MODULAR CURVES, HECKE CORRESPONDENCES, AND L-FUNCTIONS

DAVID E. ROHRLICH

In memory of my father
George F. Rohrlich
January 6, 1914 - August 21, 1995

These notes on Eichler-Shimura theory are intended for a reader who is familiar with elliptic curves and perhaps slightly acquainted with modular forms. The primary sources are [8], [19], and [20]. I am deeply indebted to Jaap Top for taking my place at the conference on very short notice and to Glenn Stevens for making the necessary arrangements with tact and understanding. I am also grateful to both of them for a careful reading of the text and for several comments which improved the final version.

1. MODULAR CURVES

Throughout, the term "curve" will mean "absolutely irreducible variety of dimension one." If $\Bbbk$ is a field, then $\Bbbk(t)$ denotes the field of rational functions over $\Bbbk$.

1.1. The modular curve $X_0(N)$. Let N be a positive integer. The modular curve $X_0(N)$ may be defined as follows. First choose an elliptic curve E over $\mathbb{Q}(t)$ such that $j(E) = t$. Then choose a point of order N on E and let C be the cyclic group which it generates. The subfield of $\overline{\mathbb{Q}(t)}$ fixed by the group

$$\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t)) : \sigma(C) = C\}$$

is a finite extension K of $\mathbb{Q}(t)$, and it turns out that K contains no proper algebraic extension of $\mathbb{Q}$: in other words, if we think of $\overline{\mathbb{Q}}$ as the algebraic closure of $\mathbb{Q}$ *inside* an algebraic closure of K, then $\overline{\mathbb{Q}} \cap K = \mathbb{Q}$. It follows that K is the function field of a smooth projective curve over $\mathbb{Q}$; this is $X_0(N)$.

The simplest nonvacuous example is the case $N = 2$. Let us choose E to be the curve

$$y^2 = 4x^3 - \frac{27t}{t-1728}x - \frac{27t}{t-1728},$$

Partially supported by NSF grant DMS-9396090

G. Cornell et al. (eds.), *Modular Forms and Fermat's Last Theorem*

so that K is the extension of $\mathbb{Q}(t)$ generated by a root of the equation

$$x^3 - \frac{27t}{4(t-1728)}x - \frac{27t}{4(t-1728)} = 0.$$

Viewed as a cubic in x, the left-hand side is an Eisenstein polynomial at the place $t = 0$ of $\mathbb{Q}(t)$ with discriminant $2^2 3^{12} t^2/(t-1728)^3 \notin \mathbb{Q}(t)^{\times 2}$. Therefore K is a nonnormal cubic extension of $\mathbb{Q}(t)$. We also see that the place $t = 0$ is totally ramified in K, while the places $t = 1728$ and $t = \infty$ each split into two places, one ramified of degree 2 and the other unramified. A calculation using the Hurwitz genus formula then shows that the genus of $X_0(2)$ is 0. By itself, this says little, because over $\mathbb{Q}$ there are infinitely many mutually nonisomorphic smooth projective curves of genus 0. However, it is easy to see that $X_0(2)$ has a rational point: for example, observe that at either place of K above $t = \infty$, the residue class field is $\mathbb{Q}$. It follows that $X_0(2)$ is isomorphic to $\mathbf{P}^1$ over $\mathbb{Q}$.

Returning to the general case, we must still verify that $\overline{\mathbb{Q}} \cap K = \mathbb{Q}$ and that up to isomorphism K is independent of the choice of E and C. The verification will ultimately lead us to modular functions.

We begin with some notation and conventions. Let $\Bbbk$ be a field of characteristic not dividing N. Given a Galois extension $\Bbbk'$ of $\Bbbk$ containing the group μ_N of N-th roots of unity, we shall write $\kappa : \mathrm{Gal}(\Bbbk'/\Bbbk) \to (\mathbb{Z}/N\mathbb{Z})^\times$ for the character giving the action of $\mathrm{Gal}(\Bbbk'/\Bbbk)$ on μ_N:

$$\sigma(\zeta) = \zeta^{\kappa(\sigma)} \qquad (\sigma \in \mathrm{Gal}(\Bbbk'/\Bbbk), \zeta \in \mu_N).$$

Suppose now that E is an elliptic curve over $\Bbbk$. Let $E[N] \subset E(\overline{\Bbbk})$ denote the subgroup of points of order dividing N, and write $\Bbbk(E[N])$ for the finite Galois extension of $\Bbbk$ generated by the coordinates relative to some generalized Weierstrass equation for E over $\Bbbk$ of the affine points on E of order dividing N. After fixing an ordered basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, we may identify the natural embedding of $\mathrm{Gal}(\Bbbk(E[N])/\Bbbk)$ in $\mathrm{Aut}(E[N])$ with a faithful representation

$$\rho : \mathrm{Gal}(\Bbbk(E[N])/\Bbbk) \hookrightarrow \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z}).$$

The formalism of the Weil pairing shows that $\Bbbk(E[N])$ contains μ_N and that the determinant of ρ is κ. In particular, if $\Bbbk$ itself contains μ_N, then κ is trivial and ρ is a representation

$$\mathrm{Gal}(\Bbbk(E[N])/\Bbbk) \hookrightarrow \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z}).$$

Theorem 1. *If E is an elliptic curve over $\mathbb{C}(t)$ with $j(E) = t$, then the representation on N-division points is an isomorphism*

$$\mathrm{Gal}(\mathbb{C}(t, E[N])/\mathbb{C}(t)) \cong \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z}).$$

Theorem 1 will be proved later. For now we derive consequences. The first consequence is that if E is an elliptic curve over $\mathbb{Q}(\mu_N)(t)$ with invariant t, then it is still true that the representation on N-division points is an isomorphism

$$\mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(\mu_N)(t)) \cong \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z}).$$

For we know that the left-hand side is *embedded* in the right-hand side, but on field-theoretic grounds we also have

$$[\mathbb{Q}(t, E[N]) : \mathbb{Q}(\mu_N)(t)] \geq [\mathbb{C}(t, E[N]) : \mathbb{C}(t)].$$

Hence the conclusion follows from Theorem 1. Next suppose that E is an elliptic curve over $\mathbb{Q}(t)$ with invariant t. Then E can be viewed as an elliptic curve over $\mathbb{Q}(\mu_N)(t)$, and consequently the representation on N-division points

$$\rho : \mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(t)) \hookrightarrow \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$$

sends $\mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(\mu_N)(t))$ onto $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$. Since $\det \rho = \kappa$, the image of ρ also contains a set of coset representatives for $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$ in $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$. Therefore the image of ρ is all of $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$, and we obtain the first part of the following assertion:

Corollary. *If E is an elliptic curve over $\mathbb{Q}(t)$ with invariant t, then the representation on N-division points is an isomorphism*

$$\mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(t)) \cong \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z}),$$

and $\overline{\mathbb{Q}} \cap \mathbb{Q}(t, E[N]) = \mathbb{Q}(\mu_N)$.

The equation $\overline{\mathbb{Q}} \cap \mathbb{Q}(t, E[N]) = \mathbb{Q}(\mu_N)$ also follows from the argument just given. Indeed, put $L = \overline{\mathbb{Q}} \cap \mathbb{Q}(t, E[N])$, and suppose that L is strictly larger than $\mathbb{Q}(\mu_N)$. Then $L(t)$ is strictly larger than $\mathbb{Q}(\mu_N)(t)$, and consequently

$$[\mathbb{Q}(t, E[N]) : L(t)] < |\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})|.$$

But as before, the left-hand side of this inequality is *a priori* greater than or equal to $[\mathbb{C}(t, E[N]) : \mathbb{C}(t)]$, and we have a contradiction.

We can now clarify some points in the definition of $X_0(N)$. Recall that the recipe for the function field K of $X_0(N)$ was as follows: Choose an elliptic curve E over $\mathbb{Q}(t)$ with $j(E) = t$, and then choose a cyclic subgroup C of $E[N]$; write G for the Galois group of $\mathbb{Q}(t, E[N])$ over $\mathbb{Q}(t)$ and H for the subgroup of G preserving C; then K is the fixed field of H. Now any point of order N on E is the second basis vector in some ordered basis for

$E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, and consequently we may identify G with $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ in such a way that H is identified with the lower triangular group

$$\left\{ \begin{pmatrix} a & 0 \\ b & d \end{pmatrix} : a, d \in (\mathbb{Z}/N\mathbb{Z})^\times, b \in \mathbb{Z}/N\mathbb{Z} \right\}.$$

Since the determinant maps the lower triangular group *onto* $(\mathbb{Z}/N\mathbb{Z})^\times$, we have $\mathbb{Q}(\mu_N) \cap K = \mathbb{Q}$. Substituting $\mathbb{Q}(\mu_N) = \overline{\mathbb{Q}} \cap \mathbb{Q}(t, E[N])$ in this equation, we obtain $\overline{\mathbb{Q}} \cap K = \mathbb{Q}$, as claimed earlier. We also see that up to isomorphism over $\mathbb{Q}(t)$, the field K is independent of the choice of C: for if we change the basis of $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, then we conjugate H inside G, and hence we conjugate (in the field-theoretic sense) K inside $\mathbb{Q}(t, E[N])$. It remains to examine the dependence of K on E. Quite generally, if E is an elliptic curve over a field $\Bbbk$ of characteristic not dividing N, let $\Bbbk(E[N]/\pm)$ denote the extension of $\Bbbk$ generated by the x-coordinates of the affine points on E of order dividing N. Then $\Bbbk(E[N]/\pm)$ is the fixed field of

$$\{\sigma \in \mathrm{Gal}(\Bbbk(E[N])/\Bbbk) : \sigma(P) = \pm P \text{ for every } P \in E[N]\}.$$

Returning to the case at hand, we can say that $\mathbb{Q}(t, E[N]/\pm)$ is the subfield of $\mathbb{Q}(t, E[N])$ corresponding to the subgroup $\{\pm I\}$ of $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$; thus

$$\mathrm{Gal}(\mathbb{Q}(t, E[N]/\pm)/\mathbb{Q}(t)) \cong \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}.$$

Suppose now that E' is another elliptic curve over $\mathbb{Q}(t)$ with $j(E') = t$. Then E and E' differ by a quadratic twist, and consequently so do the associated representations

$$\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t)) \longrightarrow \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$$

provided that bases for $E[N]$ and $E'[N]$ are chosen compatibly. It follows that the fields $\mathbb{Q}(t, E[N]/\pm)$ and $\mathbb{Q}(t, E'[N]/\pm)$ are equal and that the associated isomorphisms from

$$\mathrm{Gal}(\mathbb{Q}(t, E[N]/\pm)/\mathbb{Q}(t)) \quad \text{and} \quad \mathrm{Gal}(\mathbb{Q}(t, E'[N]/\pm)/\mathbb{Q}(t))$$

to $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$ are identical. Now the lower triangular subgroup of $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ contains $\{\pm I\}$, so that K is a subfield of $\mathbb{Q}(t, E[N]/\pm)$. Hence K can be characterized as the subfield of $\mathbb{Q}(t, E[N]/\pm)$ fixed by the image of the lower triangular subgroup in $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$, and this characterization is independent of the choice of E.

1.2. Other modular curves. More generally, let H be any subgroup of $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ satisfying two conditions:

(i) $-I \in H$.
(ii) The determinant $H \longrightarrow (\mathbb{Z}/N\mathbb{Z})^\times$ is surjective.

If K is the subfield of $\mathbb{Q}(t, E[N])$ fixed by H, then the same argument shows that K is the function field of a smooth projective curve $X(H)$ over $\mathbb{Q}$ which up to isomorphism is independent of the choice of E. As we have just seen, $X_0(N)$ corresponds to the choice

$$H = \left\{ \begin{pmatrix} a & 0 \\ b & d \end{pmatrix} : a, d \in (\mathbb{Z}/N\mathbb{Z})^\times, \ \ b \in \mathbb{Z}/N\mathbb{Z} \right\};$$

another example is the curve $X_1(N)$ corresponding to

$$H = \left\{ \begin{pmatrix} a & 0 \\ b & \pm 1 \end{pmatrix} : a \in (\mathbb{Z}/N\mathbb{Z})^\times, \ \ b \in \mathbb{Z}/N\mathbb{Z} \right\}.$$

These two examples will be the primary focus throughout, and while we shall initially emphasize $X_0(N)$, it will ultimately be $X_1(N)$ which provides the broader context for the application to L-functions. We note in passing that if P is a point of order N on E and C is the subgroup which it generates, then the function field of $X_1(N)$ can be identified with the subfield of $\overline{\mathbb{Q}(t)}$ fixed by $\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t)) : \sigma(P) = \pm P\}$, just as the function field of $X_0(N)$ coincides with the subfield of $\overline{\mathbb{Q}(t)}$ fixed by $\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t)) : \sigma(C) = C\}$.

By construction, a modular curve $X(H)$ with function field K comes equipped with a distinguished morphism to $\mathbf{P}^1$ over $\mathbb{Q}$, namely the morphism corresponding to the inclusion of $\mathbb{Q}(t)$ in K. It is conventional to refer to the finite set of points on $X(H)$ lying over the point $t = \infty$ of $\mathbf{P}^1$ as *cusps*. If we remove the cusps from $X(H)$, then we obtain an affine curve $Y(H)$, which is usually denoted $Y_0(N)$ in the case of $X_0(N)$ and $Y_1(N)$ in the case of $X_1(N)$. To illustrate the notion of a cusp, let us prove that $X_0(N)$ has at least one cusp rational over $\mathbb{Q}$. Choose an elliptic curve E over $\mathbb{Q}(t)$ with invariant t and split multiplicative reduction at the place $t = \infty$ of $\mathbb{Q}(t)$. For example, E could be the curve

$$y^2 + xy = x^3 - \frac{36}{t - 1728}x - \frac{1}{t - 1728},$$

because the covariants c_4 and c_6 of this equation satisfy $-c_4/c_6 = 1$. Denote the place $t = \infty$ of $\mathbb{Q}(t)$ simply by ∞, identify the completion of $\mathbb{Q}(t)$ at ∞ with $\mathbb{Q}((1/t))$, and pick an extension of ∞ to a place of $\overline{\mathbb{Q}(t)}$, so that the corresponding decomposition subgroup $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty$ of $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))$ is identified with the Galois group of $\overline{\mathbb{Q}((1/t))}$ over $\mathbb{Q}((1/t))$. By the theory of Tate curves, there is an isomorphism of $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty$-modules

$$E\big(\overline{\mathbb{Q}((1/t))}\big) \cong \overline{\mathbb{Q}((1/t))}^\times / q^{\mathbb{Z}},$$

where $q \in \mathbb{Q}((1/t))$ is a uniformizer and $q^{\mathbb{Z}}$ denotes the infinite cyclic group generated by q. It follows in particular that

$$E[N] \cong \mu_N \oplus (q^{1/N})^{\mathbb{Z}}/q^{\mathbb{Z}}$$

as $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty$-modules. Let $\{P_1, P_2\}$ be the basis for $E[N]$ which corresponds under the preceding isomorphism to the basis $\{q^{1/N} \bmod q^{\mathbb{Z}}, \zeta\}$, where ζ is a generator of μ_N. Relative to this basis, the action of the group $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty$ on $E[N]$ is represented by matrices of the form

$$\begin{pmatrix} 1 & 0 \\ * & \kappa(g) \end{pmatrix} \qquad (g \in \mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty).$$

In particular, the cyclic group C generated by P_2 is preserved by the decomposition group $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))_\infty$, whence the latter is contained in the subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}(t)}/\mathbb{Q}(t))$ fixing the function field of $X_0(N)$. Thus the residue class degree of the restriction of ∞ to this function field is 1, and so the field of rationality of the corresponding cusp is $\mathbb{Q}$.

1.3. Moduli interpretation of $X_0(N)$ and $X_1(N)$. Let $\Bbbk$ be an algebraically closed field, and consider pairs $(\mathcal{E}, \mathcal{C})$ consisting of an elliptic curve $\mathcal{E}$ over $\Bbbk$ and a cyclic subgroup $\mathcal{C} \subset \mathcal{E}[N]$ of order N. An isomorphism from a pair $(\mathcal{E}_1, \mathcal{C}_1)$ to a pair $(\mathcal{E}_2, \mathcal{C}_2)$ is an isomorphism from $\mathcal{E}_1$ to $\mathcal{E}_2$ sending $\mathcal{C}_1$ to $\mathcal{C}_2$. We write $[\mathcal{E}, \mathcal{C}]$ for the isomorphism class containing $(\mathcal{E}, \mathcal{C})$ and $\mathrm{Ell}_0(N)(\Bbbk)$ for the set of all isomorphism classes. Also, if S is any subset of $\mathbf{P}^1(\Bbbk)$, then $\mathrm{Ell}_0(N)(\Bbbk)_S$ denotes the set of all isomorphism classes $[\mathcal{E}, \mathcal{C}] \in \mathrm{Ell}_0(N)(\mathbb{C})$ such that $j(\mathcal{E}) \notin S$.

One can also consider pairs of the form $(\mathcal{E}, \mathcal{P})$, where $\mathcal{P}$ is a point of order N on $\mathcal{E}$. An isomorphism from $(\mathcal{E}_1, \mathcal{P}_1)$ to $(\mathcal{E}_2, \mathcal{P}_2)$ is an isomorphism from $\mathcal{E}_1$ to $\mathcal{E}_2$ sending $\mathcal{P}_1$ to $\mathcal{P}_2$. Note the equality of isomorphism classes $[\mathcal{E}, \mathcal{P}] = [\mathcal{E}, -\mathcal{P}]$. We write $\mathrm{Ell}_1(N)(\Bbbk)$ for the set of isomorphism classes and $\mathrm{Ell}_1(N)(\Bbbk)_S$ for the subset consisting of those $[\mathcal{E}, \mathcal{P}]$ such that $j(\mathcal{E}) \notin S$.

Next observe that if X is a modular curve and E is an elliptic curve over $\mathbb{Q}(t)$ with invariant t, then E can be viewed as an elliptic curve over the function field of X, because the latter field is naturally an extension of $\mathbb{Q}(t)$. In particular, since a point $x \in X(\mathbb{C})$ determines a discrete valuation ring $\mathcal{O}_x$ of the complex function field of X, one can ask whether E has good reduction at the maximal ideal $\mathfrak{m}_x$ of $\mathcal{O}_x$: if so, then reduction modulo $\mathfrak{m}_x$ yields an elliptic curve E_x over $\mathbb{C}$. After extending $\mathcal{O}_x$ to a discrete valuation ring of $\mathbb{C}(t, E[N])$ in some way, we can also consider the reduction map on N-division points $E[N] \longrightarrow E_x[N]$, which is injective. Thus if P is a point of order N on E and C is the cyclic subgroup which it generates, then the reductions P_x and C_x are defined, and are still of order N.

Given a subset S of $\mathbf{P}^1(\mathbb{C})$, let $\mathbf{P}^1(\mathbb{C})_S$ denote the complement of S in $\mathbf{P}^1(\mathbb{C})$ and $X(\mathbb{C})_S$ the inverse image of $\mathbf{P}^1(\mathbb{C})_S$ in $X(\mathbb{C})$. For example, if $S = \{\infty\}$, then $X_0(N)(\mathbb{C})_S = Y_0(N)(\mathbb{C})$.

Proposition 1. *Let E be an elliptic curve over $\mathbb{Q}(t)$ with $j(E) = t$, and let S be a subset of $\mathbf{P}^1(\mathbb{C})$ containing all places where E has bad reduction. Fix an ordered basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, let P be the second element of this basis, and let C be the cyclic group of order N generated by P. Then the*

map $x \mapsto [E_x, C_x]$ *defines a bijection of* $X_0(N)(\mathbb{C})_S$ *onto* $\mathrm{Ell}_0(N)(\mathbb{C})_S$. *The same is true if* $X_0(N)$, $\mathrm{Ell}_0(N)$, *and* C *are replaced by* $X_1(N)$, $\mathrm{Ell}_1(N)$, *and* P *respectively.*

Proof. First of all, S necessarily contains 0, 1728, and ∞. Indeed E differs from the curve

$$y^2 + xy = x^3 - \frac{36}{t - 1728}x - \frac{1}{t - 1728}$$

by a quadratic twist over $\mathbb{Q}(t)$, and therefore the discriminant of any equation for E differs from the discriminant of the above equation by a sixth power in $\mathbb{Q}(t)^\times$. Since the above equation has discriminant $t^2/(t - 1728)^3$, it follows that E has bad reduction at 0, 1728, and ∞.

