∞ is a divisor supported on the cusps. Now the simplest candidates for functions f with (f) of the required form are those for which $D = 0$, and using [3] we shall prove that the simplest candidates suffice: L_2 is generated over L_1 by pth roots of modular units. The particular group of modular units which enters here is identified precisely in Theorem 2 of Section 1.

The relevance of all of this to Theorem 1 is that a result of Boston [2] reduces the surjectivity of ρ_A to the surjectivity of ρ_A modulo $(p, X)^2$. In fact it suffices to verify the surjectivity of the associated projective representation, so we would like to see that the degree of the field extension L_2/L_1 is preserved under specialization at $j = j(A)$. Here is where our Kummer-theoretic description of this extension comes in: The surjectivity of ρ_A is now reduced to the injectivity of the specialization map on the group of modular units in Theorem 2, or rather on the quotient of this group by the subgroup of pth powers of modulo units. Finally, to see that the desired injectivity does in fact hold we use the theory of the Tate curve [11]. Here too the method is inspired by Kubert–Lang; cf. [3, Theorem 2.1, p. 182].

The main point of Theorem 1 is that it provides the "explicit example[s]" of surjectivity sought in [5]. For instance take $p = 11$ and let A be the curve $y^2 + y = x^3 - x^2$. Then $\bar{\rho}_A$ is surjective [9, p. 309] and $j(A) = -2^{12}/11$, whence ρ_A is surjective also. It should be added that another problem posed in [5], namely to derive the results of that paper within the framework of Galois deformation theory, has been solved by Boeckle [1] in a vastly more general context. On the other hand, the third problem mentioned in [5] (openness of the image of ρ_A when A does not have complex multiplication) still awaits resolution.

1. The Siegel units

As usual, $\Gamma(p)$ denotes the principal congruence subgroup of $\mathrm{SL}(2, \mathbb{Z})$ of level p. Also I is the 2×2 identity matrix and $\Gamma(p)^{\pm} = \{\pm I\}\Gamma(p)$. The compositum of all modular function fields associated to subgroups of finite index in $\mathrm{SL}(2, \mathbb{Z})$ will be denoted $\mathfrak{M}$, and the modular function field associated to a particular such subgroup Γ will be denoted $\mathfrak{M}^{\Gamma}$; however the modular function field $\mathfrak{M}^{\Gamma(p)} = \mathfrak{M}^{\Gamma(p)^{\pm}}$ will be denoted simply $\mathfrak{K}$. Put $G = \mathrm{PSL}(2, \mathbb{F}_p)$. Given $g \in G$, we define $\sigma(g) \in \mathrm{Gal}(\mathfrak{K}/\mathbb{C}(j))$ by the formula

$$\sigma(g)(f) = f \circ \gamma^{-1} \quad (f \in \mathfrak{K}),$$

where $\gamma \in \mathrm{SL}(2, \mathbb{Z})$ is any coset representative for the preimage of g under the natural identification $\mathrm{SL}(2, \mathbb{Z})/\Gamma(p)^{\pm} \cong G$. Strictly speaking, in an expression like $f \circ \gamma^{-1}$ the symbol γ actually stands for the fractional linear transformation defined by γ. In any case, the map $g \mapsto \sigma(g)$ identifies G with $\mathrm{Gal}(\mathfrak{K}/\mathbb{C}(j))$ and thus enables us to regard modules for the latter group as modules for the former. This comment applies in particular to the group U of modular units for $\Gamma(p)$. By definition, U is the subgroup of $\mathfrak{K}^{\times}$ consisting of all $f \in \mathfrak{K}^{\times}$ which are holomorphic and nowhere zero on the upper half-plane, and according to the remark just made we may view U and U/U^p as modules over $\mathbb{Z}[G]$ and $\mathbb{F}_p[G]$ respectively.

The information about U that we need here will simply be quoted from [3], but to encapsulate it in a statement suited to our application requires some additional notation. Put $R = \mathbb{F}_p^2 \backslash \{(0, 0)\}$ and let M be the $\mathbb{Z}$-module consisting of

functions $m : R \to \mathbb{Z}$ such that $m(-r) = m(r)$. We make M into a $\mathbb{Z}[G]$-module by declaring that

$$(gm)(r) = m(r\tilde{g}),$$

where r is viewed as a row vector (r_1, r_2) and $\tilde{g} \in \mathrm{SL}(2, \mathbb{F}_p)$ denotes either of the two preimages of $g \in G$. We also write Q for the submodule of M defined by the "quadratic relations" of Kubert and Lang. Thus Q consists of all $m \in M$ such that

$$\sum_{r \in R} m(r)n(r) \equiv 0 \pmod{p}$$

whenever $n \in M$ reduces mod p to the function $R \to \mathbb{F}_p$ defined by a homogeneous polynomial of degree two over $\mathbb{F}_p$. Note that Q is also characterized by the vanishing in $\mathbb{F}_p$ of the three sums $\sum \bar{m}(r)r_1^2$, $\sum \bar{m}(r)r_2^2$, and $\sum \bar{m}(r)r_1r_2$, where $\bar{m}$ denotes the reduction of m modulo p. (The vanishing of these sums does follow from the congruences $\sum m(r)n(r) \equiv 0$, because every function $R \to \mathbb{F}_p$ defined by a homogeneous polynomial of degree two can be lifted to an *even* function $R \to \mathbb{Z}$ and thus to an element $n \in M$.) Clearly $Q \supset pM$.

The reason for introducing the module Q is that it provides a description of U in terms of the "Siegel functions" g_a [3, p. 29]. Indeed given $r \in R$, write f_r for any function of the form g_a^{12} with $a \in p^{-1}\mathbb{Z}^2$ and r equal to the residue class of pa modulo $p\mathbb{Z}^2$. Then for $m \in M$ the symbolic mth power

$$f^m = \prod_{r \in R} f_r^{m(r)}$$

is invariant under $\Gamma(p)$—hence belongs to U—if and only if $m \in Q$ [3, Theorem 5.2, p. 76]. Without specifying the functions f_r more precisely we cannot quite claim that the assignment $m \mapsto f^m$ gives a map $Q \to U$, because if a is replaced by some other element of its coset modulo $\mathbb{Z}^2$ then g_a is multiplied by a pth root of unity, whence f^m is defined only up to a pth root of unity also. However since U^p contains the constant functions, the *coset* of f^m modulo U^p is uniquely determined, whence the formula

$$\Theta(m + pQ) = f^m U^p$$

does define a homomorphism $\Theta : Q/pQ \to U/U^p$. We are now in a position to summarize the key inputs needed from [3]:

Proposition 0. *View Q/pQ and U/U^p as modules over $\mathbb{F}_p[G]$ and hence in particular as vector spaces over $\mathbb{F}_p$. Then Θ is an $\mathbb{F}_p[G]$-module homomorphism and is surjective with kernel of dimension one.*

Proof. The surjectivity follows from [3, Theorem 1.3, p. 83] because the map $u \mapsto u^{12}$ is an automorphism of U/U^p. The fact that Θ intertwines the action of G on Q/pQ and U/U^p follows from [3, Formula K1, p. 27]. To see that the kernel of Θ is

one-dimensional it suffices in view of the surjectivity to compare the dimensions of domain and range. The dimension of Q/pQ over $\mathbb{F}_p$ coincides with the rank of Q as a free $\mathbb{Z}$-module and hence with the rank of M, namely $(p^2-1)/2$. On the other hand, the dimension of U/U^p is $|C(p)|/2-1$ [3, Theorem 3.2, p. 42], where $C(p)$ is the nonsplit Cartan subgroup of $\mathrm{GL}(2,\mathbb{F}_p)$, which has order (p^2-1). □

It is convenient at this juncture to move freely between the language of $\mathbb{F}_p[G]$-modules and the equivalent language of representations of G over $\mathbb{F}_p$. In particular, since G has no nontrivial one-dimensional representations we see that the representation of G on $\ker\Theta$ is trivial. Thus, if T is a $\mathbb{Z}[G]$-submodule of Q containing pQ such that the trivial representation does not occur in T/pQ then Θ maps T/pQ isomorphically onto its image in U/U^p. On the other hand, it is immediate from the definition of U that the natural map $U/U^p \to \mathfrak{K}^\times/\mathfrak{K}^{\times p}$ is also an embedding. Therefore Kummer theory gives the following statement:

Proposition 1. *Let T be a $\mathbb{Z}[G]$-submodule of Q containing pQ, and let $\mathfrak{K}_T$ be the subfield of $\mathfrak{M}$ generated over $\mathfrak{K}$ by the pth roots of the functions f^m with $m\in T$. Then $\mathfrak{K}_T$ is abelian of exponent p over $\mathfrak{K}$ and Galois over $\mathbb{C}(j)$, and if the trivial representation of G does not occur in T/pQ then $\mathrm{Gal}(\mathfrak{K}_T/\mathfrak{K})$ is isomorphic as an $\mathbb{F}_p[G]$-module to the dual of T/pQ.*

For example if $T=pM$ then the trivial representation does not occur in T/pQ because pM/pQ is irreducible of dimension >1. To verify the irreducibility consider the isomorphic module M/Q. It is easy to see that the representation of G on M/Q is nontrivial, and of course any nontrivial representation of a simple group is faithful. Furthermore, $\dim M/Q=3$ because the quadratic relations amount to the simultaneous vanishing of three linearly independent linear forms on M/pM. On the other hand, the only irreducible representation of G of dimension <3 is the trivial representation. Hence if M/Q is reducible then relative to a basis underlying a Jordan–Hölder filtration the matrices representing G are unipotent. Since G is not a p-group it follows that the representation is not faithful, a contradiction.

Returning to Proposition 1, we see that there is a Galois extension $\mathfrak{K}_{pM}$ of $\mathbb{C}(j)$ such that $\mathfrak{K}_{pM}$ is abelian of exponent p over $\mathfrak{K}$ and such that the representation of G on $\mathrm{Gal}(\mathfrak{K}_{pM}/\mathfrak{K})$ is the irreducible three-dimensional representation (which of course is self-dual).

Proposition 2. $\mathfrak{K}_{pM}=\mathfrak{M}^{\Gamma(p^2)}$.

Proof. By definition $\mathfrak{K}_{pM}=\mathfrak{K}(\{f^m : m\in M\})$, and therefore $\mathfrak{K}_{pM}\subset\mathfrak{M}^{\Gamma(p^2)}$ [3, Formula K3, p. 28]. But we have just seen that $[\mathfrak{K}_{pM}:\mathfrak{K}]=p^3=[\mathfrak{M}^{\Gamma(p^2)}:\mathfrak{K}]$. □

Remark. Another way to see that the containment $\mathfrak{K}_{pM}\subset\mathfrak{M}^{\Gamma(p^2)}$ is actually an equality is to recognize the representation of G on $\Gamma(p)/\Gamma(p^2)$ as the adjoint representation. The irreducibility of pM/pQ is then a corollary.

The next statement concerns a submodule N of M which figured implicitly in the original definition of Q. The proof will be given in Section 2.

Proposition 3. *Let N be the $\mathbb{Z}[G]$-submodule of M consisting of all $n \in M$ such that $\bar{n}$ coincides with the function $R \to \mathbb{F}_p$ defined by a homogeneous polynomial over $\mathbb{F}_p$ of degree two. Then $N \subset Q$, and the exact sequence of vector spaces over $\mathbb{F}_p$*

$$\{0\} \to pM/pQ \to N/pQ \to N/pM \to \{0\}$$

splits as an exact sequence of $\mathbb{F}_p[G]$-modules.

It is immediate from the definition of N that the representation of G on N/pM is isomorphic to the representation of G on the space of binary quadratic forms over $\mathbb{F}_p$, or in other words to the symmetric square of the standard projective representation. Of course, we already know that the representation of G on pM/pQ is likewise the irreducible three-dimensional representation. Thus we can appeal to the uniqueness of Jordan–Hölder constituents (or simply to the uniqueness of direct summands, in light of Proposition 3) to conclude that the trivial representation of G does not occur in N/pQ. Hence taking $T = N$ in Proposition 1 we obtain an extension $\mathfrak{K}_N$ of $\mathfrak{K}$, abelian of exponent p and Galois over $\mathbb{C}(j)$, such that the hypotheses of the following proposition are satisfied with $\Sigma = \mathrm{Gal}(\mathfrak{K}_N/\mathbb{C}(j))$, $A = \mathrm{Gal}(\mathfrak{K}_N/\mathfrak{K})$, and $B = \mathrm{Gal}(\mathfrak{K}_N/\mathfrak{K}_{pM})$.

Proposition 4. *Let Σ be a finite group and let A and B be abelian normal subgroups of exponent p, with $B \subset A$. Put $G = \mathrm{PSL}(2, \mathbb{Z}/p\mathbb{Z})$ and $J = \mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$, and suppose that there is an isomorphism $\Sigma/B \cong J$ mapping A/B onto the kernel of the reduction map $J \to G$. Assume also that B is irreducible and three-dimensional as a module for $G \cong \Sigma/A$. If*

$$\{0\} \to B \to A \to A/B \to \{0\}$$

splits as an exact sequence of $\mathbb{F}_p[G]$-modules then

$$\{0\} \to B \to \Sigma \to J \to \{1\}$$

splits as an exact sequence of abstract groups.

Proposition 4 will be proved in Section 3. Now put $\Lambda = \mathbb{Z}_p[[X]]$, and for $v \geqslant 1$, let Λ_v denote the quotient ring $\Lambda/(p, X)^v$. Under the hypotheses of Proposition 4 we obtain the following corollary:

Corollary 1. $\Sigma \cong \mathrm{PSL}(2, \Lambda_2)$.

Proof. The isomorphism class of B as a J-module is intrinsic to J—i.e. can be described without reference to Σ—because there is a unique isomorphism class of irreducible three-dimensional representations of G and there is also a unique normal

subgroup H of J such that $G \cong J/H$ (the uniqueness of H holds more generally for any normal subgroup of order p^{k-1} in a finite group with a nonnormal Sylow subgroup of order p^k). Now according to Proposition 4, Σ is the split extension of J by B. Hence, to deduce the corollary it will suffice to show that $\mathrm{PSL}(2, \Lambda_2)$ has the same description.