As usual, we identify $\mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(t))$ with $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$, and similarly $\mathrm{Gal}(\mathbb{C}(t, E[N])/\mathbb{C}(t))$ with $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$. Let $K \subset \mathbb{Q}(t, E[N])$ be the subfield corresponding to

$$H = \left\{ \begin{pmatrix} a & 0 \\ b & d \end{pmatrix} : a, d \in (\mathbb{Z}/N\mathbb{Z})^\times, b \in \mathbb{Z}/N\mathbb{Z} \right\},$$

so that $\mathbb{C}K$ is the function field of $X_0(N)$ over $\mathbb{C}$. Also let X be a curve with complex function field $\mathbb{C}(t, E[N])$ and $\pi : X \longrightarrow X_0(N)$ the corresponding morphism. Given a point $t_0 \in \mathbf{P}^1(\mathbb{C})_S$, a point $x_0 \in X_0(N)(\mathbb{C})$ lying over t_0, and a point $\hat{x}_0 \in X(\mathbb{C})$ lying over x_0, we have a bijection

$$\begin{aligned} \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/(H \cap \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})) &\longrightarrow \{\text{fiber of } X_0(N)(\mathbb{C}) \text{ over } t_0\} \\ g(H \cap \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})) &\longmapsto \pi(g\hat{x}_0), \end{aligned} \tag{1}$$

because the extension $\mathbb{C}(t, E[N])/\mathbb{C}(t)$ is unramified outside the places of $\mathbb{C}(t)$ where E has bad reduction, and hence in particular outside S. On the other hand, the maps

$$\begin{aligned} \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/(H \cap \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})) &\longrightarrow \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})/H \\ g(H \cap \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})) &\longmapsto gH \end{aligned} \tag{2}$$

and

$$\begin{aligned} \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})/H &\longrightarrow \{\text{cyclic subgroups of } E \text{ of order } N\} \\ gH &\longmapsto gC \end{aligned} \tag{3}$$

are also bijections, as is the map

$$\begin{aligned} &\{\text{cyclic subgroups of } E \text{ of order } N\} \\ &\qquad \longrightarrow \{\text{cyclic subgroups of } E_{x_0} \text{ of order } N\} \end{aligned} \tag{4}$$

afforded by reduction modulo $\mathfrak{m}_{x_0}$. Now if $x = \pi(g\hat{x}_0)$, then $E_x = E_{x_0}$ and $(gC)_{x_0} = C_x$ (note that these identifications are meaningful since the residue class field $\mathbb{C}$ is a subring of $\mathcal{O}_{x_0}$ and $\mathcal{O}_x$). Therefore composing the inverse of (1) with (2), (3), and (4), we see that the map

$$\begin{aligned}\{\text{fiber of } X_0(N)(\mathbb{C}) \text{ over } t_0\} &\longrightarrow \{\text{cyclic subgroups of } E_{x_0} \text{ of order } N\}\\ x &\longmapsto C_x\end{aligned}$$

is a bijection. But $j(E_{x_0}) = t_0 \neq 0, 1728$, whence $\mathrm{Aut}(E_{x_0}) = \{\pm 1\}$. Thus an automorphism of E_{x_0} sends each cyclic subgroup of E_{x_0} to itself. Consequently

$$\begin{aligned}\left\{\begin{matrix}\text{cyclic subgroups of } E_{x_0}\\ \text{of order } N\end{matrix}\right\} &\longrightarrow \{[\mathcal{E}, \mathcal{C}] \in \mathrm{Ell}_0(N)(\mathbb{C})_S : j(\mathcal{E}) = t_0\},\\ \mathcal{C} &\longmapsto [E_{x_0}, \mathcal{C}]\end{aligned}$$

is also a bijection, and composing this map with the previous one we see that $x \mapsto [E_x, C_x]$ is a bijection from the fiber of $X_0(N)(\mathbb{C})$ over $t_0 \in \mathbf{P}^1(\mathbb{C})_S$ to the subset of $\mathrm{Ell}_0(N)(\mathbb{C})_S$ consisting of isomorphism classes with j-invariant t_0.

The argument for $X_1(N)$ is similar. The only additional wrinkle is that P is not actually defined over the function field of $X_1(N)(\mathbb{C})$, but only over a quadratic extension. Thus P_x depends on a choice of a valuation ring lying over $\mathcal{O}_x$. However, if we make the alternate choice then we simply replace P_x by $-P_x$. Since $[E_x, P_x] = [E_x, -P_x]$, the map $x \mapsto [E_x, P_x]$ is still well defined.

1.4. Proof of Theorem 1: a preliminary reduction. We claim that to prove Theorem 1 it suffices to prove the following:

Proposition 2. *There exists an elliptic curve E over $\mathbb{C}(t)$ with invariant t such that*

$$[\mathbb{C}(t, E[N]/\pm) : \mathbb{C}(t)] = |\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}|.$$

Granting Proposition 2, let us see how to deduce Theorem 1. Suppose that E' is any elliptic curve over $\mathbb{Q}(t)$ with invariant t. Then E' differs from E by at most a quadratic twist, so that $\mathbb{C}(t, E[N]/\pm) = \mathbb{C}(t, E'[N]/\pm)$. Consequently

$$[\mathbb{C}(t, E'[N]/\pm) : \mathbb{C}(t)] = [\mathbb{C}(t, E[N]/\pm) : \mathbb{C}(t)] = |\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}|,$$

and the embedding $\mathrm{Gal}(\mathbb{C}(t, E'[N]/\pm)/\mathbb{C}(t)) \longrightarrow \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$ is an isomorphism. Hence the image of the representation

$$\mathrm{Gal}(\mathbb{C}(t, E'[N])/\mathbb{C}(t)) \longrightarrow \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$$

contains a set of coset representatives for $\{\pm I\}$ in $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$ and so in particular contains either the matrix

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

or its negative. Since the square of this matrix is $-I$, we find that the image is all of $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$, and Theorem 1 follows.

1.5. Modular functions. It remains to prove Proposition 2. The argument will use the theory of modular functions. We begin by recalling the relevant definitions.

Let $\mathfrak{H}$ denote the complex upper half-plane, and let $\mathrm{GL}^+(2, \mathbb{R})$ denote the subgroup of $\mathrm{GL}(2, \mathbb{R})$ consisting of matrices with positive determinant. We consider the usual action of $\mathrm{GL}^+(2, \mathbb{R})$ on $\mathfrak{H}$ by fractional linear transformations: for

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}^+(2, \mathbb{R})$$

and $z \in \mathfrak{H}$ put

$$\gamma z = \frac{az + b}{cz + d}.$$

If f is a function on $\mathfrak{H}$ we write $f \circ \gamma$ for the function $z \mapsto f(\gamma z)$. Now suppose that Γ is a subgroup of finite index in $\mathrm{SL}(2, \mathbb{Z})$. A modular function for Γ is a meromorphic function f on $\mathfrak{H}$ satisfying two conditions:

(i) $f \circ \gamma = f$ for $\gamma \in \Gamma$.

(ii) Given $\delta \in \mathrm{SL}(2, \mathbb{Z})$, let M be a positive integer such that

$$(f \circ \delta)(z + M) = (f \circ \delta)(z).$$

(Such an integer exists by (i), because Γ has finite index in $\mathrm{SL}(2, \mathbb{Z})$ and consequently some power of the matrix $\delta \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \delta^{-1}$ belongs to Γ.) Let F be the meromorphic function on the punctured unit disk $D^\circ = \{q \in \mathbb{C} : 0 < |q| < 1\}$ defined by

$$f(\delta z) = F(e^{2\pi i z/M}) \qquad (z \in \mathfrak{H}).$$

Then F extends to a meromorphic function on the full unit disk $D = \{q \in \mathbb{C} : |q| < 1\}$.

Like any meromorphic function on the punctured disk, the function F in (ii) is represented by a Laurent series

$$F(q) = \sum_{n \in \mathbb{Z}} a(n) q^n$$

for q near 0, and the content of (ii) is that F does not have an essential singularity at $q = 0$. Thus (ii) is equivalent to the condition that for every $\delta \in \mathrm{SL}(2,\mathbb{Z})$, the function $f \circ \delta$ has a Fourier series expansion of the form

$$f(\delta z) = \sum_{n \geq n_0} a(n) e^{2\pi i n z/M} \qquad (\mathrm{Im}(z) \gg 0)$$

with $n_0 \in \mathbb{Z}$. Of course if $\Gamma = \mathrm{SL}(2,\mathbb{Z})$ then $f \circ \delta = f$ by (i), and to verify (ii) it suffices to check that f itself has such a Fourier expansion.

The usual operations of addition and multiplication of meromorphic functions make the set of modular functions for Γ into a field $\mathfrak{M}(\Gamma)$. For the proof of Proposition 2 we are interested in the cases where Γ is the group $\mathrm{SL}(2,\mathbb{Z})$ (also denoted $\Gamma(1)$) or the group $\Gamma(N)$, the kernel of the reduction-modulo-N map $\mathrm{SL}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})$. The quotient group

$$\Gamma(1)/\{\pm I\}\Gamma(N) \cong \mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$$

acts as a group of automorphisms of $\mathfrak{M}(\Gamma(N))$ with fixed field $\mathfrak{M}(\Gamma(1))$, and therefore $\mathfrak{M}(\Gamma(N))$ is a Galois extension of $\mathfrak{M}(\Gamma(1))$ with Galois group isomorphic to $\mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$. For the record, we make an explicit identification

$$\theta : \mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})/\{\pm I\} \longrightarrow \mathrm{Gal}(\mathfrak{M}(\Gamma(N))/\mathfrak{M}(\Gamma(1)))$$

by declaring that if $\gamma \in \mathrm{SL}(2,\mathbb{Z})$ and $f \in \mathfrak{M}(\Gamma(N))$ then

$$\theta([\gamma])(f) = f \circ \gamma^{\mathrm{t}},$$

where $[\gamma]$ denotes the image of γ in $\mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$ and γ^{t} denotes the transpose of γ.

An essential point for the proof of Proposition 2 is that $\mathfrak{M}(\Gamma(1))$ is generated over $\mathbb{C}$ by a single nonconstant function, the j-function. Like the functions g_2 and g_3 appearing in the formula

$$j = 1728 \frac{g_2^3}{g_2^3 - 27 g_3^2}$$

which defines it, the j-function should be viewed in the first instance as a function of a lattice variable $\mathcal{L} \subset \mathbb{C}$. We obtain functions of $z \in \mathfrak{H}$ by putting $g_2(z) = g_2(\mathcal{L}_z)$, $g_3(z) = g_3(\mathcal{L}_z)$, and $j(z) = j(\mathcal{L}_z)$, where

$$\mathcal{L}_z = z\mathbb{Z} \oplus \mathbb{Z}.$$

Now as functions of lattices, g_2 and g_3 are defined by the formulas

$$g_2(\mathcal{L}) = 60 \sum_{\substack{\omega \in \mathcal{L} \\ \omega \neq 0}} \omega^{-4} \quad \text{and} \quad g_3(\mathcal{L}) = 140 \sum_{\substack{\omega \in \mathcal{L} \\ \omega \neq 0}} \omega^{-6},$$

which give at once the behavior of g_2 and g_3 under homothety:

$$g_2(\lambda\mathcal{L}) = \lambda^{-4}g_2(\mathcal{L}), \quad \text{and} \quad g_3(\lambda\mathcal{L}) = \lambda^{-6}g_3(\mathcal{L}) \quad \text{for all } \lambda \in \mathbb{C}^\times.$$

It follows that as a function of lattices, j is invariant under homothety. Furthermore, if $z \in \mathfrak{H}$ and $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$, then

$$\mathcal{L}_z = z\mathbb{Z} \oplus \mathbb{Z} = (az+b)\mathbb{Z} \oplus (cz+d)\mathbb{Z} = \lambda\mathcal{L}_{\gamma z}$$

with $\lambda = cz+d$, whence $j(\gamma z) = j(z)$. This is condition (i) in the definition of modular functions; to verify (ii) we write $z = x + iy$ and observe that

$$\lim_{y\to\infty} g_2(x+iy) = 120\sum_{n\geq 1} n^{-4} \quad \text{and} \quad \lim_{y\to\infty} g_3(x+iy) = 280\sum_{n\geq 1} n^{-6}$$

uniformly in x. Thus the holomorphic functions G_2 and G_3 on D° such that $g_2(z) = G_2(e^{2\pi iz})$ and $g_3(z) = G_3(e^{2\pi iz})$ extend holomorphically to D, and consequently the function $J = G_2^3/(G_2^3 - 27G_3^2)$ extends at least meromorphically to D. Hence j is a modular function for $\mathrm{SL}(2,\mathbb{Z})$. Now the calculation

$$(120\sum_{n\geq 1} n^{-4})^3 - 27(280\sum_{n\geq 1} n^{-6})^2 = (120\pi^4/90)^3 - 27(280\pi^6/945)^2 = 0$$

shows that $J(q)$ actually has a pole at $q = 0$, and a more thorough analysis reveals that the pole is simple with residue 1. Therefore j has an expansion of the form

$$j(z) = \frac{1}{q} + \text{ power series in } q$$

for $q = e^{2\pi iz}$ near 0. In fact the Fourier expansion of j holds for all q in the unit disk, i.e., for all $z \in \mathfrak{H}$, because j is holomorphic on $\mathfrak{H}$: indeed the properties of the Weierstrass $\wp$-function show that $g_2^3 - 27g_3^2$ is nowhere vanishing as a function of lattices and hence also as a function of $z \in \mathfrak{H}$. From the fact that j is holomorphic on $\mathfrak{H}$ with only a simple pole as a Laurent series in q, one deduces that for any $f \in \mathfrak{M}(\Gamma(1))$ there exist polynomials $P(t), Q(t) \in \mathbb{C}(t)$ with $P(t) \neq 0$ such that $P(j)f - Q(j)$ is holomorphic on $\mathfrak{H}$ and

$$\lim_{y\to\infty} P(j(z))f(z) - Q(j(z)) = 0$$

uniformly in x. An application of the maximum principle on a suitably truncated fundamental domain for $\mathrm{SL}(2,\mathbb{Z})\backslash\mathfrak{H}$ then gives $P(j)f = Q(j)$, whence $\mathfrak{M}(\Gamma(1)) = \mathbb{C}(j)$ as claimed.

1.6. Elliptic functions. To summarize, $\mathfrak{M}(\Gamma(1)) = \mathbb{C}(j)$, and $\mathfrak{M}(\Gamma(N))$ is a Galois extension of $\mathbb{C}(j)$ with Galois group

$$\mathrm{Gal}\big(\mathfrak{M}(\Gamma(N))/\mathfrak{M}(\Gamma(1))\big) \cong \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}.$$

Consider the elliptic curve

$$E : y^2 = 4x^3 - \frac{27j}{j-1728}x - \frac{27j}{j-1728}$$

over $\mathbb{C}(j)$. We will show that $\mathbb{C}(j, E[N]/\pm)$ coincides with $\mathfrak{M}(\Gamma(N))$ when both fields are viewed inside a fixed algebraic closure of $\mathbb{C}(j)$. Thsi will prove Proposition 2.

The additional ingredient needed at this point is the Weierstrass parametrization of elliptic curves over $\mathbb{C}$. Let $\mathcal{L}$ be a lattice in $\mathbb{C}$ and consider the elliptic curve

$$\mathcal{E}^{\mathrm{Wst}} : Y^2 = 4X^3 - g_2(\mathcal{L})X - g_3(\mathcal{L}).$$

We recall that the Weierstrass $\wp$-function

$$\wp(u; \mathcal{L}) = \frac{1}{u^2} + \sum_{\substack{\omega \in \mathcal{L} \\ \omega \neq 0}} \frac{1}{(u+\omega)^2} - \frac{1}{\omega^2}$$

affords a complex analytic group isomorphism

$$\begin{aligned} \mathbb{C}/\mathcal{L} &\longrightarrow \mathcal{E}^{\mathrm{Wst}}(\mathbb{C}) \\ u + \mathcal{L} &\longmapsto (\wp(u; \mathcal{L}), \wp'(u; \mathcal{L})), \end{aligned}$$

where $(\wp(u; \mathcal{L}), \wp'(u; \mathcal{L}))$ is to be interpreted as the point at infinity if $u \in \mathcal{L}$. For present purposes we must modify the classical normalization slightly. Assume that $j(\mathcal{L}) \neq 0, 1728$ and consider the elliptic curve

$$\mathcal{E} : y^2 = 4x^3 - \frac{27j(\mathcal{L})}{j(\mathcal{L})-1728}x - \frac{27j(\mathcal{L})}{j(\mathcal{L})-1728}.$$

Let $(g_2(\mathcal{L})/g_3(\mathcal{L}))^{3/2}$ denote a fixed square root of $(g_2(\mathcal{L})/g_3(\mathcal{L}))^3$. On rewriting the relation $j(\mathcal{L}) = 1728 g_2(\mathcal{L})^3/(g_2(\mathcal{L})^3 - 27 g_3(\mathcal{L})^2)$ in the form

$$\frac{g_2(\mathcal{L})^3}{g_3(\mathcal{L})^2} = \frac{27j(\mathcal{L})}{j(\mathcal{L})-1728},$$

we see that the change of variables

$$X = (g_3(\mathcal{L})/g_2(\mathcal{L}))x, \qquad Y = (g_3(\mathcal{L})/g_2(\mathcal{L}))^{3/2} y$$

transforms the equation for $\mathcal{E}$ into the equation for $\mathcal{E}^{\mathrm{Wst}}$. Thus we can replace the map $u + \mathcal{L} \mapsto (\wp(u;\mathcal{L}), \wp'(u;\mathcal{L}))$ by the map

$$u + \mathcal{L} \longmapsto \left(\frac{g_2(\mathcal{L})}{g_3(\mathcal{L})} \wp(u;\mathcal{L}), \left(\frac{g_2(\mathcal{L})}{g_3(\mathcal{L})} \right)^{3/2} \wp'(u;\mathcal{L}) \right)$$

to obtain a complex analytic group isomorphism of $\mathbb{C}/\mathcal{L}$ onto $\mathcal{E}(\mathbb{C})$. In particular, if we fix a basis $\{\omega_1, \omega_2\}$ for $\mathcal{L}$, then the numbers

$$x_{r,s}(\mathcal{L}) = \frac{g_2(\mathcal{L})}{g_3(\mathcal{L})} \wp\left(\frac{r\omega_1 + s\omega_2}{N}; \mathcal{L} \right) \qquad (r, s \in \mathbb{Z}, \ \ (r,s) \not\equiv (0,0) \bmod N)$$

are the x-coordinates of the affine N-division points on $\mathcal{E}$. Now as a function of u, $\wp(u;\mathcal{L})$ is periodic with respect to $\mathcal{L}$, even, and of degree 2 when viewed as a map $\mathbb{C}/\mathcal{L} \longrightarrow \mathbf{P}^1(\mathbb{C})$. Therefore

$$x_{r,s}(\mathcal{L}) = x_{r',s'}(\mathcal{L}) \Longleftrightarrow (r,s) \equiv \pm(r',s') \bmod N.$$

Letting $\mathcal{R}$ denote the set of orbits of $(\mathbb{Z}/N\mathbb{Z})^2 - \{(0,0)\}$ under the negation map, we see that if (r,s) runs over a set of representatives in $\mathbb{Z}^2$ for the distinct elements of $\mathcal{R}$, then the numbers $x_{r,s}(\mathcal{L})$ are distinct.

Now let $P(w; A, B) \in \mathbb{Z}[w, A, B]$ be the N-th division polynomial, a universal polynomial with the property that $P(w_0; A, B) = 0$ if and only if w_0 is the x-coordinate of an affine N-division point on the elliptic curve $y^2 = 4x^3 + Ax + B$. Applying this property to the elliptic curve $\mathcal{E}$, we find that

$$P\left(x_{r,s}(\mathcal{L}); \frac{27j(\mathcal{L})}{j(\mathcal{L}) - 1728}, \frac{27j(\mathcal{L})}{j(\mathcal{L}) - 1728} \right) = 0$$

whenever $j(\mathcal{L}) \neq 0, 1728$. In particular, let us take $\mathcal{L} = \mathcal{L}_z$, where $j(z) \neq 0, 1728$. Setting

$$f_{r,s}(z) = \frac{g_2(z)}{g_3(z)} \wp\left(\frac{r + sz}{N}; \mathcal{L}_z \right) \qquad (r, s \in \mathbb{Z}, \ \ (r,s) \not\equiv (0,0) \bmod N),$$

we have $f_{r,s}(z) = x_{r,s}(\mathcal{L}_z)$ and consequently

$$P\left(f_{r,s}(z); \frac{27j(z)}{j(z) - 1728}, \frac{27j(z)}{j(z) - 1728} \right) = 0.$$

Since this equation holds for all z such that $j(z) \neq 0, 1728$, it holds identically; in other words

$$P\left((f_{r,s}; \frac{27j}{j - 1728}, \frac{27j}{j - 1728} \right) = 0$$

in the field of meromorphic functions on $\mathfrak{H}$. Therefore the functions $f_{r,s}$ are x-coordinates of affine N-division points on the curve E over $\mathbb{C}(j)$ with which we started. In fact the functions $f_{r,s}$ comprise all such x-coordinates:

Proposition 3. *The set of x-coordinates of affine N-division points on the elliptic curve*

$$E : y^2 = 4x^3 - \frac{27j}{j-1728}x - \frac{27j}{j-1728}$$

coincides with the set of functions $f_{r,s}(z) = (g_2(z)/g_3(z))\wp\left(\frac{r+sz}{N}; \mathcal{L}_z\right)$ *in any algebraic closure of* $\mathbb{C}(j)$ *containing these functions. Therefore*

$$\mathbb{C}(j, E[N]/\pm) = \mathbb{C}(j, \{f_{r,s}\}).$$

Proof. As (r,s) runs over a set of representatives for $\mathcal{R}$ the functions $f_{r,s}$ are all distinct, because their values are distinct at any z such that $j(z) \neq 0, 1728$. Since each function $f_{r,s}$ is the x-coordinate of an affine N-division point on E, and since the number of such x-coordinates, like the number of functions $f_{r,s}$, is $|\mathcal{R}|$, we conclude that the functions $f_{r,s}$ are precisely the x-coordinates of the affine N-division points on E.

1.7. Completion of the proof. The proof of Proposition 2 and hence of Theorem 1 is completed by combining Proposition 3 with the following:

Proposition 4. $\mathfrak{M}(\Gamma(N)) = \mathbb{C}(j, \{f_{r,s}\})$.

Proof. Let us use the notations $f_{r,s}$ and $f_{(r,s)}$ interchangeably. The proof rests on two assertions:

(i) $f_{(r,s)} \circ \gamma = f_{(r,s)\gamma}$ for $\gamma \in \mathrm{SL}(2,\mathbb{Z})$.
(ii) There is a meromorphic function on D which extends the meromorphic function $F_{r,s}$ on D° defined by $f_{r,s}(z) = F_{r,s}(e^{2\pi iz/N})$.

Assertion (i) follows after a calculation from the relations

$$g_2(c\mathcal{L}) = c^{-4}g_2(\mathcal{L}), \quad g_3(c\mathcal{L}) = c^{-6}g_3(\mathcal{L}), \quad \text{and} \quad \wp(cu, c\mathcal{L}) = c^{-2}\wp(u, \mathcal{L}).$$

For (ii) one uses the definition of $\wp(u,\mathcal{L})$ as a sum over lattice points to show that $\lim_{y\to\infty} f_{r,s}(x+iy)$ exists uniformly in x. Now (i) implies that

$$f_{r,s} \circ \gamma = f_{r,s} \quad \text{for } \gamma \in \Gamma(N),$$

while (i) and (ii) together imply that if $\delta \in \mathrm{SL}(2,\mathbb{Z})$ then the meromorphic function F on D° defined by

$$f_{r,s}(\delta z) = F(e^{2\pi iz/N})$$

extends to a meromorphic function on D (put $(r',s') = (r,s)\delta$; then $F = F_{r',s'}$). Therefore $f_{r,s}$ belongs to $\mathfrak{M}(\Gamma(N))$. To see that the $f_{r,s}$ actually generate $\mathfrak{M}(\Gamma(N))$, we use (i) again: if the field inclusion $\mathbb{C}(j, \{f_{r,s}\}) \subset \mathfrak{M}(\Gamma(N))$ were proper, then $\mathbb{C}(j, \{f_{r,s}\})$ would be fixed by a nontrivial subgroup of the Galois group

$$\Gamma(1)/\{\pm I\}\Gamma(N) \cong \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}.$$

But a subgroup of $\mathrm{SL}(2,\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$ which acts trivially on $\mathcal{R}$ is trivial.

1.8. A normalized basis. The arguments just completed lead to a nearly canonical choice of basis for $E[N]$ and hence to a nearly canonical identification of $\mathrm{Gal}(\mathbb{C}(j, E[N])/\mathbb{C}(j))$ with $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$ for any elliptic curve E over $\mathbb{C}(j)$ with invariant j. To formulate the result, let us say that E has good reduction at a point $z \in \mathfrak{H}$ if E has good reduction at the place $j = j(z)$ of $\mathbb{C}(j)$. The reduction of E will be denoted E_z. Recall that we have fixed an isomorphism

$$\theta : \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\} \longrightarrow \mathrm{Gal}(\mathfrak{M}(\Gamma(N))/\mathbb{C}(j))$$

by requiring that for $\gamma \in \mathrm{SL}(2, \mathbb{Z})$ and $f \in \mathfrak{M}(\Gamma(N))$,

$$\theta([\gamma])(f) = f \circ \gamma^{\mathrm{t}},$$

where $[\gamma]$ denotes the image of γ in $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$.

Proposition 5. *Let E be an elliptic curve over $\mathbb{C}(j)$ with invariant j, and view $\mathbb{C}(j, E[N]/\pm)$ and $\mathfrak{M}(\Gamma(N))$ as subfields of a fixed algebraic closure of $\mathbb{C}(j)$.*

(i) *$\mathbb{C}(j, E[N]/\pm) = \mathfrak{M}(\Gamma(N))$. In particular, for any $z \in \mathfrak{H}$, evaluation at z defines a place of $\mathbb{C}(j, E[N]/\pm)$. Henceforth we fix a place $\hat{z}$ of $\mathbb{C}(j, E[N])$ extending evaluation at z on $\mathbb{C}(j, E[N]/\pm)$, and if E has good reduction at z and $P \in E[N]$, then $P_z \in E_z(\mathbb{C})$ denotes the reduciton of P at $\hat{z}$.*

(ii) *There is a basis $\{P_1, P_2\}$ for $E[N]$, unique up to replacement by $\{-P_1, -P_2\}$, with the following properties:*

- (a) *Let $\rho : \mathrm{Gal}(\mathbb{C}(j, E[N])/\mathbb{C}(j)) \longrightarrow \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$ be the isomorphism corresponding to $\{P_1, P_2\}$ and $\rho^{\pm} : \mathrm{Gal}(\mathbb{C}(j, E[N]/\pm)/\mathbb{C}(j)) \longrightarrow \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$ the induced isomorphism. Then $\rho^{\pm} = \theta^{-1}$.*
- (b) *If $z \in \mathfrak{H}$ is a point where E has good reduction, then there is a complex analytic group isomorphism of $\mathbb{C}/\mathcal{L}_z$ onto $E_z(\mathbb{C})$ sending $1/N + \mathcal{L}_z$ to $(P_2)_z$.*

Proof. (i) Since $\mathbb{C}(j, E[N]/\pm)$ depends on E only up to quadratic twist, this follows from Propositions 3 and 4.

(ii) Choose an equation for E over $\mathbb{C}(j)$ of the form

$$c(j)y^2 = 4x^3 - \frac{27j}{j - 1728}x - \frac{27j}{j - 1728},$$

where $c(t) \in \mathbb{C}[t]$ is a polynomial with simple zeros. Then E has good reduction at $z \in \mathfrak{H}$ if and only if $j(z) \neq 0, 1728$ and $c(j(z)) \neq 0$. Now we have seen (in the case $c(t) = 1$, and hence in general) that the x-coordinates of the affine points of order N on E are the functions $f_{r,s}$ with $(r, s) \in \mathbb{Z}^2$ and $(r, s) \not\equiv 0 \bmod N$. Thus for each such pair (r, s) there is a point $P_{r,s} \in E[N]$ such that $x(P_{r,s}) = f_{r,s}$. Of course the definition of $P_{r,s}$

represents an arbitrary choice from among two possibilities. We also make an arbitrary choice of square roots $(g_2(z)/g_3(z))^{3/2}$ and $c(j(z))^{1/2}$ at each point $z \in \mathfrak{H}$ where E has good reduction, and we let $\lambda_z : \mathbb{C}/\mathcal{L}_z \to E_z(\mathbb{C})$ denote the complex analytic group isomorphism afforded by the map

$$u \longmapsto \left(\frac{g_2(z)}{g_3(z)} \wp(u; \mathcal{L}_z), \left(\frac{g_2(z)}{g_3(z)} \right)^{3/2} \frac{\wp'(u; \mathcal{L}_z)}{c(j(z))^{1/2}} \right).$$

Now choose any point $z_0 \in \mathfrak{H}$ where E has good reduction, and let P_1 and P_2 be the preimages of $\lambda_{z_0}(z_0/N + \mathcal{L}_{z_0})$ and $\lambda_{z_0}(1/N + \mathcal{L}_{z_0})$ respectively under the isomorphism $P \mapsto P_{z_0}$ of $E[N]$ onto $E_{z_0}[N]$. Then $\{(P_1)_{z_0}, (P_2)_{z_0}\}$ is a basis for $E_{z_0}[N]$ and *a fortiori* $\{P_1, P_2\}$ is a basis for $E[N]$. We claim that

$$rP_1 + sP_2 = \pm P_{r,s}. \tag{1}$$

Since the reduction map is injective on torsion, it suffices to check that

$$(rP_1 + sP_2)_{z_0} = \pm (P_{r,s})_{z_0}.$$

The left-hand side is $\lambda_{z_0}\left(\dfrac{rz_0 + s}{N} + \mathcal{L}_{z_0} \right)$, while the right-hand side has x-coordinate $f_{r,s}(z_0)$. Therefore equality holds.

To verify (a), take $\sigma \in \mathrm{Gal}(\mathbb{C}(j, E[N]/\pm)/\mathbb{C}(j))$, choose an element $\tilde{\sigma} \in \mathrm{Gal}(\mathbb{C}(j, E[N])/\mathbb{C}(j))$ which restricts to σ, and select $\gamma \in \mathrm{SL}(2, \mathbb{Z})$ so that the image of γ in $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$ is $\rho(\tilde{\sigma})$. Then $\rho^{\pm}(\sigma) = [\gamma]$, and the identity to be proved is $\theta([\gamma]) = \sigma$. Since the $f_{r,s}$ generate $\mathfrak{M}(\Gamma(N))$ over $\mathbb{C}(j)$ (Proposition 4), it suffices to check that $\theta([\gamma])(f_{r,s}) = \sigma(f_{r,s})$. Write

$$\gamma \begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} r' \\ s' \end{pmatrix}.$$

As we have seen in the proof of Proposition 4,

$$\theta([\gamma])(f_{r,s}) = f_{r,s} \circ \gamma^{\mathrm{t}} = f_{(r',s')}.$$

On the other hand, $r'P_1 + s'P_2 = \pm P_{r',s'}$, whence

$$f_{r',s'} = x(r'P_1 + s'P_2) = x(\rho(\tilde{\sigma})(rP_1 + sP_2)) = \sigma(x(rP_1 + sP_2)).$$

By (1), the last term is $\sigma(f_{r,s})$, and (a) follows.

For (b), suppose that E has good reduction at z. Since the x-coordinate of $(P_2)_z$ is $f_{0,1}(z)$, we have $\lambda_z(1/N + \mathcal{L}_z) = \pm(P_2)_z$. Hence either λ_z or $-\lambda_z$ sends $1/N + \mathcal{L}_z$ to $(P_2)_z$.

Finally, suppose that $\{P_1', P_2'\}$ is another basis for $E[N]$ with properties (a) and (b). Choose a point $z \in \mathfrak{H}$ where E has good reduction, and let

$\lambda'_z : \mathbb{C}/\mathcal{L}_z \longrightarrow E_z(\mathbb{C})$ be a complex analytic group isomorphism sending $1/N + \mathcal{L}_z$ to $(P'_2)_z$. Then $\lambda_z^{-1} \circ \lambda'_z \in \mathrm{Aut}(\mathbb{C}/\mathcal{L}_z)$. Since E has good reduction at z we have $j(z) \neq 0, 1728$ and consequently $\mathrm{Aut}(\mathbb{C}/\mathcal{L}_z) = \{\pm 1\}$. Hence after replacing $\{P_1, P_2\}$ by $\{-P_1, -P_2\}$ if necessary, we may assume that $P'_2 = P_2$. Then the change-of-basis matrix sending $\{P_1, P_2\}$ to $\{P'_1, P'_2\}$ is a lower triangular matrix with 1 in the lower right-hand entry. Furthermore, conjugation by this matrix induces the identity on $\mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\}$, because $\{P'_1, P'_2\}$ also has property (a). It follows that the change-of-basis is the identity, as desired.

1.9. Quotients of the upper half-plane. We will use Proposition 5 to realize the modular curves as compactified quotients of $\mathfrak{H}$. First we must recall how such quotients are given the structure of a Riemann surface.

Put

$$\mathfrak{H}^* = \mathfrak{H} \cup \mathbf{P}^1(\mathbb{Q}).$$

The action of $\mathrm{SL}(2, \mathbb{Z})$ on $\mathfrak{H}$ by fractional linear transformations extends to an action on $\mathfrak{H}^*$ preserving $\mathbf{P}^1(\mathbb{Q})$, and if Γ is any subgroup of finite index in $\mathrm{SL}(2, \mathbb{Z})$, then we denote the respective orbit spaces of $\mathfrak{H}^*$, $\mathfrak{H}$, and $\mathbf{P}^1(\mathbb{Q})$ under Γ by $\Gamma\backslash\mathfrak{H}^*$, $\Gamma\backslash\mathfrak{H}$, and $\Gamma\backslash\mathbf{P}^1(\mathbb{Q})$. Thus

$$\Gamma\backslash\mathfrak{H}^* = (\Gamma\backslash\mathfrak{H}) \ \cup \ \left(\Gamma\backslash\mathbf{P}^1(\mathbb{Q})\right).$$

Since Γ has finite index in $\mathrm{SL}(2, \mathbb{Z})$ and $\mathrm{SL}(2, \mathbb{Z})$ acts transitively on $\mathbf{P}^1(\mathbb{Q})$ the set $\Gamma\backslash\mathbf{P}^1(\mathbb{Q})$ is finite.

We would like to put a topology on $\Gamma\backslash\mathfrak{H}^*$. First we put a topology on $\mathfrak{H}^*$ itself. Given $y_0 > 0$ and $c \in \mathbf{P}^1(\mathbb{Q})$, choose a matrix $\delta \in \mathrm{SL}(2, \mathbb{Z})$ such that $c = \delta\infty$, and put

$$U_{y_0} = \{x + iy : x \in \mathbb{R}, y > y_0\} \subset \mathfrak{H},$$
$$U^\circ_{c,y_0} = \delta(U_{y_0}), \quad \text{and} \quad U_{c,y_0} = U^\circ_{c,y_0} \cup \{c\}.$$

The sets U°_{c,y_0} and U_{c,y_0} depend only on c and y_0, not on the choice of δ, because U_{y_0} is preserved by the stabilizer of ∞ in $\mathrm{SL}(2, \mathbb{Z})$, namely $\left\{\pm \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\right\}$. We make $\mathfrak{H}^*$ into a topological space by choosing as a basis of open sets all sets of the following two types:

(a) open subsets U of $\mathfrak{H}$,
(b) subsets of $\mathfrak{H}^*$ of the form U_{c,y_0}.

Then the quotient toplology on $\Gamma\backslash\mathfrak{H}^*$ corresponding to the natural projection

$$\pi : \mathfrak{H}^* \longrightarrow \Gamma\backslash\mathfrak{H}^*$$

makes $\Gamma\backslash\mathfrak{H}^*$ into a compact Hausdorff space.

The next step is to make $\Gamma\backslash\mathfrak{H}^*$ into a compact Riemann surface. Let $\mathcal{F}$ be the sheaf of continuous complex-valued functions on $\Gamma\backslash\mathfrak{H}^*$, and $\mathcal{F}_x$

the stalk at a point x. We think of $\mathcal{F}_x$ as the set of equivalence classes of pairs (f, V), where V is an open neighborhood of x and f is a continuous complex-valued function on V, two pairs (f, V) and (g, W) being equivalent if f and g coincide on $V \cap W$. To make $\Gamma\backslash\mathfrak{H}^*$ into a Riemann surface, we must define a subsheaf $\mathcal{O}$ of $\mathcal{F}$ to serve as the complex structure sheaf. We define $\mathcal{O}$ by specifying that its stalk $\mathcal{O}_x$ at x is the subring of $\mathcal{F}_x$ consisting of those equivalence classes which contain a pair (f, V) of one of the following two types:

(a) There exists $z \in \mathfrak{H}$ and an open neighborhood U of z in $\mathfrak{H}$ such that $x = \pi(z)$, $V = \pi(U)$, and $f \circ \pi$ is holomorphic on U.

(b) There exists $c \in \mathbf{P}^1(\mathbb{Q})$ and $y_0 > 0$ such that $x = \pi(c)$, $V = \pi(U_{c,y_0})$, and $f \circ \pi$ satisfies the following condition. Choose $\delta \in \mathrm{SL}(2, \mathbb{Z})$ such that $c = \delta\infty$, and let M be a positive integer such that $(f \circ \pi \circ \delta)(z + M) = (f \circ \pi \circ \delta)(z)$ for $z \in U_{y_0}$. (Such an integer exists because Γ has finite index in $\mathrm{SL}(2, \mathbb{Z})$ and π is invariant under Γ.) Put $r = e^{-2\pi y_0/M}$ and let F be the function on the punctured disk $D^\circ(r) = \{q \in \mathbb{C} : 0 < |q| < r\}$ such that

$$(f \circ \pi \circ \delta)(z) = F(e^{2\pi i z/M}).$$

Then F is holomorphic on $D^\circ(r)$ and extends to a holomorphic function on the full disk $D(r) = \{q \in \mathbb{C} : |q| < r\}$.

One can check that with this definition of $\mathcal{O}$, every point x of $\Gamma\backslash\mathfrak{H}^*$ has an open neighborhood V such that the ringed space $(V, \mathcal{O}|_V)$ is isomorphic to the ringed space of an open disk in $\mathbb{C}$. (The verification requires a little care if x is the image of an elliptic fixed point of Γ, i.e., if $x = \pi(z)$ for some $z \in \mathfrak{H}$ which is fixed by an element of Γ different from $\pm I$.) Granting that this is so, we conclude that $\mathcal{O}$ gives $\Gamma\backslash\mathfrak{H}^*$ the structure of a Riemann surface. Furthermore, and this is now the key point, the definitions have been constructed in such a way that the map

$$f \longmapsto (f \circ \pi)|\mathfrak{H}$$

identifies the function field of $\Gamma\backslash\mathfrak{H}^*$ with $\mathfrak{M}(\Gamma)$. Note that both $\Gamma\backslash\mathfrak{H}^*$ and $\mathfrak{M}(\Gamma)$ depend only on $\overline{\Gamma}$, the image of Γ in $\mathrm{SL}(2, \mathbb{Z})/\{\pm I\}$.

1.10. Modular curves as quotients of the upper half-plane. Given a modular curve $X(H)$, we shall now produce a subgroup Γ of $\mathrm{SL}(2, \mathbb{Z})$ such that the Riemann surfaces $X(H)(\mathbb{C})$ and $\Gamma\backslash\mathfrak{H}^*$ are isomorphic. By assumption, H is a subgroup of $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ satisfying two conditons: $-I \in H$ and $\det : H \to (\mathbb{Z}/N\mathbb{Z})^\times$ is surjective. We let $\Gamma \subset \mathrm{SL}(2, \mathbb{Z})$ be the transpose of the inverse image of $H \cap SL(2, \mathbb{Z}/N\mathbb{Z})$ under the reduction map $\mathrm{SL}(2, \mathbb{Z}) \to \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})$.

Proposition 6. *With H and Γ as above, the Riemann surfaces $X(H)(\mathbb{C})$ and $\Gamma\backslash\mathfrak{H}^*$ are isomorphic.*

Proof. Let E be an elliptic curve over $\mathbb{Q}(j)$ with invariant j, and identify $\mathrm{Gal}(\mathbb{Q}(j, E[N])/\mathbb{Q}(j))$ with $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ using a basis for $E[N]$ as in Proposition 5. The function field of $X(H)$ over $\mathbb{Q}$ is the subfield K of $\mathbb{Q}(j, E[N]/\pm)$ fixed by H, whence the function field of the Riemann surface $X(H)(\mathbb{C})$ is $\mathbb{C}K$. Now our identification of $\mathrm{Gal}(\mathbb{Q}(j, E[N])/\mathbb{Q}(j))$ with $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ affords an identification

$$\mathrm{Gal}(\mathbb{C}(j, E[N]/\pm)/\mathbb{C}(j)) \cong \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z})/\{\pm I\},$$

and the hypotheses on H imply that $\mathbb{C}K$ is the subfield of $\mathbb{C}(j, E[N]/\pm)$ fixed by $(H \cap \mathrm{SL}(2, \mathbb{Z}/N\mathbb{Z}))/\{\pm I\}$. Applying parts (i) and (ii)(a) of Proposition 5, we deduce that $\mathbb{C}K = \mathfrak{M}(\Gamma)$, whence the result follows from the fact that a compact Riemann surface is determined up to isomorphism by its function field.

In particular, put

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z}) : c \equiv 0 \bmod N \right\}$$

and

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z}) : c \equiv 0 \bmod N, a, d \equiv 1 \bmod N \right\}.$$

Then $X_0(N)(\mathbb{C}) \cong \Gamma_0(N)\backslash\mathfrak{H}^*$ and $X_1(N)(\mathbb{C}) \cong \Gamma_1(N)\backslash\mathfrak{H}^*$ (in the latter case we use the fact that $\Gamma\backslash\mathfrak{H}^*$ depends only on $\overline{\Gamma}$). Now consider pairs $(\mathcal{T}, \mathcal{C})$ consisting of a one-dimensional complex torus $\mathcal{T}$ and a cyclic subgroup $\mathcal{C}$ of $\mathcal{T}$ of order N. An isomorphism from one pair $(\mathcal{T}_1, \mathcal{C}_1)$ to another $(\mathcal{T}_2, \mathcal{C}_2)$ is a complex analytic group isomorphism from $\mathcal{T}_1$ to $\mathcal{T}_2$ sending $\mathcal{C}_1$ onto $\mathcal{C}_2$. We denote the isomorphism class containing $(\mathcal{T}, \mathcal{C})$ by $[\mathcal{T}, \mathcal{C}]$ and the set of all isomorphism classes by $\mathrm{Tori}_0(N)$. For a point $\mathcal{P}$ of order N on $\mathcal{T}$ we make the analogous definitions of $(\mathcal{T}, \mathcal{P})$, $[\mathcal{T}, \mathcal{P}]$, and $\mathrm{Tori}_1(N)$.

Proposition 7. *Let E be an elliptic curve over $\mathbb{Q}(j)$ with invariant j, and let S be a subset of $\mathbf{P}^1(\mathbb{C})$ containing all places where E has bad reduction. Fix an ordered basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, let P be the second element of this basis, and let C be the cyclic group of order N generated by P. Then there is an isomorphism of Riemann surfaces $X_0(N)(\mathbb{C}) \cong \Gamma_0(N)\backslash\mathfrak{H}^*$ such that the diagram*

$$\begin{array}{ccc} X_0(N)(\mathbb{C})_S & \longrightarrow & \mathrm{Ell}_0(N)(\mathbb{C})_S \\ \downarrow & & \downarrow \\ \Gamma_0(N)\backslash\mathfrak{H} & \longrightarrow & \mathrm{Tori}_0(N) \end{array}$$

commutes, where:

- *The top horizontal arrow is the bijection* $x \mapsto [E_x, C_x]$ *of Proposition 1.*
- *The bottom horizontal arrow is a bijection and has the form*

$$[z] \longmapsto [\mathbb{C}/\mathcal{L}_z, \langle 1/N + \mathcal{L}_z \rangle] \qquad (z \in \mathfrak{H}),$$

  where $[z]$ *denotes the class of* z *in* $\Gamma_0(N)\backslash\mathfrak{H}$ *and* $\langle 1/N + \mathcal{L}_z \rangle$ *denotes the cyclic subgroup of* $\mathbb{C}/\mathcal{L}_z$ *generated by the coset of* $1/N + \mathcal{L}_z$.
- *The left vertical arrow is the restriction to* $X_0(N)(\mathbb{C})_S$ *of the isomorphism* $X_0(N)(\mathbb{C}) \cong \Gamma_0(N)\backslash\mathfrak{H}^*$.
- *The right vertical arrow is the restriction to* $\mathrm{Ell}_0(N)(\mathbb{C})_S$ *of the bijection from* $\mathrm{Ell}_0(N)(\mathbb{C})$ *to* $\mathrm{Tori}_0(N)$ *given by* $[\mathcal{E}, C] \mapsto [\mathcal{E}(\mathbb{C}), C]$.

The same is true if $X_0(N)$, $\mathrm{Ell}_0(N)$, $\Gamma_0(N)$, $\mathrm{Tori}_0(N)$, C, *and* $\langle 1/N + \mathcal{L}_z \rangle$ *are replaced by* $X_1(N)$, $\mathrm{Ell}_1(N)$, $\Gamma_1(N)$, $\mathrm{Tori}_1(N)$, P, *and* $1/N + \mathcal{L}_z$.