Let $\mathfrak{sl}(2, \mathbb{F}_p)$ denote the space of 2×2 matrices over $\mathbb{F}_p$ of trace 0. We view $\mathfrak{sl}(2, \mathbb{F}_p)$ as a J-module via the adjoint representation of G, so that $\mathfrak{sl}(2, \mathbb{F}_p) \cong B$. To see that $\mathrm{PSL}(2, \Lambda_2)$ is the split extension of J by $\mathfrak{sl}(2, \mathbb{F}_p)$, consider the ring homomorphism $\Lambda_2 = \mathbb{Z}[X]/(p, X)^2 \to \mathbb{Z}/p^2\mathbb{Z}$ specializing X to 0. This homomorphism has a section embedding $\mathbb{Z}/p^2\mathbb{Z}$ as a subring of Λ_2, and consequently the induced map $\mathrm{PSL}(2, \Lambda_2) \to J$ has a section also. So to complete the proof it suffices to identify the kernel of $\mathrm{PSL}(2, \Lambda_2) \to J$ with $\mathfrak{sl}(2, \mathbb{F}_p)$. An identification is provided by the map sending the matrix $D \in \mathfrak{sl}(2, \mathbb{F}_p)$ to $\pm(I + XD) \in \mathrm{PSL}(2, \Lambda_2)$, the product of X and an integer modulo p being interpreted as an element of Λ_2. □

Returning to the application at hand, we obtain the structure of $\mathrm{Gal}(\mathfrak{K}_N/\mathbb{C}(j))$:

Corollary 2. $\mathrm{Gal}(\mathfrak{K}_N/\mathbb{C}(j)) \cong \mathrm{PSL}(2, \Lambda_2)$.

Our subsequent use of this information may appear inefficient but facilitates appeals to [5], where Galois extensions with Galois group $\mathrm{PSL}(2, \Lambda_v)$ are slighted in favor of the intermediate extensions with group $\mathrm{PSL}(2, \mathbb{Z}/p^v\mathbb{Z})$. The following proposition implies that when $v = 2$ the big extension is the compositum of these intermediate extensions, a property which fails for $v \geqslant p + 1$.

Proposition 5. *The intersection of the kernels of the p distinct epimorphisms*

$$\mathrm{PSL}(2, \Lambda_2) \to \mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$$

afforded by the p possible specializations of X to an element of $p\mathbb{Z}/p^2\mathbb{Z}$ is trivial.

Proof. Given $a, b \in \mathbb{Z}_p$, one readily checks that if the specialization $X \mapsto \xi$ maps $a + bX$ to 0 for every choice of $\xi \in p\mathbb{Z}/p^2\mathbb{Z}$ then $a \in p^2\mathbb{Z}_p$ and $b \in p\mathbb{Z}_p$. It follows that the intersection of the kernels of the ring homomorphisms $\Lambda_2 \to \mathbb{Z}/p^2\mathbb{Z}$ is trivial. The corresponding statement for $\mathrm{PSL}(2, \Lambda_2) \to \mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$ is an immediate consequence. □

We now descend from $\mathbb{C}$ to the subfield

$$\widetilde{\mathbb{Q}} = \bigcup_{v \geqslant 1} \mathbb{Q}(e^{2\pi i/p^v}).$$

Let K denote the subfield of $\mathfrak{K}$ consisting of modular functions for $\Gamma(p)$ with Fourier coefficients in $\widetilde{\mathbb{Q}}$, so that K is Galois over $\widetilde{\mathbb{Q}}(j)$ with $\mathbb{C}K = \mathfrak{K}$ and $K \cap \mathbb{C}(j) = \widetilde{\mathbb{Q}}(j)$.

These last two properties amount respectively to the injectivity and surjectivity of the restriction map $\mathrm{Gal}(\mathfrak{K}/\mathbb{C}(j))\to\mathrm{Gal}(K/\widetilde{\mathbb{Q}}(j))$, so our identification of G with the first Galois group becomes an identification of G with the second as well.

Proposition 6. *Let T be a $\mathbb{Z}[G]$-submodule of Q containing pQ, and let K_T be the extension of K generated by the pth roots of the functions f^m with $m\in T$. Assume that the trivial representation of G does not occur in T/pQ. Then K_T is Galois over $\widetilde{\mathbb{Q}}(j)$ with $\mathbb{C}K_T=\mathfrak{K}_T$ and $\mathbb{C}(j)\cap K_T=\widetilde{\mathbb{Q}}(j)$, and consequently the restriction map $\mathrm{Gal}(\mathfrak{K}_T/\mathbb{C}(j))\to\mathrm{Gal}(K_T/\widetilde{\mathbb{Q}}(j))$ is an isomorphism.*

Proof. The explicit formula for the Fourier expansion of a Siegel function shows that if $m\in Q$ then $f^m\in K$. Thus K_T is a Kummer extension of K and hence in particular a Galois extension, and since T is stable under G it follows that K_T is actually Galois over $\widetilde{\mathbb{Q}}(j)$. As $\mathbb{C}K=\mathfrak{K}$ the definitions give $\mathbb{C}K_T=\mathfrak{K}_T$, so to see that $\mathbb{C}(j)\cap K_T=\widetilde{\mathbb{Q}}(j)$ it suffices to see that $[K_T:\widetilde{\mathbb{Q}}(j)]=[\mathfrak{K}_T:\mathbb{C}(j)]$. In fact since $[K:\widetilde{\mathbb{Q}}(j)]=[\mathfrak{K}:\mathbb{C}(j)]$ it suffices to see that

$$[K_T:K]=[\mathfrak{K}_T:\mathfrak{K}].$$

This last equality is equivalent by Kummer theory to $|\Psi(T/pQ)|=|\Upsilon(T/pQ)|$, where the maps $\Psi:Q/pQ\to K^\times/K^{\times p}$ and $\Upsilon:Q/pQ\to\mathfrak{K}^\times/\mathfrak{K}^{\times p}$ send the coset represented by an element $m\in Q$ to the coset of f^m in $K^\times/K^{\times p}$ and $\mathfrak{K}^\times/\mathfrak{K}^{\times p}$, respectively. Now, Υ is the composition of Θ with the natural embedding $U/U^p\to\mathfrak{K}^\times/\mathfrak{K}^{\times p}$. Since the restriction of Θ to T/pQ is by assumption injective, so is the restriction of Υ. As Υ factors through Ψ it follows that $\Psi|(T/pQ)$ is also injective, whence $|\Psi(T/pQ)|=|T/pQ|=|\Upsilon(T/pQ)|$. □

Remark. The kernel of Θ is spanned by the coset of the function $m_0\in Q$ which is identically 1 on R. This follows both from the fact that the representation of G on $\mathrm{Ker}\,\Theta$ is trivial and also from the "distribution relation" $f^{m_0}=\zeta p^{12}$, where ζ is a pth root of unity [3, p. 45]. The fact that ζp^{12} is a pth power in $\mathbb{C}$ but not in $\widetilde{\mathbb{Q}}$ shows that Proposition 6 is false without the assumption that the trivial representation does not occur in T/pQ.

Finally, let L be the Galois extension of $\widetilde{\mathbb{Q}}(j)$ defined in [5] (take $F=\widetilde{\mathbb{Q}}$ in Theorem 2 on p. 250 or in Proposition 7 on p. 278). The key properties of L are that L contains K and that $\mathrm{Gal}(L/\widetilde{\mathbb{Q}}(j))\cong\mathrm{PSL}(2,\Lambda)$. Now for a positive integer v the kernel of the reduction map $\mathrm{PSL}(2,\Lambda)\to\mathrm{PSL}(2,\Lambda_v)$ is a characteristic subgroup of $\mathrm{PSL}(2,\Lambda)$ and hence determines a fixed field $L_v\subset L$ which is independent of the choice of identification $\mathrm{Gal}(L/\widetilde{\mathbb{Q}}(j))\cong\mathrm{PSL}(2,\Lambda)$. By Galois theory we have $L_1=K$, because an open normal subgroup of $\mathrm{PSL}(2,\Lambda)$ with quotient group G is unique.

Theorem 2. *The extension L_2/L_1 is generated by pth roots of Siegel units. More precisely, $L_2 = L_1(\{(f^m)^{1/p} : m \in N\})$.*

Proof. In the notation of Proposition 6, $L_1(\{(f^m)^{1/p} : m \in N\}) = K_N$, so the assertion to be proved is that $L_2 = K_N$. Since K_N has the same degree over $\widetilde{\mathbb{Q}}(j)$ as L_2, namely $|\mathrm{PSL}(2, \Lambda_2)|$ (Corollary 1), it suffices to see that $K_N \subset L_2$. In fact it suffices to see that L_2 contains every subfield of K_N which is Galois over $\widetilde{\mathbb{Q}}(j)$ with Galois group $\mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$, because Proposition 5 implies that K_N is the compositum of such subfields. But K is the unique subfield of K_N which is Galois over $\widetilde{\mathbb{Q}}(j)$ with Galois group $\mathrm{PSL}(2, \mathbb{Z}/p\mathbb{Z})$. Hence a subfield of K_N which is Galois over $\widetilde{\mathbb{Q}}(j)$ with Galois group $\mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$ necessarily contains K and so by Kummer theory has the form K_T for some $\mathbb{Z}[G]$-submodule T of N containing pQ. The field $\mathfrak{K}_T$ is then of the form $\mathfrak{M}^\Gamma$ for some normal subgroup Γ of $\mathrm{SL}(2, \mathbb{Z})$ contained in $\Gamma(p)^{\pm}$. Consequently K_T coincides with the field K_Γ of [5, Proposition 3, p. 260] and is therefore contained in L [5, Proposition 5, p. 272]. In fact $K_T \subset L_2$, because any continuous epimorphism $\mathrm{PSL}(2, \Lambda) \to \mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$ factors through the reduction map $\mathrm{PSL}(2, \Lambda) \to \mathrm{PSL}(2, \Lambda_2)$. □

Remark. The constant subfield of the field $K = L_1$ is by definition $\widetilde{\mathbb{Q}}$, and for the application to Theorem 1 nothing smaller is needed. However, if we let K' be the subfield of K consisting of functions with Fourier coefficients in $\mathbb{Q}(e^{2\pi i/p^2})$ then the Kummer extension L_2/L_1 is simply the compositum with $\widetilde{\mathbb{Q}}$ of a Kummer extension of K' of the same degree. This follows from the explicit formula for the q-expansion of g_a [3, K4, p. 29], which shows that the functions f_r (hence also the functions f^m with $m \in N$) are elements of K'.

2. Proof of Proposition 3

To prove that $N \subset Q$ we must show that the sum $\sum_{r \in R} n(r)n'(r)$ vanishes mod p for all $n, n' \in N$. Now modulo p the expression $n(r)n'(r)$ is a homogeneous polynomial of degree four over $\mathbb{F}_p$, hence a linear combination of monomials of the form $r_1^i r_2^j$ with $i, j \geqslant 0$ and $i + j = 4$. As $p > 5$ these conditions imply in particular that $i, j < p - 1$, whence the following lemma enables us to conclude that $N \subset Q$:

Lemma. *Suppose that i and j are nonnegative integers, not both 0, satisfying $i, j < p - 1$. Then*

$$\sum_{r \in R} r_1^i r_2^j = 0,$$

with the understanding that $0^0 = 1$.

Proof. At least one of i and j is strictly positive. Say $i>0$; then

$$\sum_{r\in R} r_1^i r_2^j = \left(\sum_{r_1\neq 0} r_1^i\right)\left(\sum_{r_2\in\mathbb{F}_p} r_2^j\right).$$

Since $0<i<p-1$ the map $r_1\mapsto r_1^i$ is a nontrivial character $\mathbb{F}_p^\times\to\mathbb{F}_p^\times$, and therefore the first sum on the right is 0. □

It remains to see that the $\mathbb{F}_p[G]$-submodule pM/pQ of N/pQ has a complement. The following argument was suggested by the referee and is much more efficient than the original proof.

Let ω denote the Teichmüller character $\mathbb{F}_p^\times\to\mathbb{Z}_p^\times$, and put

$$\mathscr{M} = \mathbb{Z}_p\otimes_{\mathbb{Z}} M, \quad \mathscr{N} = \mathbb{Z}_p\otimes_{\mathbb{Z}} N \quad \text{and} \quad \mathscr{Q} = \mathbb{Z}_p\otimes_{\mathbb{Z}} Q.$$

Given a rational integer k we define a $\mathbb{Z}_p$-linear endomorphism e_k of $\mathscr{M}$ by setting

$$e_k = \frac{1}{p-1}\sum_{x\in\mathbb{F}_p^\times} \omega^{-k}(x)\langle x\rangle,$$

where $\langle x\rangle$ is the operator

$$(\langle x\rangle m)(r) = m(xr) \quad (m\in\mathscr{M}).$$

Note that we are identifying elements of $\mathscr{M}$ with functions $m: R\to\mathbb{Z}_p$ satisfying $m(-r)=m(r)$. Similarly, $\mathscr{N}$ consists of all $m\in\mathscr{M}$ such that $\bar{m}$ is the function $R\to\mathbb{F}_p$ defined by a homogeneous polynomial of degree two, and $\mathscr{Q}$ consists of all m such that $\sum_{r\in R}\bar{m}(r)\bar{n}(r)=0$ for $n\in\mathscr{N}$. The idempotents e_k afford direct sum decompositions

$$\mathscr{M} = \bigoplus_{k=0}^{p-2}\mathscr{M}_k, \quad \mathscr{N} = \bigoplus_{k=0}^{p-2}\mathscr{N}_k \quad \text{and} \quad \mathscr{Q} = \bigoplus_{k=0}^{p-2}\mathscr{Q}_k$$

with $\mathscr{M}_k = e_k\mathscr{M}$, $\mathscr{N}_k = e_k\mathscr{N}$, and $\mathscr{Q}_k = e_k\mathscr{Q}$, and since the action of G commutes with the operators $\langle x\rangle$ we see that all of these direct summands are $\mathbb{Z}_p[G]$-modules.