Proof. Without loss of generality we may assume that P is the second basis vector in a basis for $E[N]$ chosen as in Proposition 5. Then the only statement requiring proof is the bijectivity of the bottom horizontal arrow. The cases $\Gamma_0(N)$ and $\Gamma_1(N)$ are similar; we deal with the latter. Suppose that

$$[\mathbb{C}/\mathcal{L}_z, 1/N + \mathcal{L}_z] = [\mathbb{C}/\mathcal{L}_{z'}, 1/N + \mathcal{L}_{z'}].$$

Then there exists $\omega \in \mathbb{C}^\times$ so that $\mathcal{L}_{z'} = \omega\mathcal{L}_z$ and $1/N \equiv \omega/N \pmod{\omega\mathcal{L}_z}$. The first condition implies that $\{\omega, \omega z\}$ is a basis for $\mathcal{L}_{z'}$. Hence we can write

$$\begin{cases} z' = \omega(az + b) \\ 1 = \omega(cz + d) \end{cases}$$

with integers a, b, c, d satisfying $ad - bc = \pm 1$. Since $z' = (az + b)/(cz + d)$ and z and z' both have positive imaginary part, it follows that $ad - bc = 1$. Substituting $1 = \omega(cz + d)$ in the congruence $1/N \equiv \omega/N \pmod{\omega\mathcal{L}_z}$, we find that $c \equiv 0 \pmod{N}$ and $d \equiv 1 \pmod{N}$, whence $a \equiv 1 \pmod{N}$ also since $ad - bc = 1$. Thus $z' = \gamma z$ with

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N),$$

and consequently $[z] = [z']$. Next suppose that $[\mathcal{T}, \mathcal{P}] \in \mathrm{Tori}_1(N)$. Write $\mathcal{T} = \mathbb{C}/\mathcal{L}$ and $\mathcal{P} = \omega/N + \mathcal{L}$ with $\omega \in \mathcal{L}$. After replacing ω by another element of $\omega + N\mathcal{L}$, we may assume that ω is primitive, so that ω is part of a basis $\{\omega', \omega\}$ for $\mathcal{L}$. Put $z = \pm\omega'/\omega$, where the sign is chosen so that $\mathrm{Im}(z) > 0$. Multiplication by ω^{-1} gives an isomorphism of $(\mathbb{C}/\mathcal{L}, \mathcal{P} + \mathcal{L})$ onto $(\mathbb{C}/\mathcal{L}_z, 1/N + \mathcal{L}_z)$, whence $[\mathcal{T}, \mathcal{P}]$ coincides with $[\mathbb{C}/\mathcal{L}_z, 1/N + \mathcal{L}_z]$.

2. HECKE CORRESPONDENCES

By a *correspondence* on a smooth projective curve X we shall mean a triple $T = (Z, \varphi, \psi)$, where Z is a smooth projective curve and φ and ψ are nonconstant morphisms $Z \to X$. We say that T is defined over a field $\Bbbk$ if X, Z, φ, and ψ are all defined over $\Bbbk$. We view an automorphism δ of X as a special case of a correspondence by putting $Z = X$, $\varphi = \mathrm{id}_X$, and $\psi = \delta$.

2.1. The Hecke correspondences on $X_0(N)$. Let N be a positive integer, p a prime number, and M the least common multiple of N and p. Choose an elliptic curve E over $\mathbb{Q}(t)$ with invariant t, and fix a basis for $E[M]$ over $\mathbb{Z}/M\mathbb{Z}$, whence an identification of $\mathrm{Gal}(\mathbb{Q}(t, E[M])/\mathbb{Q}(t))$ with $\mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z})$. We consider the subgroup

$$H_p = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z}) : c \equiv 0 \bmod N, \quad b \equiv 0 \bmod p \right\}.$$

Since $-I \in H_p$ and $\det(H_p) = (\mathbb{Z}/M\mathbb{Z})^\times$, the fixed field of H_p is the function field of a smooth projective curve over $\mathbb{Q}$, which we shall denote $X_0(N, p)$. The Hecke correspondence T_p on $X_0(N)$ is a correspondence over $\mathbb{Q}$ of the form

$$T_p = (X_0(N, p), \varphi_p, \psi_p),$$

where the morphisms $\varphi_p, \psi_p : X_0(N, p) \to X_0(N)$ must now be defined.

The definition of φ_p is straightforward. Let K_p and K denote the fixed fields of H_p and

$$H = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z}) : c \equiv 0 \bmod N \right\}$$

respectively. Then $H_p \subset H$, whence $K \subset K_p$. The latter inclusion is an inclusion of function fields and so corresponds to a morphism of curves

$$\varphi_p : X_0(N, p) = X(H_p) \longrightarrow X(H).$$

But $X(H)$ is $X_0(N)$, because the kernel of the reduction map

$$\mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z}) \to \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$$

is a subgroup of H and the image of H in $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$ is the lower triangular group. Therefore φ_p is a morphism from $X_0(N, p)$ to $X_0(N)$.

The definition of ψ_p is more subtle. It corresponds to an inclusion of function fields $K' \hookrightarrow K_p$, where K' is a subfield of K_p which is isomorphic to, but distinct from, K. To define K', let us recall once again that our identification of $\mathrm{Gal}(\mathbb{Q}(t, E[M])/\mathbb{Q}(t))$ with $\mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z})$ rests on a choice

of basis for $E[M]$ over $\mathbb{Z}/M\mathbb{Z}$ and hence in particular on a decomposition of $E[M]$ as a direct sum of cyclic subgroups of order M:

$$E[M] = C_1 \oplus C_2.$$

Let C denote the cyclic subgroup of C_2 of order N, and let Π denote the cyclic subgroup of C_1 of order p. Then C and Π are stable under H_p, hence defined over K_p. In particular, since Π is defined over K_p there is an elliptic curve E/Π defined over K_p together with an isogeny

$$\lambda : E \longrightarrow E/\Pi$$

over K_p with kernel Π. Furthermore, E/Π has a cyclic subgroup of order N defined over K_p, namely the subgroup $\lambda(C)$. Now put

$$t' = j(E/\Pi) \in K_p,$$

and let E' be an elliptic curve over $\mathbb{Q}(t')$ with invariant t'. Then there is an isomorphism $\theta : E/\Pi \to E'$ over $\overline{K_p}$, and θ is unique up to sign because t' is transcendental, hence $\neq 0, 1728$. It follows that the group $C' = \theta(\lambda(C))$ is a cyclic subgroup of E' of order N which is independent of the choice of θ. Furthermore, C' is defined over K_p because $\lambda(C)$ is defined over K_p and $\sigma \circ \theta \circ \sigma^{-1} = \pm\theta$ for $\sigma \in \mathrm{Gal}(\overline{K_p}/K_p)$. Thus K_p contains the field K' fixed by

$$\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(t')}/\mathbb{Q}(t')) : \sigma(C') = C'\}.$$

Since K' is isomorphic to the function field of $X_0(N)$ we obtain the desired morphism ψ_p from $X_0(N,p)$ to $X_0(N)$.

2.2. The Hecke correspondences on $X_1(N)$. *Mutatis mutandis*, the same construction yields a correspondence

$$T_p = (X_1(N,p), \varphi_p, \psi_p)$$

on $X_1(N)$, where $X_1(N,p)$ is the modular curve determined by the subgroup

$$H_p = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z}) : \begin{matrix} c \equiv 0 \bmod N, & b \equiv 0 \bmod p \\ d \equiv \pm 1 \bmod N & \end{matrix} \right\}$$

of $\mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z})$. Put

$$H = \left\{ \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/M\mathbb{Z}) : c \equiv 0 \bmod N, \quad d \equiv \pm 1 \bmod N \right\}$$

and write K_p and K for the subfields of $\mathbb{Q}(t, E[M])$ fixed by H_p and H. Then φ_p is the morphism $X_1(N,p) \to X_1(N)$ corresponding to the inclusion

of K in K_p. To define ψ_p, let $\{P_1, P_2\}$ be our chosen basis for $E[M]$ and put $P = (M/N)P_2$. Also let Π be the group of order p generated by $(M/p)P_1$. As before, there is an elliptic curve E/Π over K_p and an isogeny $\lambda : E \to E/\Pi$ over K_p with kernel Π. Since K is contained in K_p and the set $\{\pm P\}$ is stable under $\mathrm{Gal}(\mathbb{Q}(t, E[M])/K)$, it follows that $\{\pm\lambda(P)\}$ is stable under $\mathrm{Gal}(\mathbb{Q}(t, E[M])/K_p)$. Putting $t' = j(E/\Pi)$ as before, we see that if E' is any elliptic curve over $\mathbb{Q}(t')$ with invariant t' and $\theta : E/\Pi \to E'$ is any isomorphism over $\overline{K_p}$, then the point $P' = \theta(\lambda(P))$ has order N and $\{\pm P\}$ is defined over K_p. Hence K_p contains K', the field fixed by

$$\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(t')}/\mathbb{Q}(t')) : \sigma(P') = \pm P'\}.$$

Since K' is isomorphic to the function field of $X_1(N)$ we obtain a morphism ψ_p from $X_1(N, p)$ to $X_1(N)$.

2.3. Moduli interpretation of the Hecke correspondences. We denote the free abelian group on a set W by $\mathrm{Div}(W)$. In particular, if X is a smooth projective curve over an algebraically closed field $\Bbbk$, then $\mathrm{Div}(X(\Bbbk))$ is the usual group of divisors on $X(\Bbbk)$. Given a correspondence $T = (Z, \varphi, \psi)$ on X, we use the same letter T to denote the map

$$\begin{aligned} X(\Bbbk) &\longrightarrow \mathrm{Div}(X(\Bbbk)) \\ x &\longmapsto \sum_{\substack{z \in Z \\ \varphi(z) = x}} (\mathrm{mult}_z \varphi)\psi(z), \end{aligned}$$

where $\mathrm{mult}_z\varphi$ is the ramification index of φ at z. In the case of the Hecke correspondence T_p we shall give a formula for this map which displays its effect on isomorphism classes $[\mathcal{E}, \mathcal{C}]$ and $[\mathcal{E}, \mathcal{P}]$. First a point of notation.

Suppose that $\mathcal{E}$ is an elliptic curve over an algebraically closed field $\Bbbk$ and Λ is a subgroup of $\mathcal{E}$ of order p. In keeping with our usage thus far, we shall write $\mathcal{E}/\Lambda$ for an elliptic curve which is the image of a separable isogeny with domain $\mathcal{E}$ and kernel Λ. Note that $\mathcal{E}/\Lambda$ is unique up to isomorphism. Now if $\lambda : \mathcal{E} \to \mathcal{E}/\Lambda$ is a separable isogeny with kernel Λ and $\mathcal{C}$ is a cyclic subgroup of $\mathcal{E}$ of order N which intersects Λ trivially (a vacuous condition if N is prime to p), then we obtain a well-defined isomorphism class

$$[\mathcal{E}/\Lambda, (\mathcal{C} + \Lambda)/\Lambda] \in \mathrm{Ell}_0(N)(\Bbbk)$$

by putting $[\mathcal{E}/\Lambda, (\mathcal{C} + \Lambda)/\Lambda] = [\lambda(\mathcal{E}), \lambda(\mathcal{C})]$. To see that $[\mathcal{E}/\Lambda, (\mathcal{C} + \Lambda)/\Lambda]$ is independent of the choice of λ, suppose that $\lambda' : \mathcal{E} \to \mathcal{E}/\Lambda$ is another such isogeny. Then there is an automorphism θ of $\mathcal{E}/\Lambda$ such that $\lambda' = \theta \circ \lambda$, whence $[\lambda'(\mathcal{E}), \lambda'(\mathcal{C})] = [\lambda(\mathcal{E}), \lambda(\mathcal{C})]$. Similarly, if $\mathcal{P}$ is a point of order N on $\mathcal{E}$ such that the cyclic subgroup $\langle\mathcal{P}\rangle$ generated by $\mathcal{P}$ intersects Λ trivially, then we define

$$[\mathcal{E}/\Lambda, \mathcal{P} + \Lambda] \in \mathrm{Ell}_1(N)(\Bbbk)$$

by putting $[\mathcal{E}/\Lambda, \mathcal{P} + \Lambda] = [\lambda(\mathcal{E}), \lambda(\mathcal{P})]$.

Proposition 8. *Let E be an elliptic curve over $\mathbb{Q}(t)$ with invariant t. Let S, S', and S'' be subsets of $\mathbf{P}^1(\mathbb{C})$ containing all places where E has bad reduction and such that*

$$\varphi_p^{-1}(X_0(N)(\mathbb{C})_S) \subset X_0(N,p)(\mathbb{C})_{S'}$$

and

$$\psi_p(X_0(N,p)(\mathbb{C})_{S'}) \subset X_0(N)(\mathbb{C})_{S''}.$$

Fix an ordered basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$, let P be the second element of this basis, and let C be the cyclic group generated by P. Then the diagram

$$\begin{array}{ccc} X_0(N)(\mathbb{C})_S & \xrightarrow{T_p} & \mathrm{Div}(X_0(N)(\mathbb{C})_{S''}) \\ \downarrow & & \downarrow \\ \mathrm{Ell}_0(N)(\mathbb{C}) & \longrightarrow & \mathrm{Div}(\mathrm{Ell}_0(N)(\mathbb{C})) \end{array}$$

commutes, where the left vertical arrow is the map $x \mapsto [E_x, C_x]$ of Proposition 1, the right vertical arrow is the corresponding homomorphism between free abelian groups, and the bottom horizontal arrow is the map

$$[\mathcal{E}, \mathcal{C}] \longmapsto \sum_{\substack{[\mathcal{E}[p]:\Lambda]=p \\ \mathcal{C}\cap\Lambda=\{0\}}} [\mathcal{E}/\Lambda, (\mathcal{C}+\Lambda)/\Lambda],$$

the sum being taken over subgroups Λ of index p in $\mathcal{E}[p]$ which intersect $\mathcal{C}$ trivially. The same is true if $X_0(N)$ is replaced by $X_1(N)$, $\mathrm{Ell}_0(N)$ by $\mathrm{Ell}_1(N)$, the left vertical arrow by $x \mapsto [E_x, P_x]$, and the bottom horizontal arrow by

$$[\mathcal{E}, \mathcal{P}] \longmapsto \sum_{\substack{[\mathcal{E}[p]:\Lambda]=p \\ \langle\mathcal{P}\rangle\cap\Lambda=\{0\}}} [\mathcal{E}/\Lambda, \mathcal{P}+\Lambda],$$

where $\langle\mathcal{P}\rangle$ denotes the cyclic subgroup generated by $\mathcal{P}$.

Proof. For $x \in X_0(N)(\mathbb{C})_S$ the formula

$$T_p(x) = \sum_{\substack{z\in Z \\ \varphi_p(z)=x}} (\mathrm{mult}_z\varphi_p)\psi_p(z)$$

can be written simply as

$$T_p(x) = \sum_{z\in\varphi_p^{-1}(x)} \psi_p(z),$$

because the morphism $\varphi_p : X_0(N,p) \to X_0(N)$ is unramified outside S: indeed the corresponding extension of function fields K_p/K is contained inside the extension $\mathbb{Q}(t, E[M])/\mathbb{Q}(t)$ and is therefore unramified outside the places where E has bad reduction. Here M denotes the least common multiple of N and p, as before.

Consider triples $(\mathcal{E}, \mathcal{C}, \Lambda)$, where $\mathcal{E}$ is an elliptic curve over $\mathbb{C}$, $\mathcal{C}$ is a cyclic subgroup of $\mathcal{E}$ of order N, and Λ is a cyclic subgroup of $\mathcal{E}$ of order p which intersects $\mathcal{C}$ trivially. We write $[\mathcal{E}, \mathcal{C}, \Lambda]$ for the isomorphism class of $(\mathcal{E}, \mathcal{C}, \Lambda)$ and $\mathrm{Ell}_0(N,p)(\mathbb{C})$ for the set of isomorphism classes. If we define maps φ and ψ from $\mathrm{Ell}_0(N,p)(\mathbb{C})$ to $\mathrm{Ell}_0(N)(\mathbb{C})$ by

$$\varphi([\mathcal{E}, \mathcal{C}, \Lambda]) = [\mathcal{E}, \mathcal{C}]$$

and

$$\psi([\mathcal{E}, \mathcal{C}, \Lambda]) = [\mathcal{E}/\Lambda, (\mathcal{C} + \Lambda)/\Lambda],$$

then the map

$$[\mathcal{E}, \mathcal{C}] \longmapsto \sum_{\substack{[\mathcal{E}[p]:\Lambda]=p \\ \mathcal{C}\cap\Lambda=\{0\}}} [\mathcal{E}/\Lambda, (\mathcal{C} + \Lambda)/\Lambda],$$

in the statement of the proposition has the form

$$x \longmapsto \sum_{z \in \varphi^{-1}(x)} \psi(z).$$

Therefore it suffices to check that the diagrams

$$\begin{array}{ccc} X_0(N,p)(\mathbb{C})_{S'} & \xrightarrow{\varphi_p} & X_0(N)(\mathbb{C})_S \\ \downarrow & & \downarrow \\ \mathrm{Ell}_0(N,p)(\mathbb{C}) & \xrightarrow[\varphi]{} & \mathrm{Ell}_0(N)(\mathbb{C}) \end{array}$$

and

$$\begin{array}{ccc} X_0(N,p)(\mathbb{C})_{S'} & \xrightarrow{\psi_p} & X_0(N)(\mathbb{C})_{S''} \\ \downarrow & & \downarrow \\ \mathrm{Ell}_0(N,p)(\mathbb{C}) & \xrightarrow[\psi]{} & \mathrm{Ell}_0(N)(\mathbb{C}) \end{array}$$

commute, where the right vertical arrows are the maps $x \mapsto [E_x, C_x]$ of Proposition 1 and the left vertical arrows are given by

$$\begin{aligned} X_0(N,p)(\mathbb{C})_{S'} &\longrightarrow \mathrm{Ell}_0(N,p)(\mathbb{C}) \\ z &\longmapsto [E_z, C_z, \Pi_z]. \end{aligned}$$

As before, Π denotes a subgroup of E of order p which intersects C trivially, and the subscript z indicates reduction modulo the maximal ideal of the discrete valuation ring corresponding to z in the complex function field of $X_0(N,p)$. Now the commutativity of the first diagram amounts to the equation

$$[E_z, C_z] = [E_{\varphi_p(z)}, C_{\varphi_p(z)}],$$

which follows from the compatibility of reduction at a good place with base extension. To verify that the second diagram commutes put $t' = j(E/\Pi)$, choose an elliptic curve E' over $\mathbb{Q}(t')$ with invariant t', the image of $(C+\Pi)/\Pi$ under an isomorphism $E/\Pi \to E'$. We must check that

$$[E_z/\Pi_z, (C_z+\Pi_z)/\Pi_z] = [E'_{\psi_p(z)}, C'_{\psi_p(z)}]. \tag{1}$$

We verify this equation in two steps. First,

$$[E_z/\Pi_z, (C_z+\Pi_z)/\Pi_z] = [(E/\Pi)_z, ((C+\Pi)/\Pi)_z] \tag{2}$$

by the compatibility of reduction with isogenies. Next let θ be an isomorphism from E/Π to E', so that $C' = \theta((C+\Pi)/\Pi)$. Then the reduction of θ gives an isomorphism from $(E/\Pi)_z$ to $(E')_{\psi_p(z)}$ sending $((C+\Pi)/\Pi)_z$ to $(C')_{\psi_p(z)}$, whence

$$[(E/\Pi)_z, ((C+\Pi)/\Pi)_z] = [(E')_{\psi_p(z)}, (C')_{\psi_p(z)}]. \tag{3}$$

Together, (2) and (3) give (1). The argument for $X_1(N)$ is similar.

2.4. The Hecke correspondences on the upper half-plane. Given $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ we write $\langle d \rangle$ to denote any element of $\Gamma_0(N)$ with lower right-hand entry congruent to d modulo N. Also, if $d \in \mathbb{Z}$ is an integer prime to N then we write $\langle d \rangle$ for $\langle d \bmod N \rangle$.

Proposition 9. *There is a commutative diagram*

$$\begin{array}{ccc} X_0(N)(\mathbb{C}) & \xrightarrow{T_p} & \mathrm{Div}(X_0(N)(\mathbb{C})) \\ \downarrow & & \downarrow \\ \Gamma_0(N)\backslash\mathfrak{H}^* & \longrightarrow & \mathrm{Div}(\Gamma_0(N)\backslash\mathfrak{H}^*), \end{array}$$

where the left vertical arrow is the isomorphism of Proposition 7, the right vertical arrow is the corresponding homomorphism of free abelian groups, and the bottom horizontal arrow is the map

$$[z] \longmapsto \begin{cases} \sum_{\nu=0}^{p-1}[(z+\nu)/p] + [pz] & \text{if } p \nmid N \\ \sum_{\nu=0}^{p-1}[(z+\nu)/p] & \text{if } p|N. \end{cases}$$

The same is true if $X_0(N)$ is replaced by $X_1(N)$ and $\Gamma_0(N)$ by $\Gamma_1(N)$ provided that the bottom horizontal arrow is modified as follows: in the case $p \nmid N$, the term $[pz]$ is replaced by $[\langle p\rangle pz]$.