Proposition 7. *Let k be an integer satisfying $0\leqslant k\leqslant p-2$.*

(i) *If $k\neq 2$ then $\mathscr{N}_k = p\mathscr{M}_k$.*
(ii) *If $k\neq p-3$ then $\mathscr{M}_k = \mathscr{Q}_k$.*

Note that $p-3\neq 2$ as $p>5$. Thus granting Proposition 7 we have

$$\mathscr{N}/p\mathscr{Q}\cong\bigoplus_{k=0}^{p-2}\mathscr{N}_k/p\mathscr{Q}_k\cong\mathscr{N}_2/p\mathscr{M}_2\oplus p\mathscr{M}_{p-3}/p\mathscr{Q}_{p-3}\cong\mathscr{N}/p\mathscr{M}\oplus p\mathscr{M}/p\mathscr{Q}.$$

Proposition 3 follows, for we have canonical identifications $\mathscr{N}/p\mathscr{Q}\cong N/pQ$, $\mathscr{N}/p\mathscr{M}\cong N/pM$, and $p\mathscr{M}/p\mathscr{Q}\cong pM/pQ$.

Proof of Proposition 7. (i) If we think of the operators $\langle x\rangle$ as giving an action of $\mathbb{F}_p^\times$ on $\mathscr{M}$ then the submodules $\mathscr{M}_k$, $\mathscr{N}_k$, and $\mathscr{Q}_k$ are the ω^k-isotypic components of $\mathscr{M}$, $\mathscr{N}$, and $\mathscr{Q}$, respectively. On the other hand, let V be the vector space over $\mathbb{F}_p$ consisting of functions $v: R\to\mathbb{F}_p$ satisfying $v(-r)=v(r)$, and let V_k be the subspace consisting of v such that $v(xr)=x^kv(r)$ for $x\in\mathbb{F}_p^\times$. Then V is the direct sum of the subspaces V_k for $0\leqslant k\leqslant p-2$, and the natural isomorphism $\mathscr{M}/p\mathscr{M}\cong V$ identifies $\mathscr{M}_k/p\mathscr{M}_k$ with V_k. But this isomorphism also identifies $\mathscr{N}/p\mathscr{M}$ with the space of functions $R\to\mathbb{F}_p$ defined by homogeneous binary polynomials of degree two over $\mathbb{F}_p$, and the latter space is contained in V_2. We conclude that $\mathscr{N}_k/p\mathscr{M}_k=\{0\}$ if $k\neq 2$.

(ii) Given $\alpha\in\mathbf{P}^1(\mathbb{F}_p)$ let $r_\alpha\in R$ be a point on the line represented by α. For $m\in\mathscr{M}_k$ and $n\in\mathscr{N}$ we have

$$\sum_{r\in R}\bar{m}(r)\bar{n}(r)=\sum_{\alpha\in\mathbf{P}^1(\mathbb{F}_p)}\sum_{x\in\mathbb{F}_p^\times}\bar{m}(xr_\alpha)\bar{n}(xr_\alpha)=\sum_{\alpha\in\mathbf{P}^1(\mathbb{F}_p)}\bar{m}(r_\alpha)\bar{n}(r_\alpha)\sum_{x\in\mathbb{F}_p^\times}x^{k+2}.$$

If $k\neq p-3$ then $k+2\not\equiv 0 \bmod p-1$ and consequently the inner sum is 0, whence $m\in\mathscr{Q}_k$. □

3. Proof of Proposition 4

We must show that the extension class $c\in H^2(J,B)$ determined by the isomorphism $\Sigma/B\cong J$ is 0. Let P be a Sylow p-subgroup of J. The restriction–corestriction sequence shows that the restriction map from $H^2(J,B)$ to $H^2(P,B)$ is injective, and consequently it suffices to see that the image of c under this map, say c_P, is 0.

Let Π be the extension of P by B afforded by c_P, and view Π as a subgroup of Σ. Then Π is a Sylow p-subgroup and so contains the normal p-subgroup A. Furthermore Π/A is isomorphic to the image of Π in G and hence to a Sylow p-subgroup of G. Therefore, Π/A is cyclic of order p. Choose $\pi\in\Pi$ so that πB has order p^2 in Π/B and maps to a generator of Π/A (the latter condition actually implies the former). Since A is a normal subgroup of Π and a vector space over $\mathbb{F}_p$ we may view π as a linear automorphism of A. We claim that there is a π-stable subspace C of A such that $\pi^p\in C$ and $A=B\oplus C$. The proof of this claim will complete the

argument, for the claim implies that the subgroup generated by C and π is a complement to B inside Π, whence $c_P = 0$.

By hypothesis there is an $\mathbb{F}_p[G]$-submodule B' of A such that $A = B \oplus B'$. Furthermore, both B and B' realize the irreducible three-dimensional representation of G, which sends an element of order p in G to a matrix with minimal polynomial $(X-1)^3$. It follows that B and B' are free modules of rank 1 over the local ring $\mathbb{F}_p[\pi] \cong \mathbb{F}_p[X]/((X-1)^3)$. Let b and b' be generators of B and B' respectively over $\mathbb{F}_p[\pi]$. Then $(\pi-1)^2 b$ and $(\pi-1)^2 b'$ are a basis over $\mathbb{F}_p$ for the 1-eigenspace of π on A, and consequently π^p is a linear combination of $(\pi-1)^2 b$ and $(\pi-1)^2 b'$. On the other hand π^p is not simply a multiple of $(\pi-1)^2 b$, for otherwise $\pi^p \in B$, contradicting the fact that the image of π in Π/B has order p^2. It follows that

$$\pi^p = \alpha(\pi-1)^2 b + \alpha'(\pi-1)^2 b'$$

with $\alpha, \alpha' \in \mathbb{F}_p$ and $\alpha' \neq 0$. Now put $c = \alpha b + \alpha' b'$, and consider the classes of b and c in $A/(\pi-1)A$. Since the classes of b and b' constitute a basis for $A/(\pi-1)A$ over $\mathbb{F}_p$ the same is true of b and c (here we use the fact that $\alpha' \neq 0$). Hence b and c constitute a basis for A over $\mathbb{F}_p[\pi]$ by Nakayama's lemma. We conclude that the subspace C of A spanned by c, $(\pi-1)c$, and $(\pi-1)^2 c = \pi^p$ has the required properties.