Proof. By a continuity argument it suffices to check that the diagram commutes when $X_0(N)(\mathbb{C})$ is replaced by $X_0(N)(\mathbb{C})_S$ for some finite set S. Propositions 7 and 8 then reduce the problem to the following: Given $[\mathcal{T},\mathcal{C}] \in \mathrm{Tori}_0(N)$ with $\mathcal{T} = \mathbb{C}/\mathcal{L}_z$ and $\mathcal{C} = \langle 1/N + \mathcal{L}_z\rangle$, show that

$$\sum_{\substack{[\mathcal{T}[p]:\Lambda]=p \\ \mathcal{C}\cap\Lambda=\{0\}}} [\mathcal{T}/\Lambda, (\mathcal{C}+\Lambda)/\Lambda]$$

coincides with

$$\sum_{\nu=0}^{p-1} [\mathbb{C}/\mathcal{L}_{(z+\nu)/p}, \langle 1/N + \mathcal{L}_{(z+\nu)/p}\rangle] + \ [\mathbb{C}/\mathcal{L}_{pz}, \langle 1/N + \mathcal{L}_{pz}\rangle],$$

the last term being omitted if p divides N. Now the subgroups of order p in $\mathbb{C}/\mathcal{L}_z$ which intersect $\langle 1/N + \mathcal{L}_z\rangle$ trivially are

$$\langle (z+\nu)/p + \mathcal{L}_z\rangle \qquad (0 \le \nu \le p-1)$$

and also

$$\langle 1/p + \mathcal{L}_z\rangle$$

if $p \nmid N$. Furthermore, for $\mathcal{T} = \mathbb{C}/\mathcal{L}_z$ and $\Lambda = \langle (z+\nu)/p + \mathcal{L}_z\rangle$ we have $\mathcal{T}/\Lambda \cong \mathbb{C}/\mathcal{L}_{(z+\nu)/p}$. Hence the only point to check is that if $p \nmid N$ then

$$[\mathbb{C}/(z\mathbb{Z} \oplus p^{-1}\mathbb{Z}), \langle 1/N + (z\mathbb{Z} \oplus p^{-1}\mathbb{Z})\rangle] = [\mathbb{C}/\mathcal{L}_{pz}, \langle 1/N + \mathcal{L}_{pz}\rangle].$$

This holds because multiplication by p maps $\mathbb{C}/(z\mathbb{Z}\oplus p^{-1}\mathbb{Z})$ isomorphically onto $\mathbb{C}/\mathcal{L}_{pz}$ and because the cosets of $1/N$ and p/N generate the same subgroup of $\mathbb{C}/\mathcal{L}_{pz}$ for $p \nmid N$.

The argument for $\Gamma_1(N)$ is much the same, except that in the case $p \nmid N$ one must check that

$$[\mathbb{C}/\mathcal{L}_{pz}, p/N + \mathcal{L}_{pz}] = [\mathbb{C}/\mathcal{L}_{\langle p\rangle pz}, 1/N + \mathcal{L}_{\langle p\rangle pz}].$$

Put $w = pz$ and $\gamma = \langle p\rangle$, and write $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Since $\mathcal{L}_w = w\mathbb{Z} \oplus \mathbb{Z}$ and $\mathcal{L}_{\gamma w} = \frac{aw+b}{cw+d}\mathbb{Z} \oplus \mathbb{Z}$, multiplication by $cw+d$ defines an isomorphism from $\mathbb{C}/\mathcal{L}_{\gamma w}$ to $\mathbb{C}/\mathcal{L}_w$ sending $1/N + \mathcal{L}_{\gamma w}$ to $(cw+d)/N + \mathcal{L}_w$. The latter coset coincides with $p/N + \mathcal{L}_w$, because

$$(cw+d)/N - p/N = (c/N)w + (d-p)/N \in \mathcal{L}_w.$$

2.5. The diamond automorphisms. The next construction pertains only to $X_1(N)$. Choose an elliptic curve E over $\mathbb{Q}(t)$ with invariant t, and make the usual identification of $\mathrm{Gal}(\mathbb{Q}(t, E[N])/\mathbb{Q}(t))$ with $\mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z})$, and of the function field K of $X_1(N)$ with the fixed field of

$$H = \left\{ \begin{pmatrix} a & 0 \\ b & \pm 1 \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/N\mathbb{Z}) : a \in (\mathbb{Z}/N\mathbb{Z})^\times, b \in (\mathbb{Z}/N\mathbb{Z}) \right\}.$$

Since H is normal in the lower triangular subgroup B, the quotient group B/H acts as a group of automorphisms of K and hence of $X_1(N)$. We shall identify B/H with $(\mathbb{Z}/N\mathbb{Z})^\times/\{\pm 1\}$ via the map sending the coset of $\begin{pmatrix} a & 0 \\ b & d \end{pmatrix}$ modulo H to the coset of d modulo $\{\pm 1\}$. The automorphism of $X_1(N)$ corresponding to the coset of $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ modulo $\{\pm 1\}$ will be denoted $\langle d \rangle$. Of course we have already used the symbol $\langle d \rangle$ to denote an element of $\Gamma_0(N)$. The next two propositions show that the notations are consistant and that $\langle d \rangle$ may also be used for the bijection from $\mathrm{Ell}_1(N)(\mathbb{C})$ to itself given by $[\mathcal{E}, \mathcal{P}] \mapsto [\mathcal{E}, d\mathcal{P}]$. The proofs are left to the reader.

Proposition 10. *Let E be an elliptic curve over $\mathbb{Q}(t)$ with invariant t, and let S be a subset of $\mathbf{P}^1(\mathbb{C})$ containing all places where E has bad reduction. Fix an ordered basis for $E[N]$ over $\mathbb{Z}/N\mathbb{Z}$ and let P be the second element of this basis. For $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ the diagram*

$$\begin{array}{ccc} X_1(N)(\mathbb{C})_S & \xrightarrow{\langle d \rangle} & X_1(N)(\mathbb{C})_S \\ \downarrow & & \downarrow \\ \mathrm{Ell}_1(N)(\mathbb{C}) & \longrightarrow & \mathrm{Ell}_1(N)(\mathbb{C}) \end{array}$$

commutes, where the bottom horizontal arrow is the map $[\mathcal{E}, \mathcal{P}] \mapsto [\mathcal{E}, d\mathcal{P}]$ and the vertical arrows are the map $x \mapsto [E_x, C_x]$ of Proposition 1.

Proposition 11. *There is a commutative diagram*

$$\begin{array}{ccc} X_1(N)(\mathbb{C}) & \xrightarrow{\langle d \rangle} & X_1(N)(\mathbb{C}) \\ \downarrow & & \downarrow \\ \Gamma_1(N)\backslash\mathfrak{H}^* & \longrightarrow & \Gamma_1(N)\backslash\mathfrak{H}^*, \end{array}$$

where the bottom horizontal arrow is the map $[z] \mapsto [\langle d \rangle z]$ and the vertical arrows are the isomorphism of Proposition 7.

For the application to L-functions toward which we are heading it is enough to consider the curve $X_1(N)$, and henceforth $X_0(N)$ will drop out of sight.

2.6. Hecke correspondences and the Frobenius automorphism. Let p be a prime not dividing N, and fix a prime ideal $\mathfrak{p}$ of $\overline{\mathbb{Q}}$ lying over p. Write $\overline{\mathbb{F}}_p$ for the residue class field of $\mathfrak{p}$. Reduction modulo $\mathfrak{p}$ will be denoted by a tilde: for example, if $\mathcal{E}$ is an elliptic curve over $\overline{\mathbb{Q}}$ with good reduction at $\mathfrak{p}$ then $\tilde{\mathcal{E}}$ denotes the elliptic curve over $\overline{\mathbb{F}}_p$ obtained from $\mathcal{E}$ by reduction modulo $\mathfrak{p}$. We define sets $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})$ and $\mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p)$ by replacing $\mathbb{C}$ by $\overline{\mathbb{Q}}$ or $\overline{\mathbb{F}}_p$ respectively in the definition of $\mathrm{Ell}_1(N)(\mathbb{C})$. Thus $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})$ can be identified with the subset of $\mathrm{Ell}_1(N)(\mathbb{C})$ consisting of classes $[\mathcal{E}, \mathcal{P}]$ such that $j(\mathcal{E}) \in \overline{\mathbb{Q}}$. Also, we let $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{gd}}$ denote the subset of $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})$ consisting of classes $[\mathcal{E}, \mathcal{P}]$ such that $\mathcal{E}$ has good reduction at $\mathfrak{p}$. Under our assumption that p does not divide N we have a well-defined map

$$\begin{aligned} \mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{gd}} &\longrightarrow \mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p) \\ [\mathcal{E}, \mathcal{P}] &\longmapsto [\tilde{\mathcal{E}}, \tilde{\mathcal{P}}], \end{aligned}$$

because reduction modulo $\mathfrak{p}$ is injective on N-torsion.

We recall that an elliptic curve $\mathcal{E}$ over $\overline{\mathbb{Q}}$ is said to have *ordinary* good reduction at $\mathfrak{p}$ if $\tilde{\mathcal{E}}[p]$ has order p. If $\mathcal{E}$ has ordinary reduction at $\mathfrak{p}$ then reduction modulo $\mathfrak{p}$ defines a surjective map

$$\mathcal{E}[p] \longrightarrow \tilde{\mathcal{E}}[p]$$

with kernel a subgroup of $\mathcal{E}[p]$ of order (or index) p. In addition $\mathcal{E}[p]$ has exactly p other subgroups of index p.

We use a superscript p to indicate the image of an object under the Frobenius automorphism of $\overline{\mathbb{F}}_p$ and a superscript p^{-1} to indicate the image under the inverse of the Frobenius automorphism.

Proposition 12. *Let $\mathcal{E}$ be an elliptic curve over $\overline{\mathbb{Q}}$ with ordinary reduction at $\mathfrak{p}$, and let Λ_0 be the kernel of the reduction map*

$$\mathcal{E}[p] \longrightarrow \tilde{\mathcal{E}}[p].$$

Let $\mathcal{P}$ be a point of order N on $\mathcal{E}$. If Λ is a subgroup of $\mathcal{E}[p]$ of index p, then

$$[\widetilde{\mathcal{E}/\Lambda}, \widetilde{\mathcal{P}+\Lambda}] = \begin{cases} [\tilde{\mathcal{E}}^p, \tilde{\mathcal{P}}^p] & \text{if } \Lambda = \Lambda_0, \\ [\tilde{\mathcal{E}}^{p^{-1}}, p\tilde{\mathcal{P}}^{p^{-1}}] & \text{if } \Lambda \neq \Lambda_0. \end{cases}$$

Proof. Let $\lambda_0 : \mathcal{E} \to \mathcal{E}/\Lambda_0$ be an isogeny with kernel Λ_0 and $\mu_0 : \mathcal{E}/\Lambda_0 \to \mathcal{E}$ the dual isogeny. The image of $(\mathcal{E}/\Lambda_0)[p]$ under μ_0 is Λ_0. Indeed, since μ_0 is a p-isogeny, the image of $(\mathcal{E}/\Lambda_0)[p]$ under μ_0 is a group of order p; but if this group were not contained in Λ_0 then the image of $(\mathcal{E}/\Lambda_0)[p]$ under $\lambda_0 \circ \mu_0$ would not be zero, contradicting the fact that $\lambda_0 \circ \mu_0$ is multiplication by p. Now consider the commutative diagram

$$\begin{array}{ccccc} \mathcal{E}[p] & \xrightarrow{\lambda_0} & (\mathcal{E}/\Lambda_0)[p] & \xrightarrow{\mu_0} & \mathcal{E}[p] \\ \downarrow & & \downarrow & & \downarrow \\ \tilde{\mathcal{E}}[p] & \xrightarrow{\widetilde{\lambda_0}} & \widetilde{\mathcal{E}/\Lambda_0}[p] & \xrightarrow{\widetilde{\mu_0}} & \tilde{\mathcal{E}}[p], \end{array}$$

where the vertical arrows are reduction modulo $\mathfrak{p}$ and hence surjective. Since the image of $(\mathcal{E}/\Lambda_0)[p]$ under μ_0 is Λ_0, the commutativity of the right-hand square shows that $\widetilde{\mu_0}$ is zero on $\widetilde{\mathcal{E}/\Lambda_0}[p]$. Now $\tilde{\mathcal{E}}$ is ordinary by assumption, and since $\widetilde{\mathcal{E}/\Lambda_0}$ is isogenous to $\tilde{\mathcal{E}}$, it too is ordinary. Hence the fact that $\widetilde{\mu_0}$ is zero on $\widetilde{\mathcal{E}/\Lambda_0}[p]$ means that $\widetilde{\mu_0}$ is a *separable* p-isogeny. But multiplication by p is inseparable. Consequently $\widetilde{\lambda_0}$ is an inseparable (and hence purely inseparable) isogeny of degree p, and we can write $\widetilde{\lambda_0} = \theta \circ \beta$, where β is the Frobenius endomorphism of degree p and $\theta : E^p \to \widetilde{\mathcal{E}/\Lambda_0}$ is an isomorphism. Therefore

$$[\widetilde{\lambda_0}(\tilde{\mathcal{E}}), \widetilde{\lambda_0}(\tilde{\mathcal{P}})] = [\beta(\tilde{\mathcal{E}}), \beta(\tilde{\mathcal{P}})].$$

But the left-hand side is $[\widetilde{\mathcal{E}/\Lambda_0}, \widetilde{\mathcal{P} + \Lambda_0}]$, and the right-hand side is $[\tilde{\mathcal{E}}^p, \tilde{\mathcal{P}}^p]$. Hence we get the stated equality.

Now suppose that $\Lambda \neq \Lambda_0$. Choose an isogeny $\lambda : \mathcal{E} \to \mathcal{E}/\Lambda$ with kernel Λ, and consider the curve $\lambda(\mathcal{E})$ $(= \mathcal{E}/\Lambda)$ together with its subgroup $\lambda(\Lambda_0)$ of order p. Since $\lambda(\mathcal{E})$ is isogenous to $\mathcal{E}$, it has ordinary reduction at $\mathfrak{p}$, and consequently the kernel of reduction mod $\mathfrak{p}$ on $\lambda(\mathcal{E})[p]$ is a subgroup of order p. The calculation

$$\widetilde{\lambda(\Lambda_0)} = \tilde{\lambda}(\widetilde{\Lambda_0}) = \tilde{\lambda}(\{0\}) = \{0\}$$

shows that $\lambda(\Lambda_0)$ is contained in this subgroup and hence coincides with it, since both have order p. Therefore we can apply to $\lambda(\mathcal{E})$ and $\lambda(\Lambda_0)$ what we have already proved for $\mathcal{E}$ and Λ_0:

$$[\widetilde{\lambda(\mathcal{E})/\lambda(\Lambda_0)}, \widetilde{\lambda(\mathcal{P}) + \lambda(\Lambda_0)}] = [\widetilde{\lambda(\mathcal{E})}^p, \widetilde{\lambda(\mathcal{P})}^p]. \tag{1}$$

But if $\mu : \lambda(\mathcal{E}) \to \lambda(\mathcal{E})/\lambda(\Lambda_0)$ is any isogeny with kernel $\lambda(\Lambda_0)$, then by definition

$$[\lambda(\mathcal{E})/\lambda(\Lambda_0), \lambda(\mathcal{P}) + \lambda(\Lambda_0)] = [(\mu \circ \lambda)(\mathcal{E}), (\mu \circ \lambda)(\mathcal{P})].$$

We choose μ to be the isogeny dual to λ, and then $[(\mu \circ \lambda)(\mathcal{E}), (\mu \circ \lambda)(\mathcal{P})]$ becomes $[\mathcal{E}, p\mathcal{P}]$. Thus the left-hand side of (1) is $[\tilde{\mathcal{E}}, p\tilde{\mathcal{P}}]$, and (1) can be rewritten

$$[\tilde{\mathcal{E}}, p\tilde{\mathcal{P}}] = [\widetilde{\mathcal{E}/\Lambda}^p, \widetilde{\mathcal{P} + \Lambda}^p]. \tag{2}$$

Applying the inverse Frobenius automorphism to (2) we obtain the stated formula.

Let $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}}$ be the subset of $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{gd}}$ consisting of classes $[\mathcal{E}, \mathcal{P}]$ such that $\mathcal{E}$ has ordinary reduction at $\mathfrak{p}$. Reduction of isomorphism classes

$$\begin{aligned} \mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}} &\longrightarrow \mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p) \\ [\mathcal{E}, \mathcal{P}] &\longmapsto [\tilde{\mathcal{E}}, \tilde{\mathcal{P}}], \end{aligned}$$

extends uniquely to a homomorphism

$$\mathrm{red}_{\mathfrak{p}} : \mathrm{Div}(\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}}) \longrightarrow \mathrm{Div}(\mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p)).$$

Proposition 13. *Let $\sigma_\mathfrak{p} \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be a Frobenius element at $\mathfrak{p}$. Then*

$$T_p = \sigma_\mathfrak{p} + p\langle p\rangle\sigma_\mathfrak{p}^{-1}$$

when both sides are regarded as maps

$$\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}} \longrightarrow \mathrm{Div}(\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}})/\mathrm{Ker}(\mathrm{red}_\mathfrak{p}).$$

Proof. Let $\mathcal{E}$ be an elliptic curve over $\overline{\mathbb{Q}}$ with ordinary reduction at $\mathfrak{p}$, and let $\mathcal{P}$ be a point of order N on $\mathcal{E}$. We must show that $T_p([\mathcal{E},\mathcal{P}])$ and $(\sigma_\mathfrak{p} + p\langle p\rangle\sigma_\mathfrak{p}^{-1})([\mathcal{E},\mathcal{P}])$ have the same image under $\mathrm{red}_\mathfrak{p}$. Now Proposition 8 gives

$$T_p([\mathcal{E},\mathcal{P}]) = \sum_\Lambda [\mathcal{E}/\Lambda, \mathcal{P}+\Lambda],$$

where the sum runs over subgroups $\Lambda \subset \mathcal{E}[p]$ of index p which have trivial intersection with the cyclic group generated by $\mathcal{P}$. On the other hand,

$$(\sigma_\mathfrak{p} + p\langle p\rangle\sigma_\mathfrak{p}^{-1})([\mathcal{E},\mathcal{P}]) = [\mathcal{E}^{\sigma_\mathfrak{p}}, \mathcal{P}^{\sigma_\mathfrak{p}}] + p[\mathcal{E}^{\sigma_\mathfrak{p}^{-1}}, p\mathcal{P}^{\sigma_\mathfrak{p}^{-1}}].$$

Thus we must check that

$$\sum_\Lambda [\widetilde{\mathcal{E}/\Lambda}, \widetilde{\mathcal{P}+\Lambda}] = [\tilde{\mathcal{E}}^p, \tilde{\mathcal{P}}^p] + p[\tilde{\mathcal{E}}^{p^{-1}}, p\tilde{\mathcal{P}}^{p^{-1}}].$$

This follows from Proposition 12, because the sum on the left-hand side has $p+1$ terms, exactly one of which coincides with the kernel of $\mathcal{E}[p] \to \tilde{\mathcal{E}}[p]$.

Given a smooth projective curve X and a correspondence $T = (Z, \varphi, \psi)$ on X, we use the same letter T to denote the endomorphism $\psi_* \circ \varphi^*$ of its Jacobian variety $\mathrm{Jac}(X)$. If $\Bbbk$ is an algebraically closed field then the map on points $\mathrm{Jac}(X)(\Bbbk) \to \mathrm{Jac}(X)(\Bbbk)$ can be obtained from $T : X(\Bbbk) \to \mathrm{Div}(X(\Bbbk))$ by extending the latter map to $\mathrm{Div}(X(\Bbbk))$, restricting to the subgroup $\mathrm{Div}^0(X(\Bbbk))$ of divisors of degree 0, and then passing to divisor classes.

We are concerned with the case $X = X_1(N)$, $T = T_p$. We denote the Jacobian variety $\mathrm{Jac}(X_1(N))$ simply by $J_1(N)$. Also, the automorphism of $J_1(N)$ induced by the diamond automorphism $\langle d\rangle$ of $X_1(N)$ will be denoted by the same symbol $\langle d\rangle$. If ℓ is a prime and n a positive integer, then the ring of endomorphisms of $J_1(N)$ acts on the abelian group $J_1(N)[\ell^n]$. So does $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, because $J_1(N)$ is defined over $\mathbb{Q}$. In sketching a proof of the next statement we shall simply quote what we need from the work of Igusa [10], in particular the fact that $X_1(N)$ has good reduction at primes not dividing N.

Theorem 2. *Let $\sigma_{\mathfrak{p}} \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be a Frobenius element at $\mathfrak{p}$. Then for $\ell \neq p$ and $n \geq 1$,*

$$T_p = \sigma_{\mathfrak{p}} + p\langle p \rangle {\sigma_{\mathfrak{p}}}^{-1}$$

as endomorphisms of $J_1(N)[\ell^n]$.

Proof. Let $\widetilde{X_1(N)}$ denote the reduction of $X_1(N)$ modulo p. As a map on points, the reduction map $X_1(N)(\overline{\mathbb{Q}}) \to \widetilde{X_1(N)}(\overline{\mathbb{F}}_p)$ is compatible with the map $[\mathcal{E}, \mathcal{P}] \mapsto [\tilde{\mathcal{E}}, \tilde{\mathcal{P}}]$ in a sense which we shall now describe.

Let $\mathbb{Z}[t]_{(p)}$ denote the localization of $\mathbb{Z}[t]$ at the prime ideal generated by p. We say that an elliptic curve E over $\mathbb{Q}(t)$ has good reduction at p if there is a generalized Weierstrass equation for E over $\mathbb{Z}[t]_{(p)}$ with discriminant a unit of $\mathbb{Z}[t]_{(p)}$. The reduction of this equation modulo $p\mathbb{Z}[t]_{(p)}$ then defines an elliptic curve $\tilde{E}$ over $\mathbb{F}_p(t)$. Now let E be an elliptic curve over $\mathbb{Q}(t)$ with invariant t and good reduction at p. For example we can take E to be the curve defined by the equation

$$y^2 + xy = x^3 - \frac{36}{t - 1728}x - \frac{1}{t - 1728}$$

of discriminant $t^2/(t - 1728)^3$. Let P be a point of order N on E, let $\tilde{P} \in \tilde{E}[N]$ be its reduction modulo $\mathfrak{p}$, and let $\tilde{K}$ be the fixed field of

$$\{\sigma \in \mathrm{Gal}(\overline{\mathbb{F}_p(t)}/\mathbb{F}_p(t)) : \sigma(P) = \pm P\},$$

where $\overline{\mathbb{F}_p(t)}$ denotes a separable algebraic closure of $\mathbb{F}_p(t)$. Then $\widetilde{X_1(N)}$ is characterized up to isomorphism as the smooth projective curve over $\mathbb{F}_p$ with function field $\tilde{K}$. Furthermore, by viewing $\tilde{E}$ as an elliptic curve over $\tilde{K}$ one obtains a reduction map

$$\begin{aligned} \tilde{E}(\overline{\mathbb{F}}_p)_{S'} &\longrightarrow \mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p)_{S'} \\ x &\longmapsto [(\tilde{E})_x, (\tilde{P})_x] \end{aligned}$$

for any subset S' of $\mathbf{P}^1(\overline{\mathbb{F}})$ containing the places where $\tilde{E}$ has bad reduction. Let S be the inverse image of S' under the reduction map $\mathbf{P}^1(\overline{\mathbb{Q}}) \longrightarrow \mathbf{P}^1(\overline{\mathbb{F}})$. Then the diagram of reduction maps

$$\begin{array}{ccc} X_1(N)(\overline{\mathbb{Q}})_S & \longrightarrow & \mathrm{Ell}_1(N)(\overline{\mathbb{Q}}) \\ \downarrow & & \downarrow \\ \widetilde{X_1(N)}(\overline{\mathbb{F}}_p)_{S'} & \longrightarrow & \mathrm{Ell}_1(N)(\overline{\mathbb{F}}_p) \end{array}$$

commutes.

Henceforth we take S' to be the set of places where $\tilde{E}$ has bad or supersingular reduction. Note that S' is a finite set. The commutativity of the

above diagram allows us to replace $\mathrm{Ell}_1(N)(\overline{\mathbb{Q}})_{\mathrm{ord}}$ and $\mathrm{Ell}_1(N)(\overline{\mathbb{F}})$ in the statement of Proposition 13 by $X_1(N)(\overline{\mathbb{Q}})_S$ and $\widetilde{X_1(N)}(\overline{\mathbb{F}}_p)_{S'}$ respectively.

Now let $\widetilde{J_1(N)}$ denote the reduction of $J_1(N)$ modulo p, identifiable with the Jacobian of $\widetilde{X_1(N)}$. There is a commutative diagram

$$\begin{array}{ccc} \mathrm{Div}^0(X_1(N)(\overline{\mathbb{Q}})_S) & \longrightarrow & J_1(N)(\overline{\mathbb{Q}}) \\ \downarrow & & \downarrow \\ \mathrm{Div}^0(\widetilde{X_1(N)}(\overline{\mathbb{F}}_p)_{S'}) & \xrightarrow{\alpha} & \widetilde{J_1(N)}(\overline{\mathbb{F}}_p) \end{array}$$

in which the vertical arrows are reduction modulo $\mathfrak{p}$ and the horizontal arrows send a divisor to the point on the Jacobian representing its divisor class. Since S' is finite, α is surjective. Let $L \in J_1(N)(\overline{\mathbb{Q}})$ be a torsion point of ℓ-power order; we must show that

$$(T_p - \sigma_{\mathfrak{p}} - p\langle p\rangle\sigma_{\mathfrak{p}}^{-1})(L) = 0.$$

In fact it is enough to show that this equation holds after reduction modulo $\mathfrak{p}$, because reduction mod $\mathfrak{p}$ is injective on ℓ-torsion. Write $\tilde{L} = \alpha(\tilde{D})$ with $D \in \mathrm{Div}^0(X_1(N)(\overline{\mathbb{Q}})_S)$. According to Proposition 13, the point $(T_p - \sigma_{\mathfrak{p}} - p\langle p\rangle\sigma_{\mathfrak{p}}^{-1})(D)$ reduces to 0 modulo $\mathfrak{p}$, and consequently so does the divisor $(T_p - \sigma_{\mathfrak{p}} - p\langle p\rangle\sigma_{\mathfrak{p}}^{-1})(L)$.

3. L-FUNCTIONS

Theorem 2 is at best an approximation to the Eichler-Shimura relations, because it refers only to Frobenius *elements* of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, not to the Frobenius *correspondence* in characteristic p (cf. [19], p. 17, formulas (I) and (II)). Nevertheless, it suffices for the application to L-functions, to which we now turn.

3.1. The Hasse-Weil conjecture. Originally conceived of as an assertion about the zeta function of a smooth projective variety over a number field, the conjecture has since evolved into a more general statement about L-functions of motives. Here we shall restrict our attention to motives of a very special kind, namely motives afforded by H^1 of an abelian variety over $\mathbb{Q}$ and more generally products of such motives with Artin motives. To begin with we take the Artin motive to be trivial. Let A be an abelian variety of dimension g defined over $\mathbb{Q}$, and recall that for every prime number ℓ one has an ℓ-adic representation

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(V_\ell(A)) \cong \mathrm{GL}(2g, \mathbb{Q}_\ell),$$

where $V_\ell(A) = \mathbb{Q}_\ell \otimes_{\mathbb{Z}_\ell} T_\ell(A)$ and $T_\ell(A)$ is the Tate module of A:

$$T_\ell(A) = \varprojlim_n A[\ell^n].$$

We let ρ_ℓ^* denote the *contragredient* representation on the *dual space* $V_\ell^*(A)$ of $V_\ell(A)$. Given a prime number p, one defines a polynomial $P_p(A,t) \in \mathbb{Z}[t]$ by the formula

$$P_p(A,t) = \det\left(1 - t\rho_\ell^*(\sigma_\mathfrak{p}^{-1})|V_\ell^*(A)^{I(\mathfrak{p})}\right),$$

where ℓ is any prime number different from p, $I(\mathfrak{p})$ and $\sigma_\mathfrak{p}$ denote respectively the inertia group and a Frobenius element of some prime ideal $\mathfrak{p}$ of $\overline{\mathbb{Q}}$ lying over p, and

$$V_\ell^*(A)^{I(\mathfrak{p})} = \{v \in V_\ell^*(A) : \rho_\ell^*(g)v = v \text{ for all } g \in I(\mathfrak{p})\}.$$

That $P_p(A,t)$ is independent of the choice of $\mathfrak{p}$ and $\sigma_\mathfrak{p}$ follows by a straightforward verification from the conjugacy under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of the prime ideals lying over a given rational prime. Far deeper is the fact that $P_p(A,t)$ belongs to $\mathbb{Z}[t]$ and is independent of the choice of $\ell \neq p$. Indeed we are able to make this assertion for *all* p, and not just for the p where A has good reduction, precisely because we have confined ourselves to the case of abelian varieties, for which Grothendieck's semistable reduction theorem [9] is available: in the case of an arbitrary smooth projective variety, the analogues of $P_p(A,t)$ – defined using ℓ-adic cohomology groups $H_\ell^i(*)$ rather than the Tate module – are not yet known to be independent of ℓ when p is a prime of bad reduction and $i > 1$ (for $i = 1$ the ℓ-adic cohomology group is dual to the Tate module of the Albanese, so we are back to the case of abelian varieties). Now write

$$P_p(A,t) = \prod_{i=1}^{2g}(1 - \alpha_{i,p}t)$$

with complex numbers $\alpha_{i,p}$. One has

$$|\alpha_{i,p}| \leq \sqrt{p} \quad (1 \leq i \leq 2g)$$

with equality if p is a prime of good reduction, whence the Euler product

$$L(A,s) = \prod_p P_p(A,p^{-s})^{-1}$$

converges in the region $\mathrm{Re}(s) > 3/2$. Another consequence of the semistable reduction theorem is that one can associate to A a well-defined conductor $N(A)$ and sign $W(A) = \pm 1$ (cf. [17]). The definition of the "root number" $W(A)$ requires the theory of local epsilon factors [6].

Conjecture 1. *Put*

$$\Lambda(A,s) = N(A)^{s/2}((2\pi)^{-s}\Gamma(s))^g L(A,s).$$

Then $\Lambda(A,s)$ has an analytic continuation to an entire function of order one satisfying the functional equation

$$\Lambda(A,s) = W(A)\Lambda(A,2-s).$$

It is also useful to have at hand a slightly less precise formulation of the conjecture, evocative of the state of affairs which prevails when $H^1_\ell(A)$ is replaced by the cohomology of an arbitrary smooth projective variety:

Conjecture 1*. *There exist:*

- *a finite set S of prime numbers containing all primes where A has bad reduction,*
- *for each $p \in S$, a polynomial*

$$P_p^*(A,t) = \prod_{i=1}^{2g}(1-\alpha_{i,p}^* t) \in \mathbb{Z}[t]$$

  with $|\alpha_{i,p}^*| < p$ for all i,
- *a positive integer $N^*(A)$, and*
- *a sign $W^*(A) \in \{\pm 1\}$,*

such that if

$$L^*(A,s) = \prod_{p \notin S} P_p(A,p^{-s})^{-1} \cdot \prod_{p \in S} P_p^*(A,p^{-s})^{-1}$$

and

$$\Lambda^*(A,s) = N^*(A)^{s/2}((2\pi)^{-s}\Gamma(s))^g L^*(A,s)$$

then $\Lambda^*(A,s)$ has an analytic continuation to an entire function of order one satisfying the functional equation

$$\Lambda^*(A,s) = W^*(A)\Lambda^*(A,2-s).$$

We have included a bound on $\alpha_{i,p}^*$ in the statement of Conjecture 1* to ensure that if Conjecture 1 is true then $N^*(A)$, $W^*(A)$, and $P_p^*(A,t)$ coincide respectively with $N(A)$, $W(A)$, and $P_p(A,t)$. Indeed for all good p (and hence in particular for all $p \notin S$) we already have the stronger information that $|\alpha_{i,p}| = \sqrt{p}$, so that the stated bound on $\alpha_{i,p}^*$ affords a uniform estimate

$$p > \begin{cases} |\alpha_{i,p}| & (p \notin S) \\ |\alpha_{i,p}^*| & (p \in S); \end{cases}$$

but a remark of Deligne-Serre ([7], p. 515, Lemme 4.9) then shows that $N^*(A)$, $W^*(A)$, and the $P_p^*(A,t)$ are uniquely determined by the functional equation, whence these quantities coincide with the corresponding quantities in Conjecture 1 whenever the latter conjecture is satisfied.

3.2. Modular forms. Quite apart from its significance for the arithmetic of abelian varieties, Conjecture 1 asserts the existence of a class of Dirichlet series with Euler products and functional equations. Such Dirichlet series arise naturally in the theory of modular forms.

Let k be a positive integer. Given a holomorphic function f on $\mathfrak{H}$ and a matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}^+(2, \mathbb{R}),$$

we put

$$(f|_k\gamma)(z) = \frac{\det(\gamma)}{(cz+d)^k} f(\gamma z).$$

This formula defines a right action of $\mathrm{GL}^+(2, \mathbb{R})$ on the space of holomorphic functions on $\mathfrak{H}$. Now let Γ be a subgroup of finite index in $\mathrm{SL}(2, \mathbb{Z})$. A *modular form of weight k for Γ* is a holomorphic function f on $\mathfrak{H}$ satisfying two conditions:

(i) $f|_k\gamma = f$ for $\gamma \in \Gamma$.

(ii) For every $\delta \in \mathrm{SL}(2, \mathbb{Z})$ the function $f|_k\delta$ has a Fourier expansion of the form

$$(f|_k\delta)(z) = \sum_{n \geq 0} a(n) e^{2\pi i n z/M}.$$

If for every $\delta \in \mathrm{SL}(2, \mathbb{Z})$ the coefficient $a(0)$ in (ii) is 0 then f called a *cusp form.* The vector space of modular forms of weight k for Γ will be denoted $\mathrm{M}_k(\Gamma)$ and the subspace of cusp forms $\mathrm{S}_k(\Gamma)$. These spaces are finite-dimensional. We remark in passing that in condition (ii) the phrase "$\delta \in \mathrm{SL}(2, \mathbb{Z})$" can be replaced by "$\delta \in \mathrm{GL}^+(2, \mathbb{Q})$", where $\mathrm{GL}^+(2, \mathbb{Q}) = \mathrm{GL}(2, \mathbb{Q}) \cap \mathrm{GL}^+(2, \mathbb{R})$. This is simply a matter of writing an element of $\mathrm{GL}^+(2, \mathbb{Q})$ as the product of an element of $\mathrm{SL}(2, \mathbb{Z})$ and an upper triangular matrix. It follows in particular that if Γ is normalized by a matrix $\delta \in \mathrm{GL}^+(2, \mathbb{Q})$ then the spaces $\mathrm{M}_k(\Gamma)$ and $\mathrm{S}_k(\Gamma)$ are stable under the map $f \mapsto f|_k\delta$.

Let us now specialize to the case $\Gamma = \Gamma_1(N)$. In this case we denote the spaces $\mathrm{M}_k(\Gamma)$ and $\mathrm{S}_k(\Gamma)$ simply by $\mathrm{M}_k(N)$ and $\mathrm{S}_k(N)$. Furthermore, given a character χ of $(\mathbb{Z}/N\mathbb{Z})^\times$ we let $\mathrm{M}_k(N, \chi)$ and $\mathrm{S}_k(N, \chi)$ be the subspaces of $\mathrm{M}_k(N)$ and $\mathrm{S}_k(N)$ consisting of f such that

$$f|_k\gamma = \chi(d) f$$

for

$$\gamma = \begin{pmatrix} a & b \\ cN & d \end{pmatrix} \in \Gamma_0(N).$$

(Implicit in the notation $\chi(d)$ is the usual identification of characters of $(\mathbb{Z}/N\mathbb{Z})^\times$ with Dirichlet characters modulo N.) Another way to describe

the subspaces $M_k(N,\chi)$ and $S_k(N,\chi)$ is to say that they are the χ-eigenspaces for the "diamond operators" $f \mapsto f|_k\langle d\rangle$. In this approach d denotes an element of $(\mathbb{Z}/N\mathbb{Z})^\times$, and the operator $\langle d\rangle$ is defined by setting

$$f|_k\langle d\rangle = f|_k\gamma$$

for any $\gamma \in \Gamma_0(N)$ which reduces modulo N to a matrix with d as lower right-hand entry. In view of the isomorphism

$$\begin{aligned}\Gamma_0(N)/\Gamma_1(N) &\longrightarrow (\mathbb{Z}/N\mathbb{Z})^\times \\ \text{coset of } \begin{pmatrix} a & b \\ cN & d \end{pmatrix} &\longmapsto d \pmod N\end{aligned}$$

the diamond operators give a well-defined action of $(\mathbb{Z}/N\mathbb{Z})^\times$ on $M_k(N)$ and $S_k(N)$, and consequently we have eigenspace decompositions

$$M_k(N) = \oplus_\chi M_k(N,\chi)$$

and

$$S_k(N) = \oplus_\chi S_k(N,\chi)$$

where χ runs over Dirichlet characters modulo N. Note that if χ is the trivial character then $M_k(N,\chi)$ and $S_k(N,\chi)$ coincide with $M_k(\Gamma_0(N))$ and $S_k(\Gamma_0(N))$ respectively.

Henceforth we restrict our attention to cusp forms. To see why cusp forms give rise to Dirichlet series with functional equations, observe that the matrix

$$W_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$$

normalizes $\Gamma_1(N)$, whence $f|_k W_N \in S_k(N)$ if $f \in S_k(N)$. In fact

$$\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}\begin{pmatrix} a & b \\ cN & d \end{pmatrix}\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}^{-1} = \begin{pmatrix} d & -c \\ -bN & a \end{pmatrix},$$

so that $f|_k W_N \in S_k(N,\overline{\chi})$ if $f \in S_k(N,\chi)$. Now write

$$f(z) = \sum_{n\geq 1} a(n)e^{2\pi inz}$$

and

$$(f|_k W_N)(z) = \sum_{n\geq 1} b(n)e^{2\pi inz},$$

and put

$$A(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)\sum_{n\geq 1} a(n)n^{-s}$$

and

$$B(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)\sum_{n\geq 1} b(n)n^{-s}.$$

As Hecke observed, these Dirichlet series converge absolutely in some right half-plane and can be analytically continued by a method which goes back to Riemann's paper on the Riemann zeta function: The usual interchange of summation and integration shows that

$$A(s) = \int_0^\infty f(it/\sqrt{N})t^s \frac{dt}{t},$$

whence

$$\begin{aligned} A(s) &= \int_0^1 f(it/\sqrt{N})t^s \frac{dt}{t} + \int_1^\infty f(it/\sqrt{N})t^s \frac{dt}{t} \\ &= \int_1^\infty \left(f(i/(t\sqrt{N}))t^{-s} + f(it/\sqrt{N})t^s\right)\frac{dt}{t} \end{aligned}$$

on making the change of variables $t \mapsto 1/t$ in the integral from 0 to 1. Since

$$f\left(\frac{i}{t\sqrt{N}}\right) = (it)^k (f|_k W_N)(it/\sqrt{N})$$

we obtain

$$A(s) = \int_1^\infty \left(i^k (f|_k W_N)(it/\sqrt{N})t^{k-s} + f(it/\sqrt{N})t^s\right)\frac{dt}{t}.$$

But $(W_N)^2 = -NI$, and consequently $f|_k(W_N)^2 = (-1)^k f$. Hence one can repeat the preceding calculation with $A(s)$ replaced by $B(s)$, f by $f|_k W_N$, and $f|_k W_N$ by $(-1)^k f$, and a comparison of the resulting expressions for $A(s)$ and $B(s)$ yields:

Proposition 14. *The functions $A(s)$ and $B(s)$ have analytic continuations to entire functions of order one satisfying the functional equation $A(s) = i^k B(k-s)$.*

We have avoided calling the Dirichlet series $\sum a(n)n^{-s}$ and $\sum b(n)n^{-s}$ as L-functions, because as yet we have imposed no condition to guarantee the existence of an Euler product. For this we need the Hecke operators.

3.3. Hecke operators. Given a prime number p, let Δ_p denote the set of 2×2 matrices with integer coefficients and determinant p which are congruent modulo N to a matrix of the form

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

It is immediate from the definition that Δ_p is stable under left and right multiplication by $\Gamma_1(N)$, and elementary calculations show that if $\Gamma = \Gamma_1(N)$ and

$$\delta_p = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$$

then Δ_p has the one-term double-coset decomposition

$$\Delta_p = \Gamma \delta_p \Gamma$$

and the following decomposition as a disjoint union of right cosets:

$$\Delta_p = \begin{cases} \bigcup_{\nu=0}^{p-1} \Gamma \begin{pmatrix} 1 & \nu \\ 0 & p \end{pmatrix} \ \cup \ \Gamma\langle p \rangle \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } p \nmid N \\ \bigcup_{\nu=0}^{p-1} \Gamma \begin{pmatrix} 1 & \nu \\ 0 & p \end{pmatrix}, & \text{if } p | N \end{cases}$$

(recall that if p does not divide N then $\langle p \rangle$ denotes an arbitrary element of $\Gamma_0(N)$ with lower right-hand entry congruent to p modulo N). Of course if $\gamma \in \Gamma_1(N)$ and $\{\delta\}$ is any set of representatives for the distinct right cosets of $\Gamma_1(N)$ in Δ_p then $\{\delta\gamma\}$ is another such set, because Δ_p is stable under right multiplication by $\Gamma_1(N)$.

The p-th Hecke operator

$$T_p : \mathrm{S}_k(N) \longrightarrow \mathrm{S}_k(N)$$

is defined by the formula

$$f|_k T_p = p^{k/2-1} \sum_{\delta} f|_k \delta,$$

where δ runs over a set of representatives for the distinct right cosets of $\Gamma_1(N)$ in Δ_p. The definition is independent of the choice of coset representatives because $f \in \mathrm{S}_k(N)$. Furthermore, $f|_k T_p$ does belong to $\mathrm{S}_k(N)$, because right multiplication by any $\gamma \in \Gamma_1(N)$ sends one set of right coset representatives to another. For much the same reason, T_p commutes with the diamond operators $\langle d \rangle$, whence each subspace $\mathrm{S}_k(N, \chi)$ is stable under T_p: since $\Gamma_0(N)$ normalizes both Δ_p and $\Gamma_1(N)$, conjugation by an element of $\Gamma_0(N)$ sends one set of right coset representatives for $\Gamma_1(N)$ in Δ_p to another. To exhibit the effect of T_p on Fourier expansions, suppose that $f \in \mathrm{S}_k(N, \chi)$ and write

$$f(z) = \sum_{n \geq 1} a(n) q^n$$

with $q = e^{2\pi i z}$. A straightforward calculation using the right coset representatives listed above gives

$$(f|_k T_p)(z) = \sum_{n \geq 1} a(pn) q^n + \chi(p) p^{k-1} \sum_{n \geq 1} a(n) q^{pn}.$$

Note that if p divides N then $\chi(p)$ is to be interpreted as 0 in keeping with the usual conventions for Dirichlet characters modulo N. In the literature T_p is often denoted U_p in this case and the preceding formula is written

$$(f|_k U_p)(z) = \sum_{n\geq 1} a(pn)q^n \qquad (p|N).$$

The notation U_p has the advantage of forestalling an ambiguity which in principle could arise when $N = pM$, $p \nmid M$, and $f \in \mathrm{S}_k(M)$: in this situation the expression $f|_k T_p$ can have two possible meanings depending on whether we regard f as belonging to $\mathrm{S}_k(M)$ or to $\mathrm{S}_k(N)$. Nevertheless, we shall continue to use the notation T_p for all primes p, leaving the appropriate interpretation to context.

By a *Hecke eigenform* we shall mean a nonzero element of $\mathrm{S}_k(N,\chi)$ which is an eigenvector of the operators T_p for all primes p. If $f = \sum a(n)q^n$ is a Hecke eigenform and λ_p is the eigenvalue of T_p on f, then the above formula for $f|_k T(p)$ gives

$$a(pn) - \lambda_p a(n) + \chi(p)p^{k-1}a(n/p) = 0 \qquad (n \geq 1),$$

where $a(n/p)$ is understood to be 0 if n is not divisible by p. Taking $n = 1$ we see that $a(p) = \lambda_p a(1)$, so that $a(1) = 0$ implies $a(p) = 0$. More generally, using induction on the total number of prime factors of n one finds that if $a(1) = 0$ then $a(n) = 0$ for all $n \geq 1$, whence $f = 0$. Therefore:

Proposition 15. *If $f = \sum_{n\geq 1} a(n)q^n$ is a Hecke eigenform then $a(1) \neq 0$.*

A Hecke eigenform $f = \sum a(n)q^n$ is said to be *normalized* if $a(1) = 1$. The proposition implies that if f is any Hecke eigenform then some scalar multiple of f is normalized. For a normalized eigenform the relation $a(p) = \lambda_p a(1)$ becomes $\lambda_p = a(p)$, whence the recursion formula for $a(n)$ becomes

$$a(pn) - a(p)a(n) + \chi(p)p^{k-1}a(n/p) = 0.$$

Taking $n = p^{\nu-1}$ with $\nu \geq 1$ one sees that

$$a(p^\nu) - a(p)a(p^{\nu-1}) + \chi(p)p^{k-1}a(p^{\nu-2}) = 0,$$

and then taking $n = p^{\nu-1}m$ with m relatively prime to p one deduces by induction on ν that $a(p^\nu m) = a(p^\nu)a(m)$. A further induction on the number of distinct prime factors of some l relatively prime to m shows that $a(lm) = a(l)a(m)$. In other words, the function $n \mapsto a(n)$ is multiplicative; the associated formal Dirichlet series has an Euler product:

$$\sum_{n\geq 1} a(n)n^{-s} = \prod_p (\sum_{\nu\geq 0} a(p^\nu)p^{-\nu s}).$$

On the other hand, the recursion relation for $a(p^\nu)$ amounts to the formal identity

$$\sum_{\nu\geq 0} a(p^\nu)p^{-\nu s} = (1 - a(p)p^{-s} + \chi(p)p^{k-1}p^{-2s})^{-1},$$

and substitution in the preceding equation gives one direction of the following equivalence (the other is obtained by reversing the argument):

Proposition 16. *For an element* $f(z) = \sum_{n\geq 1} a(n)e^{2\pi inz}$ *of* $S_k(N)$, *the following are equivalent:*

(i) f *is a normalized Hecke eigenform.*
(ii) $\sum_{n\geq 1} a(n)n^{-s} = \prod_p (1 - a(p)p^{-s} + \chi(p)p^{k-1-2s})^{-1}$.