4. Proof of Theorem 1: reduction to a local statement

It is a standard remark, valid for any integral domain Δ, that the only subgroup of $\mathrm{SL}(2,\Delta)$ with projective image $\mathrm{PSL}(2,\Delta)$ is $\mathrm{SL}(2,\Delta)$ itself. Thus to prove that ρ_A is surjective it suffices to verify the surjectivity of the associated projective representation $\mathrm{P}\rho_A$. The latter depends only on $j(A)$ and can be described as follows. Let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j)) \to \mathrm{PSL}(2,\Lambda)$ be the epimorphism with kernel $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/L)$ defined on p. 302 of [6]. Choose a place $\varphi : \overline{\mathbb{Q}(j)} \to \overline{\mathbb{Q}} \cup \{\infty\}$ extending the place $j = j(A)$ of $\widetilde{\mathbb{Q}}(j)$, and write D and I for the corresponding decomposition and inertia subgroups of $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\widetilde{\mathbb{Q}}(j))$. Since the map ρ of [6] is unramified outside $\{0, 1728, \infty\}$ and hence in particular at $j(A)$, the restriction of ρ to D factors through D/I. Up to equivalence, $\mathrm{P}\rho_A$ is simply the composition of the isomorphism $\mathrm{Gal}(\overline{\mathbb{Q}}/\widetilde{\mathbb{Q}}) \cong D/I$ induced by φ and the homomorphism $D/I \to \mathrm{PSL}(2,\Lambda)$ induced by $\rho|D$. Thus to verify that $\mathrm{P}\rho_A$ is surjective it suffices to see that the image of D in $\mathrm{Gal}(L/\widetilde{\mathbb{Q}}(j))$ is all of $\mathrm{Gal}(L/\widetilde{\mathbb{Q}}(j))$. In fact by the criterion of Boston [2, Proposition 2, p. 262] it suffices to verify that the image of D in $\mathrm{Gal}(L_2/\widetilde{\mathbb{Q}}(j))$ is all of $\mathrm{Gal}(L_2/\widetilde{\mathbb{Q}}(j))$. Equivalently, we want $[L_2 : \widetilde{\mathbb{Q}}(j)] = [\ell_2 : \widetilde{\mathbb{Q}}]$, where ℓ_v is the residue class field of $\varphi|L_v$ (i.e. $\ell_v = \varphi(L_v)\backslash\{\infty\}$). Our assumption that $\bar{\rho}_A$ is surjective implies that $[L_1 : \widetilde{\mathbb{Q}}(j)] = [\ell_1 : \widetilde{\mathbb{Q}}]$, so it suffices to prove that $[L_2 : L_1] = [\ell_2 : \ell_1]$.

There is actually a further reduction. For simplicity, let k be an alternate notation for ℓ_1 just as K is an alternate notation for L_1. Also write ψ for the map

$Q/pQ \to k^\times/k^{\times p}$ sending the coset represented by an element $m \in Q$ to the coset represented by $\varphi(f^m) \in k^\times$. Then Theorem 2 and Kummer theory give $[L_2 : L_1] = |N/pQ|$ and $[\ell_2 : \ell_1] \geqslant |\psi(N/pQ)|$. Hence Theorem 1 will follow if we prove that the restriction of ψ to N/pQ is injective. But we have identified G with $\mathrm{Gal}(K/\widetilde{\mathbb{Q}}(j))$ and hence via φ with $\mathrm{Gal}(k/\widetilde{\mathbb{Q}})$, and under these identifications ψ becomes a G-map. Furthermore the quotient modules in the filtration $\{0\} \subsetneqq pM/pQ \subsetneqq N/pQ$ both afford the irreducible three-dimensional representation of G. Thus to see that ψ is injective on N/pQ it suffices to see that $\{1\} \subsetneqq \psi(pM/pQ) \subsetneqq \psi(N/pQ)$. Equivalently, our task is to show that

$$k \subsetneqq k_{pM} \subsetneqq k_N,$$

where for any $\mathbb{Z}[G]$-submodule T of Q containing pQ we write k_T to denote the residue class field of $\varphi|K_T$. Of course the only point requiring proof is the *strictness* of the above inclusions, in other words the fact that $k \neq k_{pM} \neq k_N$.

The first of these inequalities follows from our assumption that $\bar{\rho}_A$ is surjective. Indeed, the surjectivity of $\bar{\rho}_A$ gives the surjectivity of the representation $\mathrm{Gal}(\overline{\mathbb{Q}}/\widetilde{\mathbb{Q}}) \to \mathrm{SL}(2, \mathbb{Z}_p)$ afforded by the p-adic Tate module of A [8, Lemma 30, p. IV-23] hence also the surjectivity of the associated projective representation and of its reduction mod p^2. As the extension of $\widetilde{\mathbb{Q}}$ generated by the x-coordinates of the points of order p^2 on A coincides by Proposition 2 with k_{pM}, it follows that $\mathrm{Gal}(k_{pM}/\widetilde{\mathbb{Q}}) \cong \mathrm{PSL}(2, \mathbb{Z}/p^2\mathbb{Z})$. In particular, $k \neq k_{pM}$.

To prove that $k_{pM} \neq k_N$ we work locally. After embedding $\overline{\mathbb{Q}}$ in an algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ we may form the composita $\mathbf{k} = \mathbb{Q}_p k$ and $\mathbf{k}_T = \mathbb{Q}_p k_T$, and then it suffices to prove that $\mathbf{k}_{pM} \neq \mathbf{k}_N$. Thus we will be done if we show that for some $m \in N$ we have $\varphi(f^m)^{1/p} \notin \mathbf{k}_{pM}$, where $\varphi(f^m)^{1/p}$ denotes an arbitrary pth root of $\varphi(f^m)$. Now we have already noted that k_{pM} is the extension of $\widetilde{\mathbb{Q}}$ generated by the x-coordinates of the points of order p^2 on A, and this statement remains true if A is replaced by any other elliptic curve over $\mathbb{Q}$ with the same modular invariant. In fact an analogous statement holds over the compositum $\widetilde{\mathbb{Q}}_p = \mathbb{Q}_p\widetilde{\mathbb{Q}}$: since $v_p(j(A)) < 0$ there is a unique Tate curve B over $\mathbb{Q}_p$ with $j(B) = j(A)$, and $\mathbf{k}_{pM}$ is the extension of $\widetilde{\mathbb{Q}}_p$ generated by the x-coordinates of the points of order p^2 on B. In particular,

$$\mathbf{k}_{pM} \subset \widetilde{\mathbb{Q}}_p(B[p^\infty]),$$

where $B[p^\infty] = \bigcup_{v \geqslant 1} B[p^v]$ and $B[p^v]$ is the group of points on B of order dividing p^v. As $\widetilde{\mathbb{Q}}_p(B[p^\infty]) = \mathbb{Q}_p(B[p^\infty])$ the assertion that $\varphi(f^m)^{1/p} \notin \mathbf{k}_{pM}$ for some $m \in N$ will follow if we prove that $\varphi(f^m)^{1/p} \notin \mathbb{Q}_p(B[p^v])$ for some $m \in N$ and for all sufficiently large v.