If f is a normalized Hecke eigenform then the Dirichlet series in (ii) is called the L-function of f and denoted $L(f,s)$. ¿From Proposition 14 we know that there is a functional equation relating the L-function of f to a Dirichlet series associated to $f|_k W_N$, but we do not know that the latter Dirichlet series has an Euler product. Thus it remains to find conditions under which both f and $f|_k W_N$ are Hecke eigenforms. Such conditions are provided by the theory of new forms. The starting point is to define a suitable inner product on $S_k(N)$.

3.4. The Petersson inner product. Put

$$S_k = \bigcup_{\Gamma} S_k(\Gamma),$$

where the union is taken over all subgroups of finite index in $\mathrm{SL}(2,\mathbb{Z})$. We define an inner product $(*,*)$ on S_k as follows. Given $f, g \in S_k$, choose a subgroup Γ of finite index in $\mathrm{SL}(2,\mathbb{Z})$ such that f and g both belong to $S_k(\Gamma)$, and put

$$(f,g) = [\mathrm{SL}(2,\mathbb{Z}) : \Gamma]^{-1} \int_{\Gamma\backslash \mathrm{GL}^+(2,\mathbb{R})} (f|_k r)(i)\overline{(g|_k r)(i)}\, dr,$$

where dr denotes the measure on $\Gamma\backslash\mathrm{GL}^+(2,\mathbb{R})$ afforded by a Haar measure on $\mathrm{GL}^+(2,\mathbb{R})$ (recall that $\mathrm{GL}^+(2,\mathbb{R})$ is unimodular – a left Haar measure is a right Haar measure). Using the fact that f and g are cusp forms, one can check that the integral is absolutely convergent. Furthermore, the factor $[\mathrm{SL}(2,\mathbb{Z}) : \Gamma]^{-1}$ in front of the integral guarantees that the value of (f,g) is independent of the choice of Γ. Now if

$$r = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}$$

with $y > 0$ then

$$(f|_k r)(i)\overline{(f|_k r)(i)} = |f(x+iy)|^2 y^k,$$

and consequently $(f,f) > 0$ if $f \neq 0$. Thus $(*,*)$ is in fact an inner product. Since we have not specified a choice of Haar measure on $\mathrm{GL}^+(2,\mathbb{R})$, we have defined $(*,*)$ only up to a scalar multiple; this suffices for our purposes.

Next we observe that if $f, g \in S_k$ and $\delta \in \mathrm{GL}^+(2,\mathbb{Q})$ then $f|_k\delta$ and $g|_k\delta^{-1}$ both belong to S_k and

$$(f|_k\delta, g) = (f, g|_k\delta^{-1}).$$

Indeed choose Γ of finite index in $\mathrm{SL}(2,\mathbb{Z})$ so that $f, g \in \mathrm{S}_k(\Gamma)$. Then the groups $\Gamma' = \Gamma \cap \delta^{-1}\Gamma\delta$ and $\Gamma'' = \Gamma \cap \delta\Gamma\delta^{-1}$ also have finite index in $\mathrm{SL}(2,\mathbb{Z})$ and satisfy $\delta\Gamma'\delta^{-1} = \Gamma''$. Hence we can express $(f|_k\delta, g)$ and $(f, g|_k\delta^{-1})$ as integrals over $\Gamma'\backslash\mathrm{GL}^+(2,\mathbb{R})$ and $\Gamma''\backslash\mathrm{GL}^+(2,\mathbb{R})$ respectively, and the stated formula follows from the left-invariance of Haar measure on $\mathrm{GL}^+(2,\mathbb{R})$. More generally, taking $\gamma, \gamma' \in \Gamma$ and replacing δ by $\delta\gamma'$, we find that

$$(f|_k\gamma\delta\gamma', g) = (f, g|_k\delta^{-1}),$$

because $f|_k\gamma = f$ and $g|_k(\gamma')^{-1} = g$.

Let us apply the preceding formula with $\Gamma = \Gamma_1(N)$ and $\delta \in \Delta_p$, where p is a prime not dividing N. Since Δ_p is equal to a single double coset of Γ, we have $\Delta_p = \Gamma\delta\Gamma$ and consequently

$$(f|_k\delta', g) = (f, g|_k\delta^{-1})$$

for any $\delta' \in \Delta_p$. It follows that

$$(f|_kT_p, g) = (p+1)p^{k/2-1}(f, g|_k\delta^{-1}), \tag{1}$$

because $f|_kT_p$ is the sum of $p+1$ terms of the form $p^{k/2-1}f|_k\delta'$. Take $\delta = \delta_p$ in (1), and as usual, let $\langle p\rangle$ denote any element of $\Gamma_0(N)$ with lower right-hand entry congruent to p modulo N (and hence with upper left-hand entry congruent to p^{-1} modulo N). Since

$$(pI)\langle p\rangle(\delta_p)^{-1} \in \Delta_p$$

formula (1) becomes

$$(f|_kT_p, g) = (p+1)p^{k/2-1}(f, (g|_k\langle p\rangle^{-1})|_k\delta) \tag{2}$$

with some new element δ of Δ_p. On the other hand, repeating a previous argument we see that $(f, (g|_k\langle p\rangle^{-1})|_k\gamma\delta\gamma')$ is independent of $\gamma, \gamma' \in \Gamma$, and consequently that

$$(p+1)p^{k/2-1}(f, (g|_k\langle p\rangle^{-1})|_k\delta) = (f, (g|_k\langle p\rangle^{-1})|_kT_p). \tag{3}$$

Together, (2) and (3) give

$$\check{T}_p = \langle p\rangle^{-1}T_p,$$

where $\check{T}_p$ denotes the adjoint of T_p on $\mathrm{S}_k(N)$ with respect to $(*,*)$. Since the diamond operators commute with the Hecke operators, we conclude that for p not dividing N the operators T_p are normal. We also obtain:

Proposition 17. *Let $f \in S_k(N, \chi)$ be a Hecke eigenform and p a prime not dividing N. If λ_p is the eigenvalue of T_p on f then $\overline{\lambda_p} = \overline{\chi}(p)\lambda_p$.*

As a commuting family of normal operators, the operators T_p ($p \nmid N$) are simultaneously diagonalizable on $S_k(N)$. However, simultaneous diagonalization of the T_p for all primes p, including those dividing N, is a more delicate matter and is possible in general only on a subspace of $S_k(N)$, the subspace of new forms.

3.5. New forms. Consider positive integers M and r such that M divides N properly and r divides N/M, and put

$$V_r = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}.$$

The calculation

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ cN & d \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} a & br \\ cN/r & d \end{pmatrix}$$

shows that the map $f \mapsto f|_k V_r$ sends $S_k(M)$ to $S_k(N)$, indeed each subspace $S_k(M, \chi)$ to the corresponding subspace of $S_k(N)$. In fact a glance at Fourier expansions shows that if p does not divide N then $(f|_k T_p)|_k V_r = (f|_k V_r)|_k T_p$, so that $f \mapsto f|_k V_r$ sends eigenvectors of T_p to eigenvectors of T_p. The need for a distinction between "old forms" and "new forms" arises because this last assertion fails for p dividing N.

The *space of old forms of level N* is by definition the subspace $S_k(N)^{\text{old}}$ of $S_k(N)$ spanned by the images of the maps $f \mapsto f|_k V_r$ as M and r vary over all integers satisfying the divisibility conditions stated above. In other words,

$$S_k(N)^{\text{old}} = \text{span}\left(\bigcup_{\substack{M|N \\ M<N}} \bigcup_{r|N/M} \{f|_k V_r : f \in S_k(M)\} \right).$$

A Hecke eigenform belonging to $S_k(N)^{\text{old}}$ is called an *old form of level N*. The *space of new forms of level N*, denoted $S_k(N)^{\text{new}}$, is the orthogonal complement of $S_k(N)^{\text{old}}$ in $S_k(N)$ relative to the Petersson inner product. A Hecke eigenform belonging to $S_k(N)^{\text{new}}$ is called a *new form of level N*, and a normalized new form of level N is called a *primitive form of level N*. Let $\text{Prim}_k(N)$ denote the set of primitive forms of weight k and level N. One of the main theorems of the theory of new forms is that $\text{Prim}_k(N)$ is a basis for $S_k(N)^{\text{new}}$; as a corollary one deduces that the set

$$\bigcup_{M|N} \bigcup_{r|N/M} \{f|_k V_r : f \in \text{Prim}_k(M)\}$$

is a basis for all of $S_k(N)$. Results such as these are important to mention here because they show that the theory of new forms is nonvacuous, but for present purposes the result of primary interest is the following theorem, which will lead us to a functional equation for the L-function of a primitive form:

Theorem 3. *Given $f \in \mathrm{Prim}_k(N)$, $g \in S_k(N)$, and a finite set S of prime numbers such that g is an eigenvector of T_p for $p \notin S$, suppose that the eigenvalues of T_p on f and g coincide for $p \notin S$. Then g is a scalar multiple of f.*

For the proof, the reader is referred to the literature on new forms: Atkin-Lehner [2], Casselman [4], Li [14], and Miyake [15]. The application to L-functions starts from the observation that if $f \in S_k(N, \chi)$ and we set

$$\check{f}(z) = \overline{f(-\bar{z})}$$

then $\check{f} \in S_k(N, \overline{\chi})$. This follows from the identity $-\overline{\gamma z} = \gamma'(-\bar{z})$, where $\gamma \in \mathrm{GL}^+(2, \mathbb{R})$ and γ' is obtained from γ by negating the diagonal entries. One also verifies that the map $f \mapsto \check{f}$ is unitary with respect to $(*, *)$ and preserves $S_k(N)^{\mathrm{old}}$, whence it preserves $S_k(N)^{\mathrm{new}}$ as well. Now at the level of Fourier expansions the map $f \mapsto \check{f}$ has the form

$$\sum_{n \geq 1} a(n) e^{2\pi i n z} \longmapsto \sum_{n \geq 1} \overline{a(n)} e^{2\pi i n z}.$$

Hence on applying complex conjugation to the formal identity in part (ii) of Proposition 16, we see that if f is a normalized Hecke eigenform, then so is $\check{f}$. Since $S_k(N)^{\mathrm{new}}$ is stable under $f \mapsto \check{f}$ we conclude that $\mathrm{Prim}_k(N)$ is stable under this map also.

Suppose now that $f \in S_k(N, \chi)$. We shall compare $\check{f}$ and $f|_k W_N$. For a prime p not dividing N, let Δ'_p denote the set consisting of 2×2 matrices with integer coefficients and determinant p which are congruent modulo N to a matrix of the form

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

A calculation shows that

$$W_N \Delta_p W_N^{-1} = \Delta'_p = \Delta_p \langle p \rangle^{-1}.$$

Since W_N normalizes $\Gamma_1(N)$ we deduce that if $\{\delta\}$ is a set of representatives for the distinct right cosets of $\Gamma_1(N)$ in Δ_p then both $\{W_N \delta W_N^{-1}\}$ and $\{\delta \langle p \rangle^{-1}\}$ are sets of representatives for the distinct right cosets of $\Gamma_1(N)$ in Δ'_p. It follows that

$$f|_k W_N|_k T_p = \overline{\chi(p)} f|_k T_p|_k W_N.$$

Thus f is an eigenvector of T_p for all primes p not dividing N then so is $f|_k W_N$, and if λ_p is the eigenvalue of T_p on f then $\lambda'_p = \overline{\chi(p)}\lambda_p$ is the eigenvalue of T_p on $f|_k W_N$. Referring to Proposition 17, we see that $\lambda'_p = \overline{\lambda_p}$, and then Theorem 3 implies that $f|_k W_N$ is a scalar multiple of $\check{f}$. We shall write the scalar in question as $i^{-k}W(f)$, so that

$$f|_k W_N = i^{-k}W(f)\check{f}.$$

Then Proposition 14 gives:

Proposition 18. *Given $f \in \mathrm{Prim}_k(N)$, put*

$$\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s).$$

Then $\Lambda(f,s)$ has an analytic continuation to an entire function of order one satisfying the functional equation $\Lambda(f,s) = W(f)\Lambda(\check{f}, k-s)$.

We have reached the limits of what can be done to suggest a possible connection between modular forms and Conjecture 1 on the basis of formal analytic properties alone. The next step is to make a connection between modular forms and modular curves, or at least between cusp forms of weight 2 and regular differentials on modular curves.

3.6. Differentials and cusp forms of weight 2. To begin with let Γ be any subgroup of finite index in $\mathrm{SL}(2,\mathbb{Z})$ and let π denote the restriction to $\mathfrak{H}$ of the natural map $\mathfrak{H}^* \to \Gamma\backslash\mathfrak{H}^*$. If ω is a regular differential on $\Gamma\backslash\mathfrak{H}^*$ then $\pi^*\omega = f(z)dz$ for some function f on $\mathfrak{H}$. We claim that the functions f which arise in this way are characterized by the following conditions:

(o) f is holomorphic.
(i) $f(\gamma z)d(\gamma z) = f(z)dz$ for $\gamma \in \Gamma$.
(ii) Suppose that $\delta \in \mathrm{SL}(2,\mathbb{Z})$, and let M be a positive integer such that

$$(f \circ \delta)(z+M)d(\delta(z+M)) = (f \circ \delta)(z)d(\delta z)$$

(such an integer exists by (i)). Let F be the holomorphic function on the punctured unit disk $D^\circ = \{q \in \mathbb{C} : 0 < |q| < 1\}$ defined by

$$f(\delta z)\frac{d}{dz}\delta z = F(e^{2\pi i z/M}) \qquad (z \in \mathfrak{H}).$$

Then F extends to a holomorphic function on the full unit disk $D = \{q \in \mathbb{C} : 0 < |q| < 1\}$ vanishing at 0.

Indeed (i) says that the differential $f(z)dz$ on $\mathfrak{H}$ descends to a differential on $\Gamma\backslash\mathfrak{H}$, while (o) is the condition for the descended differential to be holomorphic (at an elliptic fixed point of Γ the equivalence between the

holomorphy of ω and the holomorphy of f requires a small verification). As for (ii), its content is that the descended differential extends holomorphically from $\Gamma\backslash\mathfrak{H}$ to $\Gamma\backslash\mathfrak{H}^*$. Again there is a small verification: if we assume without loss of generality that M is minimal, then the change of variables $w = \delta z$, $q = e^{2\pi i \delta^{-1} w/M}$ defines a local parameter at $\delta\infty$, and condition (ii) is a consequence of the fact that $dw = \dfrac{dq}{q} \cdot \dfrac{d}{dz}\delta z \cdot \dfrac{M}{2\pi i}$. Thus (o), (i), and (ii) do characterize the functions on $\mathfrak{H}$ obtained by pulling back regular differentials from $\Gamma^*\backslash\mathfrak{H}$. Now if $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an element of $\mathrm{GL}^+(2,\mathbb{R})$ then

$$d(\gamma z) = \frac{\det(\gamma)}{(cz+d)^2} dz.$$

Therefore condition (i) can be rewritten

$$f|_2\gamma = f \qquad (\gamma \in \Gamma),$$

while in (ii) the requirement is the existence, for any $\delta \in \mathrm{SL}(2,\mathbb{Z})$, of a Fourier series expansion of the form

$$(f|_2\delta)(z) = \sum_{n\geq 1} a(n) e^{2\pi i n z/M}.$$

Returning to the equation $\pi^*\omega = f(z)dz$, we conclude that as ω runs over the space of regular differentials on $\Gamma\backslash\mathfrak{H}^*$ the function f runs over the space of cusp forms of weight 2 for Γ.

Let us now specialize to the case $\Gamma = \Gamma_1(N)$. We shall write $H^0(\Omega^1_{X_1(N)})$ for the space of regular differentials on $X_1(N)$ defined over $\mathbb{Q}$, and similarly $H^0(\Omega^1_{X_1(N)/\mathbb{C}})$ for the corresponding space over $\mathbb{C}$, so that

$$H^0(\Omega^1_{X_1(N)/\mathbb{C}}) = \mathbb{C} \otimes_{\mathbb{Q}} H^0(\Omega^1_{X_1(N)}).$$

The isomorphism just described gives an identification

$$H^0(\Omega^1_{X_1(N)/\mathbb{C}}) \cong \mathrm{S}_2(\Gamma_1(N)),$$

and on the right-hand side we have an action of the Hecke operators T_p. As we shall now explain, the Hecke correspondences determine operators on the left-hand side (to be denoted T_p also) such that the above isomorphism respects the action of T_p. Quite generally, if $T = (Z, \varphi, \psi)$ is a correspondence on a smooth projective curve X, then T gives rise to the operator

$$\begin{aligned} H^0(\Omega^1_X) &\longrightarrow H^0(\Omega^1_X) \\ \omega &\longmapsto \mathrm{tr}_\varphi(\psi^*\omega), \end{aligned}$$

where tr_φ is the trace on differentials associated to the morphism φ: if $K \subset L$ is the inclusion of function fields afforded by φ then any $\alpha \in H^0(\Omega^1_Z)$ has the form $\alpha = udv$ with $u \in L$ and $v \in K$, and by definition, $\mathrm{tr}_\varphi(\alpha) = \mathrm{tr}_{L/K}(u)dv$. Returning to the case at hand, we see that the Hecke correspondence $T_p = (X_1(N,p), \varphi_p, \psi_p)$ on $X_1(N)$ determines an operator T_p on $H^0(\Omega^1_{X_1(N)})$ and hence by extension of scalars an operator on $H^0(\Omega^1_{X_1(N)/\mathbb{C}})$.

Proposition 19. *The canonical isomorphism*

$$H^0(\Omega^1_{X_1(N)/\mathbb{C}}) \cong \mathrm{S}_2(N)$$

commutes with the action of T_p.

Proof. We begin with a general remark. Suppose that X is a smooth projective curve over a subfield of $\mathbb{C}$ and $T = (Z, \varphi, \psi)$ is a correspondence on X. Let d be the degree of φ. Since the base field is a subfield of $\mathbb{C}$, we can discuss the correspondence T in the language of Riemann surfaces, and in particular we can speak of the local analytic sections $(\varphi^{-1})_{x_0,i}$ $(1 \le i \le d)$ of φ in a neighborhood of some unramified point $x_0 \in X(\mathbb{C})$. Then

$$\mathrm{tr}_\varphi(\alpha) = \sum_{1 \le i \le d} ((\varphi^{-1})_{x_0,i})^* \alpha$$

locally at x_0. Thus for $\omega \in H^0(\Omega^1_{X/\mathbb{C}})$ we have

$$T(\omega) = \sum_{1 \le i \le d} (\varphi_i^{-1})^*_{x_0,i} \psi^* \omega.$$

This formula makes it possible to compute the action of T on differentials directly from a knowledge of the map $T : X(\mathbb{C}) \longrightarrow \mathrm{Div}(X(\mathbb{C}))$. Indeed if the latter map is given locally by

$$x \longmapsto \sum_{1 \le i \le d} (\rho_{x_0,i}(x))$$

with analytic functions $\rho_{x_0,i}$, then after a permutation of indices we have $\rho_{x_0,i} = \psi \circ (\varphi_i^{-1})_{x_0,i}$ and consequently

$$T(\omega) = \sum_{1 \le i \le d} \rho^*_{x_0,i} \omega$$

locally at x_0.

In the case at hand we can identify $X_1(N)(\mathbb{C})$ with $\Gamma_1(N) \backslash \mathfrak{H}^*$, and in the coordinate z of $\mathfrak{H}$ the map $T_p : X_1(N)(\mathbb{C}) \longrightarrow \mathrm{Div}(X_1(N)(\mathbb{C}))$ is given by

$$[z] \longmapsto \begin{cases} \sum_{\nu=0}^{p-1} [(z+\nu)/p] + [\langle p \rangle pz] & \text{if } p \nmid N \\ \sum_{\nu=0}^{p-1} [(z+\nu)/p] & \text{if } p | N \end{cases}$$

(Proposition 9). This map can be written simply as

$$[z] \longmapsto \sum_{\delta} [\delta z],$$

where δ runs over a set of representatives for the right cosets of $\Gamma_1(N)$ in Δ_p. Now we have already observed that if f is a function on $\mathfrak{H}$ and $\gamma \in \mathrm{GL}^+(2,\mathbb{R})$ then $f(\gamma z)d(\gamma z) = (f|_2\gamma)(z)dz$. Thus for $f \in \mathrm{S}_2(N)$ we can write

$$\sum_{\delta} f(\delta z)d(\delta z) = \sum_{\delta} (f|_2\delta)(z)dz = (f|_2 T_p)(z)dz,$$

the factor $p^{k/2-1}$ in the definition of T_p being 1 for $k = 2$. This proves Proposition 19.

Using Proposition 11 one proves the analogous statement for the diamond operators:

Proposition 20. *The canonical isomorphism*

$$H^0(\Omega^1_{X_1(N)/\mathbb{C}}) \cong \mathrm{S}_2(N)$$

commutes with the action of $\langle d \rangle$ *for* $d \in (\mathbb{Z}/N\mathbb{Z})^\times$.