We have used the phrase "for all sufficiently large v" in preference to "for all $v \geqslant 1$" because the functions f^m has until now been specified only up to multiplication by a pth root of unity, so that for $v = 1$ the validity of the statement

$\varphi(f^m)^{1/p} \notin \mathbb{Q}_p(B[p^v])$ may appear to depend on the particular pth root of $\varphi(f^m)$ chosen. We can eliminate the ambiguity by making an explicit choice of f_r for $r \in R$. Our understanding all along has been that $f_r = g_a^{12}$ with $a \in p^{-1}\mathbb{Z}^2$ and $r = pa \bmod p\mathbb{Z}^2$, but henceforth we demand that $0 \leqslant a_1, a_2 < 1$. With this requirement the functions f_r are uniquely determined, as is therefore f^m, and the proof of Theorem 1 is reduced to the statement that there is a choice of $m \in N$ such that $\varphi(f^m)^{1/p} \notin \mathbb{Q}_p(B[p^v])$ for $v \geqslant 1$.

5. Proof of Theorem 1: the Tate curve

The preceding reduction holds for any place $\varphi : \overline{\mathbb{Q}(j)} \to \overline{\mathbb{Q}} \cup \{\infty\}$ extending the place $j = j(A)$ of $\widetilde{\mathbb{Q}}(j)$. However we now impose an additional condition. As usual, let $\Gamma_1(p)$ denote the subgroup of $\mathrm{SL}(2, \mathbb{Z})$ consisting of matrices which are strictly upper triangular modulo p. Put

$$K^{\Gamma_1(p)} = K \cap \mathfrak{M}^{\Gamma_1(p)}$$

and write $\mathcal{O}$ for the integral closure of $\widetilde{\mathbb{Q}}[j]$ in $K^{\Gamma_1(p)}$. Then $\mathcal{O}$ is the subring of $K^{\Gamma_1(p)}$ consisting of functions which are holomorphic on the upper half-plane. We also put $q_\tau = e^{2\pi i \tau}$, where τ denotes an element of the upper half-plane. Every $h \in K$ is represented for $\Im(\tau)$ sufficiently large by a finite-tailed Laurent series in $q_{\tau/p}$, and h belongs to $K^{\Gamma_1(p)}$ if and only this series is actually a Laurent series in q_τ. Thus given $h \in K^{\Gamma_1(p)}$ we can write

$$h(\tau) = \sum_{n \geqslant n_0} a(n) q_\tau^n \quad (\Im(\tau) \gg 0)$$

with $n_0 \in \mathbb{Z}$ and $a(n) \in \widetilde{\mathbb{Q}}$ for all n. In fact there exists some finite extension of $\mathbb{Q}$ inside $\widetilde{\mathbb{Q}}$ which contains $a(n)$ for all n. Furthermore, if $h \in \mathcal{O}$ then there is a positive integer d such that $da(n)$ is integral for all n and hence lies in the ring of integers of some finite extension of $\mathbb{Q}$. Consequently if $h \in \mathcal{O}$ and $q \in p\mathbb{Z}_p \backslash \{0\}$ then the series $\sum_{n \geqslant n_0} a(n) q^n$ converges in $\overline{\mathbb{Q}}_p$. Henceforth we take q to be the unique element of $p\mathbb{Z}_p \backslash \{0\}$ such that $j(q) = j(B) = j(A)$, where $j(q) = q^{-1} + 744 + 196884q + \cdots$ as usual (note that q belongs to $p\mathbb{Z}_p^\times$ and not merely to $p\mathbb{Z}_p$ because $v_p(j(A)) = -1$ by assumption). The "additional condition" on φ alluded to at the beginning of the paragraph is that $\varphi|\mathcal{O}$ is required to be the map

$$\sum_{n \geqslant n_0} a(n) q_\tau^n \quad \mapsto \quad \sum_{n \geqslant n_0} a(n) q^n.$$

This map sends j to $j(A)$ and extends uniquely to a place of the quotient field $K^{\Gamma_1(p)}$ of $\mathcal{O}$. For our purposes the extension from $K^{\Gamma_1(p)}$ to $\overline{\mathbb{Q}}(j)$ can be chosen arbitrarily. Note that the residue class field of φ is the subfield $\overline{\mathbb{Q}}$ of $\overline{\mathbb{Q}}_p$, as required.

We now return to the task of showing that $\varphi(f^m)^{1/p} \notin \mathbb{Q}_p(B[p^v])$ for some $m \in N$ and all $v \geqslant 1$. The following proposition will be proved in Section 6.

Proposition 8. *There exists $m \in N$ such that*

$$\varphi(f^m) = q^\mu(1-uq)(1-*q^2)$$

with $\mu \in \mathbb{Z}$, $u \in \mathbb{Z}_p^\times$, and $* \in \mathbb{Z}_p$. In particular, $\varphi(f^m) \in \mathbb{Q}_p$.

Denoting f^m simply by f, and bearing in mind that $v_p(q) = 1$, we see that $q^{-\mu}\varphi(f)$ is a topological generator of the subgroup $1 + p\mathbb{Z}_p$ of $\mathbb{Z}_p^\times$. Hence the group $\mathbb{Q}_p^\times / \mathbb{Q}_p^{\times p^v}$ is generated by the cosets of q and $\varphi(f)$. On the other hand, let F_v denote the extension of $\mathbb{Q}_p$ generated by the p^vth roots of unity. An elementary cohomological argument (using for example [7, Proposition 2.7, p. 60]) shows that the natural map $\mathbb{Q}_p^\times / \mathbb{Q}_p^{\times p^v} \to F_v^\times / F_v^{\times p^v}$ is injective, whence the subgroup of $F_v^\times / F_v^{\times p^v}$ generated by the cosets of q and $\varphi(f)$ has the same order as $\mathbb{Q}_p^\times / \mathbb{Q}_p^{\times p^v}$, namely p^{2v}. It follows that

$$[F_v(q^{1/p^v}, \varphi(f)^{1/p^v}) : F_v] = p^{2n}.$$

Since $[F_v(q^{1/p^v}, \varphi(f)^{1/p^v}) : F_v(q^{1/p^v})]$ and $[F_v(q^{1/p^v}) : F_v]$ are a priori $\leqslant p^v$ we deduce that

$$[F_v(q^{1/p^v}, \varphi(f)^{1/p^v}) : F_v(q^{1/p^v})] = p^v.$$

But the theory of the Tate curve shows that $F_v(q^{1/p^v}) = \mathbb{Q}_p(B[p^v])$, so we are saying that $\varphi(f)^{1/p^v}$ has degree p^v over $\mathbb{Q}_p(B[p^v])$. Therefore $\varphi(f)^{1/p}$ has degree p over $\mathbb{Q}_p(B[p^v])$, and in particular $\varphi(f)^{1/p} \notin \mathbb{Q}_p(B[p^v])$.

6. Proof of Proposition 8

As in [3], if c is a rational number then q_τ^c will mean $e^{2\pi i c\tau}$. Furthermore, given an element $r \in R$ (temporarily regarded as fixed) we put $z = a_1\tau + a_2$ and $q_z = e^{2\pi i z}$, where $a = (a_1, a_2) \in p^{-1}\mathbb{Z}^2$ is characterized by the properties $r = pa \bmod p\mathbb{Z}^2$ and $0 \leqslant a_1, a_2 < 1$. With these conventions we have

$$f_r(\tau) = q_\tau^{6a_1^2 - 6a_1 + 1} e^{12\pi i a_2(a_1 - 1)} (1 - q_z)^{12} \prod_{n \geqslant 1} (1 - q_\tau^n q_z)^{12} (1 - q_\tau^n / q_z)^{12}$$

[3, Formula K4, p. 29]. Let $\zeta = e^{2\pi i/p}$ and write $a_l = b_l/p$ $(l = 1, 2)$, so that b_l is the integer which coincides modulo p with r_l and satisfies $0 \leqslant b_l \leqslant p - 1$. Setting

$$f_r^{\mathrm{I}} = q_\tau^{6a_1^2 - 6a_1 + 1} e^{12\pi i b_2(a_1 - 1)/p} \tag{15}$$

and

$$f_r^{\mathrm{II}} = (1-q_\tau^{a_1}\zeta^{r_2})^{12}\prod_{n\geqslant 1}(1-q_\tau^{n+a_1}\zeta^{r_2})^{12}(1-q_\tau^{n-a_1}\zeta^{-r_2})^{12}, \tag{16}$$

we have $f_r = f_r^{\mathrm{I}} f_r^{\mathrm{II}}$.

Given $x\in\mathbb{F}_p$, let $\omega(x)$ be the unique even integer defined by the requirements $|\omega(x)|\leqslant p-1$ and $x=\omega(x) \bmod p$. (Thus in contrast to our notation in Section 2, ω is no longer the Teichmüller character but rather a $\mathbb{Z}$-valued approximation to the Teichmüller character.) Since $\omega(-x)=-\omega(x)$, the function $m: R\to\mathbb{Z}$ given by

$$m(r)=\omega(r_1)^2$$

is even and therefore belongs to M. In fact since m reduces mod p to the function $r\mapsto r_1^2$ we see that $m\in N$. Put $f=f^m$. Since $\omega(0)=0$ we can write

$$f=\prod_{r\in R} f_r^{\omega(r_1)^2} = \prod_{r_1\neq 0}\left(\prod_{r_2\in\mathbb{F}_p} f_r\right)^{\omega(r_1)^2}, \tag{17}$$

and if we define f^{I} and f^{II} by the analogous iterated product with f_r replaced by f_r^{I} and f_r^{II} respectively then $f=f^{\mathrm{I}}f^{\mathrm{II}}$. To prove Proposition 8 it will suffice to see that f^{I} is an integral power of q_τ and that f^{II} is a prime-to-p power of $(1-q_\tau)$ times an infinite product of factors of the form $(1-q_\tau^n)$ with integers $n\geqslant 2$. Indeed these assertions imply first of all that $f\in K^{\Gamma_1(p)}$, whence $f\in U\cap K^{\Gamma_1(p)}$ and in particular $f\in\mathcal{O}$. Thus we can evaluate $\varphi(f)$ by formally replacing q_τ with q in the q_τ-expansion of f. The result is an integral power of q times a prime-to-p power of $(1-q)$ times an infinite product of factors of the form $(1-q^n)$ with integers $n\geqslant 2$. Proposition 8 follows.

It remains to see that f^{I} and f^{II} do have the required form. First consider f^{I}. For $r_1\neq 0$ a routine calculation using (15) gives

$$\prod_{r_2\in\mathbb{F}_p} f_r^{\mathrm{I}} = q_\tau^{p(6a_1^2-6a_1+1)}\zeta^{-3r_1}.$$

Hence the definition of f^{I} gives

$$f^{\mathrm{I}} = q_\tau^{\Sigma_1}\zeta^{\Sigma_2}$$

with $\Sigma_1 = p\sum_{r_1\neq 0}\omega(r_1)^2(6a_1^2-6a_1+1)$ and $\Sigma_2=-3\sum_{r_1\neq 0}r_1^3$. The latter sum is 0 because $x\mapsto x^3$ is a nontrivial character $\mathbb{F}_p^\times\to\mathbb{F}_p^\times$. As for Σ_1, all we have to verify is its integrality, and since $p(-6a_1+1)$ is integral as it stands we need only check that $\sum_{r_1\neq 0} r_1^4=0$. But $p>5$, so again we have the sum of the values of a nontrivial character of $\mathbb{F}_p^\times$. Thus f^{I} has the required form.

Next we turn to f^{II}. Using (16) one readily computes that if $r_1 \neq 0$ then

$$\prod_{r_2 \in \mathbb{F}_p} f_r^{\mathrm{II}} = (1 - q_\tau^{b_1})^{12} \prod_{n \geqslant 1} (1 - q_\tau^{pn+b_1})^{12} (1 - q_\tau^{pn-b_1})^{12}. \tag{18}$$

Let us write $[\cdots]$ to denote a product of factors of the form $(1 - q_\tau^{\mu})$ with $\mu \geqslant 2$. If $r_1 = 1$ (or what amounts to the same, $b_1 = 1$) or if $r_1 = -1$ (i.e. $b_1 = p - 1$) then the right-hand side of (18) has the form $(1 - q_\tau)^{12}[\cdots]$. In all other cases the right-hand side of (18) has the form $[\cdots]$. Thus the definition of f^{II} gives

$$f^{\mathrm{II}} = (1 - q_\tau)^{12(\omega(1)^2 + \omega(-1)^2)}[\cdots].$$

But $\omega(\pm 1) = \mp(p-1)$, so we find that $f^{\mathrm{II}} = (1 - q_\tau)^{24(p-1)^2}[\cdots]$, which is of the required form.

Acknowledgments

I thank John Tate, Álvaro Lozano Robledo, and the referee for their comments.

References

[1] G. Boeckle, Deformations and the rigidity method, preprint.
[2] N. Boston, Appendix to [4], Compos. Math. 59 (1986) 261–264.
[3] D.S. Kubert, S. Lang, Modular Units, Grundlehren Mathematical Wissen, Vol. 244, Springer, New York, 1981.
[4] B. Mazur, A. Wiles, On p-adic analytic families of Galois representations, Compos. Math. 59 (1986) 231–264.
[5] D.E. Rohrlich, False division towers of elliptic curves, J. Algebra 229 (2000) 249–279.
[6] D.E. Rohrlich, A deformation of the Tate module, J. Algebra 229 (2000) 280–313.
[7] H. Sah, Automorphisms of finite groups, J. Algebra 10 (1968) 47–68.
[8] J.-P. Serre, Abelian ℓ-adic Representations and Elliptic Curves, McGill University, Lecture Notes written with the Collaboration of W. Kuyk, J. Labute, W.A. Benjamin, New York, 1968.
[9] J.-P. Serre, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, Invent. Math. 15 (1972) 259–331.
[10] J.-P. Serre, in: M. Brown (Ed.), Lectures on the Mordell–Weil theorem, Transl., from notes by Michel Waldschmidt, 3rd Edition, Aspects of Mathematics, Vol. E15, Fridr, Viehweg & Sohn, Braunschweig/Wiesbaden, 1997.
[11] J. Tate, A review of non-archimedean elliptic functions, in: J. Coates, S.T. Yau (Eds.), Elliptic curves, Modular Forms and Fermat's Last Theorem, International Press, Cambridge, MA, 1995, pp. 162–184.