One consequence of Propositions 19 and 20 is that $\mathrm{S}_2(N)$ has a $\mathbb{Q}$-form stable under the operators T_p and $\langle d \rangle$, because $H^0(\Omega^1_{X_1(N)/\mathbb{C}})$ has such a $\mathbb{Q}$-form, namely $H^0(\Omega^1_{X_1(N)})$. It follows that if $\mathrm{S}_2(N,\chi)$ contains a Hecke eigenform on which the operator T_p has eigenvalue λ_p then for any $\sigma \in \mathrm{Aut}(\mathbb{C})$ the space $\mathrm{S}_2(N,\chi^\sigma)$ contains a Hecke eigenform on which the operator T_p has eigenvalue λ_p^σ. Now the Fourier coefficients of a normalized Hecke eigenform are polynomials with integer coefficients in the Hecke eigenvalues and the character values. Hence we can define an action of $\mathrm{Aut}(\mathbb{C})$ on the set of normalized Hecke eigenforms in $\mathrm{S}_2(N)$ by the rule

$$f = \sum_{n\geq 1} a(n)q^n \longmapsto f^\sigma = \sum_{n\geq 1} a(n)^\sigma q^n.$$

We claim that if f is a new form of level N then so is f^σ. Suppose on the contrary that f^σ is an old form. Then there is a proper divisor M of N and an element $g = \sum_{n\geq 1} b(n)q^n$ of $\mathrm{Prim}_2(M)$ such that $a(p)^\sigma = b(p)$ for $p \nmid N$. Then $a(p) = b(p)^{\sigma^{-1}}$ for $p \nmid N$, whence $g^{\sigma^{-1}}$ is a normalized Hecke eigenform in $\mathrm{S}_2(M)$ which has the same Hecke eigenvalues as f for $p \nmid N$. This contradicts Theorem 3, proving the claim. We conclude that the map $f \mapsto f^\sigma$ defines an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathrm{Prim}_2(N)$. Since $\mathrm{Prim}_2(N)$ is finite, it follows that if $f \in \mathrm{Prim}_2(N)$ then the field generated by the Fourier coefficients of f has finite degree over $\mathbb{Q}$. We denote this field $\mathbb{E}_f$.

3.7. The Hecke algebra. Given a smooth projective curve X, we will let $\mathrm{Corr}(X)$ denote the free abelian group on the set of isomorphism classes of correspondences on X. If $T = (Z, \varphi, \psi)$ and $T' = (Z', \varphi', \psi')$ are correspondences on X we define the product of their isomorphism classes $[T'] \cdot [T]$ by the formula

$$[T'] \cdot [T] = \sum_W [\widetilde{W}, \varphi \circ \mathrm{pr}_Z \circ \nu_W, \psi' \circ \mathrm{pr}_{Z'} \circ \nu_W],$$

where W runs over the irreducible components of

$$Z'' = \{(z, z') \in Z \times Z' : \psi(z) = \varphi'(z')\},$$

$\nu_W : \widetilde{W} \to W$ is the normalization map, and $\mathrm{pr}_Z : Z'' \to Z$, $\mathrm{pr}_{Z'} : Z'' \to Z'$ are the projections. Extending this product to $\mathrm{Corr}(X)$ by $\mathbb{Z}$-linearity, we make $\mathrm{Corr}(X)$ into a $\mathbb{Z}$-algebra. We shall view $\mathrm{Aut}(X)$ as a subgroup of the multiplicative group of $\mathrm{Corr}(X)$ by identifying $\psi \in \mathrm{Aut}(X)$ with the isomorphism class of the correspondence (X, id_X, ψ) on X.

In the case of $X_1(N)$ we are interested in the subalgebra of $\mathrm{Corr}(X_1(N))$ generated over $\mathbb{Z}$ by the isomorphism classes of all Hecke correspondences T_p and all diamond automorphisms $\langle d \rangle$. We denote this subalgebra by $\mathbb{T}$, and refer to it as the Hecke algebra (of level N). Furthermore, we use the same symbol $\mathbb{T}$ and the same term "Hecke algebra" for the image of $\mathbb{T}$ under the canonical embedding of $\mathrm{Corr}(X_1(N))$ in $\mathrm{End}(J_1(N))$, and we likewise identify the opposite algebra $\mathbb{T}^{\mathrm{opp}}$ with its image in $\mathrm{End}(H^0(\Omega^1_{X_1(N)}))$. Alternatively, we can view $\mathbb{T}$ itself as acting on the dual space of $H^0(\Omega^1_{X_1(N)})$, or we can consider $\mathbb{T}$ to be acting on $H^0(\Omega^1_{X_1(N)})$ on the right. This last point of view is consistent with our identification of $H^0(\Omega^1_{X_1(N)/\mathbb{C}})$ with $\mathrm{End}(\mathrm{S}_2(N))$ (Propositions 19 and 20), and we may therefore think of $\mathbb{T}$ as the subring of $\mathrm{End}(\mathrm{S}_2(N))$ generated over $\mathbb{Z}$ by the Hecke operators and diamond operators on $\mathrm{S}_2(N)$. It follows in particular that $\mathbb{T}$ is commutative, so that $\mathbb{T}$ and $\mathbb{T}^{\mathrm{opp}}$ are canonically isomorphic and every left $\mathbb{T}$-module is a right $\mathbb{T}$-module.

The next step is to associate a quotient ring $\mathbb{T}_f$ of $\mathbb{T}$ to each $f \in \mathrm{Prim}_2(N)$. Consider the ring homomorphism $\lambda_f : \mathbb{T} \to \mathbb{C}$ such that $f|_2 T = \lambda_f(T) f$ for $T \in \mathbb{T}$, and let $\mathbb{I}_f$ be the kernel of λ_f. We set

$$\mathbb{T}_f = \mathbb{T}/\mathbb{I}_f.$$

Thus $\mathbb{T}_f$ is the quotient of $\mathbb{T}$ by the annihilator ideal of f. Write $f(z) = \sum_{n \geq 1} a(n) q^n$, and recall that $\mathbb{E}_f = \mathbb{Q}(\{a(n) : n \geq 1\})$. If $\mathrm{S}_2(N, \chi)$ is the character space to which f belongs then λ_f induces an isomorphism

$$\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{T}_f \longrightarrow \mathbb{E}_f$$

sending $T_p + \mathbb{I}_f$ to $a(p)$ and $\langle d \rangle + \mathbb{I}_f$ to $\chi(d)$. Let A_f be the abelian variety over $\mathbb{Q}$ defined by

$$A_f = J_1(N)/\mathbb{I}_f J_1(N).$$

The action of $\mathbb{T}$ on $J_1(N)$ induces an action of $\mathbb{T}_f$ on A_f and hence on each $V_\ell(A_f)$.

Proposition 21. *The image of $\mathbb{T}_f$ in $\mathbb{E}_f$ is an order of $\mathbb{E}_f$, and $V_\ell(A_f)$ is a free module of rank two over $\mathbb{Q}_\ell \otimes_{\mathbb{Z}} \mathbb{T}_f$. In particular, A_f is an abelian variety of dimension $[\mathbb{E}_f : \mathbb{Q}]$, and A_f is an elliptic curve if and only if the Fourier coefficients of f are rational.*

Proof. The second statement is contained in the first because $V_\ell(A_f)$ is a vector space of dimension $2\dim(A_f)$ over $\mathbb{Q}_\ell$, while

$$\dim_{\mathbb{Q}_\ell} \mathbb{Q}_\ell \otimes_{\mathbb{Z}} \mathbb{T}_f = \operatorname{rank}_{\mathbb{Z}} \mathbb{T}_f = [\mathbb{E}_f : \mathbb{Q}].$$

To prove the first statement we start with the observation that as a subring of $\operatorname{End}(J_1(N))$, the Hecke algebra $\mathbb{T}$ acts on $H_1(J_1(N)(\mathbb{C}), \mathbb{Z})$ and consequently also on $H_1(X_1(N)(\mathbb{C}), \mathbb{Z})$, the two homology groups being isomorphic via the map on homology induced by the embedding of $X_1(N)(\mathbb{C})$ in $J_1(N)(\mathbb{C})$. Denoting the complex dual of $H^0(\Omega^1_{X_1(N)/\mathbb{C}})$ by $H^0(\Omega^1_{X_1(N)/\mathbb{C}})^*$, we see that the standard isomorphism of complex tori

$$J_1(N)(\mathbb{C}) \cong \frac{H^0(\Omega^1_{X_1(N)/\mathbb{C}})^*}{H_1(X_1(N)(\mathbb{C}), \mathbb{Z})}$$

is actually an isomorphism of $\mathbb{T}$-modules. Hence so is the isomorphism

$$J_1(N)(\mathbb{C}) \cong \mathrm{S}_2(N)^*/\Lambda, \tag{1}$$

where Λ is the image of $H_1(X_1(N)(\mathbb{C}), \mathbb{Z})$ when we identify $H^0(\Omega^1_{X_1(N)/\mathbb{C}})^*$ with $\mathrm{S}(2, N)^*$. The fact that the lattice Λ in $\mathrm{S}_2(N)^*$ is stable under $\mathbb{T}$ already shows that the eigenvalues of $\mathbb{T}$ on $\mathrm{S}_2(N)$ are algebraic integers, because eigenvalues are preserved under transpose. It follows that the image of $\mathbb{T}_f$ in $\mathbb{E}_f$ is an order.

Next put

$$V_f = \mathrm{S}_2(N)/(\mathrm{S}_2(N)|_2\mathbb{I}_f),$$

where $\mathrm{S}_2(N)|_2\mathbb{I}_f$ denotes the space of all $g|_2T$ with $g \in \mathrm{S}_2(N)$ and $T \in \mathbb{T}$. We identify V_f^* with the quotient of $\mathrm{S}_2(N)^*$ by $\mathbb{I}_f\mathrm{S}_2(N)^*$, and we let Λ_f be the lattice in V_f^* corresponding to $\Lambda/\mathbb{I}_f\Lambda$ under this identification. Then (1) induces an isomorphism of $\mathbb{T}_f$-modules

$$A_f(\mathbb{C}) \cong V_f^*/\Lambda_f. \tag{2}$$

We claim that V_f (hence also V_f^*) is a free module of rank one over $\mathbb{C} \otimes \mathbb{T}_f$. Granting the claim, we deduce that V_f^* is free of rank two over $\mathbb{R} \otimes \mathbb{T}_f$. Since $V_f^* = \mathbb{R} \otimes \Lambda_f$ it follows that there is a sublattice $\Lambda_f' \subset \Lambda_f$ which is free of rank two over $\mathbb{T}_f$. But (2) gives

$$T_\ell(A_f) \cong \varprojlim_n (\ell^{-n}\Lambda_f)/\Lambda_f \cong \mathbb{Z}_\ell \otimes \Lambda_f.$$

Therefore

$$V_\ell(A_f) \cong \mathbb{Q}_\ell \otimes \Lambda_f \cong \mathbb{Q}_\ell \otimes \Lambda'_f,$$

and the proposition follows.

It remains to prove the claim. The semisimple ring $\mathbb{C} \otimes \mathbb{E}_f$ is canonically a product

$$\mathbb{C} \otimes \mathbb{E}_f = \prod_\sigma \mathbb{C},$$

where the factors are indexed by the distinct embeddings of $\mathbb{E}_f$ in $\mathbb{C}$. Projection onto the factor corresponding to σ gives a character $\mathrm{pr}_\sigma : \mathbb{C} \otimes \mathbb{E}_f \to \mathbb{C}$ sending $T_p + \mathbb{I}_f$ to $a(p)^\sigma$ and $\langle d \rangle + \mathbb{I}_f$ to $\chi(d)^\sigma$, and a simple $\mathbb{C} \otimes \mathbb{E}_f$-module is a one-dimensional complex vector space on which $\mathbb{C} \otimes \mathbb{E}_f$ acts through one of the characters pr_σ. Now as a finitely generated $\mathbb{C} \otimes \mathbb{E}_f$-module V_f is a direct sum of simple modules and is therefore spanned over $\mathbb{C}$ by eigenvectors with eigencharacters of the form pr_σ. Suppose that $v \in V_f$ is such an eigenvector. Then v is in particular an eigenvector for the family of operators $\mathcal{T}_f = \{T_p + \mathbb{I}_f : p \nmid N\}$. But the action of $\mathcal{T}_f$ on $V_f = \mathrm{S}_2(N)/(\mathrm{S}_2(N)|_2\mathbb{I}_f)$ is induced by the action of $\mathcal{T} = \{T_p : p \nmid N\}$ on $\mathrm{S}_2(N)$, and as a commuting family of normal operators $\mathcal{T}$ acts semisimply on $\mathrm{S}_2(N)$. It follows that v is the image in V_f of some $\mathcal{T}$-eigenvector $g \in \mathrm{S}_2(N)$. Then Theorem 3 implies that g is a scalar multiple of one of the cusp forms f^σ. It follows that the restriction to $\oplus_\sigma \mathbb{C} f^\sigma$ of the natural map of $S_2(N)$ onto V_f is surjective. But the restriction is also injective, because

$$(\oplus_\sigma \mathbb{C} f^\sigma) \cap (\mathrm{S}_2(N)|_2\mathbb{I}_f) \subset (\oplus_\sigma \mathbb{C} f^\sigma) \cap (\mathrm{S}_2(N)^{\mathrm{new}}|_2\mathbb{I}_f) = \{0\}$$

by the theory of new forms. Therefore V_f is isomorphic to $\oplus_\sigma \mathbb{C} f^\sigma$ as a $\mathbb{C} \otimes \mathbb{E}_f$-module and is consequently free of rank one.

We are now ready to compute the Euler factor of A_f at a prime of good reduction:

Theorem 4. *Let $f \in \mathrm{S}_2(N, \chi)$ be a primitive cusp form of level N, with Fourier expansion*

$$f(z) = \sum_{n \geq 1} a(n) e^{2\pi i n z}.$$

If p is a prime not dividing N then

$$P_p(A_f, t) = \prod_\sigma (1 - a(p)^\sigma t + \chi(p)^\sigma p t^2),$$

where σ runs over the distinct embeddings of $\mathbb{E}_f$ in $\mathbb{C}$.

Proof. Fix a prime ℓ and let ρ_ℓ denote the natural representation

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow V_\ell(A_f).$$

It will suffice to prove that for a prime p not dividing ℓN we have

$$\det(xI - \rho_\ell(\sigma_\mathfrak{p})) = \prod_\sigma (x^2 - a(p)^\sigma x + \chi(p)^\sigma p), \tag{1}$$

where x is an indeterminate, $\mathfrak{p}$ is a prime ideal of $\overline{\mathbb{Q}}$ lying over p, and $\sigma_\mathfrak{p} \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is the Frobenius automorphism at $\mathfrak{p}$. Indeed the left-hand side of (1) coincides with the characteristic polynomial of $\rho_\ell^*(\sigma_\mathfrak{p}^{-1})$ on $V_\ell(A_f)$, because a matrix and its transpose have the same characteristic polynomial. Also $V_\ell(A_f) = V_\ell(A_f)^{I(\mathfrak{p})}$ by the criterion of Néron-Ogg-Shafarevich [18]. Hence if x is replaced by $1/t$ and the equation multiplied by $t^{2[\mathbb{E}_f:\mathbb{Q}]}$ then (1) becomes the stated formula for $P_p(A_f, t)$, valid for any prime p not dividing $N\ell$. Since ℓ was arbitrary the stated formula follows for any prime p not dividing N.

To prove (1) we recall a fact from linear algebra. Suppose that B is an $(mn) \times (mn)$ matrix which can be written as an $m \times m$ block matrix $B = \left(B^{ij}\right)$ with $n \times n$ blocks B^{ij}. Suppose further that the ring generated over $\mathbb{Z}$ by the matrices B^{ij} is commutative. Then

$$\det B = \det(\det_{m\times m}(B)),$$

where $\det_{m\times m}(B)$ denotes the determinant of the $m \times m$ matrix over $\mathbb{Z}[B^{ij}]$ with ij-entry equal to B^{ij}. On replacing B by $xI - B$ we obtain the formula

$$\det(xI - B) = \det(\det_{m\times m}(xI - B))$$

for the characteristic polynomial of B.

To apply this formula, recall that $V_\ell(A_f)$ is a free module of rank two over $\mathbb{Q}_\ell \otimes \mathbb{T}_f$ and observe that $\rho_\ell(\sigma_\mathfrak{p})$ is a $\mathbb{Q}_\ell \otimes \mathbb{T}_f$-linear transformation of $V_\ell(A_f)$. Let $\det_{\mathbb{Q}_\ell\otimes\mathbb{T}_f}(xI - \rho_\ell(\sigma_\mathfrak{p}))$ denote the characteristic polynomial of $\rho_\ell(\sigma_\mathfrak{p})$ as a $\mathbb{Q}_\ell \otimes \mathbb{T}_f$-linear map. Then

$$\det(xI - \rho_\ell(\sigma_\mathfrak{p})) = N_{\mathbb{Q}_\ell\otimes\mathbb{T}_f/\mathbb{Q}_\ell}(\det_{\mathbb{Q}_\ell\otimes\mathbb{T}_f}(xI - \rho_\ell(\sigma_\mathfrak{p}))),$$

where $N_{\mathbb{Q}_\ell\otimes\mathbb{T}_f/\mathbb{Q}_\ell}$ is the norm from $\mathbb{Q}_\ell \otimes \mathbb{T}_f[x]$ to $\mathbb{Q}_\ell[x]$ (which coincides on $\mathbb{T}_f[x]$ with the norm from $\mathbb{T}_f[x]$ to $\mathbb{Q}[x]$). To prove (1) it suffices to show that

$$\det_{\mathbb{Q}_\ell\otimes\mathbb{T}_f}(xI - \rho_\ell(\sigma_\mathfrak{p})) = x^2 - T_p x + \langle p\rangle p, \tag{2}$$

because $N_{\mathbb{T}_f/\mathbb{Q}}(x^2 - T_p x + \langle p\rangle p)$ is the right-hand side of (1).

Write

$$\mathbb{Q}_\ell \otimes \mathbb{T}_f = \prod_{\lambda|\ell} \mathbb{E}_{f,\lambda},$$

where λ runs over the places of $\mathbb{E}_f$ dividing ℓ and $\mathbb{E}_{f,\lambda}$ idenotes the completion of $\mathbb{E}_f$ at λ. Also put $\rho_\lambda = \mathrm{pr}_\lambda \circ \rho_\ell$, where pr_λ is the projection map

from $\mathbb{Q}_\ell \otimes \mathbb{T}_f$ to the factor $\mathbb{E}_{f,\lambda}$ on the right-hand side. Then equation (2) is equivalent to a system of equations indexed by the places λ, namely the equations

$$\det(xI - \rho_\lambda(\sigma_\mathfrak{p})) = x^2 - T_{p,\lambda}x + \langle p\rangle_\lambda p \tag{3}$$

with $T_{p,\lambda} = \mathrm{pr}_\lambda(T_p)$ and $\langle p\rangle_\lambda = \mathrm{pr}_\lambda(\langle p\rangle)$. It follows from Theorem 2 that the right-hand side of (3) annihilates $\rho_\lambda(\sigma_\mathfrak{p})$. Furthermore, if a nonscalar 2×2 matrix over a field is annihilated by a monic polynomial of degree 2 then that polynomial is its characteristic poynomial. Therefore (3) holds whenever $\rho_\lambda(\sigma_\mathfrak{p})$ is nonscalar. Now fix a place λ_0 dividing ℓ and let P_0 be the set of primes p not dividing $N\ell$ such that $\rho_{\lambda_0}(\sigma_\mathfrak{p})$ is scalar (note that this condition is independent of the choice of $\mathfrak{p}$). It remains to show that (3) holds for $\lambda = \lambda_0$ and all $p \in P_0$.

Let $e_0 \in \mathbb{Q}_\ell \otimes \mathbb{T}_f$ be the idempotent which generates the kernel of the map

$$\prod_{\lambda \neq \lambda_0} \mathrm{pr}_\lambda : \mathbb{Q}_\ell \otimes \mathbb{T}_f \longrightarrow \prod_{\lambda \neq \lambda_0} \mathbb{E}_{f,\lambda},$$

and choose an integer $\nu \geq 0$ such that the element $d_0 = \ell^\nu e_0$ belongs to $\mathbb{Z}_\ell \otimes \mathbb{T}_f$. Since $V_\ell(A_f)$ is free of rank two over $\mathbb{Q}_\ell \otimes \mathbb{T}_f$ it follows that the $\mathbb{Z}_\ell$-module

$$d_0 T_\ell(A_f) \cong \varprojlim_n d_0 A_f[\ell^n]$$

has a $\mathbb{Z}_\ell$-submodule of finite index which is free of rank $2[\mathbb{E}_{f,\lambda_0} : \mathbb{Q}_\ell]$. In particular, putting

$$A_f[\ell^\infty] = \bigcup_{n \geq 1} A_f[\ell^n],$$

we see that $d_0 A_f[\ell^\infty]$ is infinite.

Put $L = \mathbb{Q}(d_0 A_f[\ell^\infty])$. Then the torsion subgroup of $A_f(L)$ contains $d_0 A_f[\ell^\infty]$ and is consequently infinite. Hence a theorem of Ribet [16] implies that L is not contained in the maximal cyclotomic extension of $\mathbb{Q}$. Therefore the group $G = \mathrm{Gal}(L/\mathbb{Q})$ is nonabelian. Let $\mathrm{Frob}_L(P_0)$ be the set of Frobenius elements of prime ideals of L lying over primes in P_0, and let H be the closure of the subgroup of G generated by $\mathrm{Frob}_L(P_0)$. We claim that H is abelian, whence H is a proper subgroup of G. Indeed ρ_{λ_0} can be viewed as a faithful representation of G on $d_0 V_\ell(A_f)$, and since the restriction of ρ_{λ_0} to H is scalar the claim follows. Now let $\overline{P_0}$ be the complement of P_0 in the set of prime numbers not dividing $N\ell$. Also let $\mathrm{Frob}_L(\overline{P_0}) \subset G$ be the set of Frobenius elements of prime ideals of L lying over primes in $\overline{P_0}$. Then the Chebotarev density theorem implies that the set $G - H$ is contained in the closure of $\mathrm{Frob}_L(\overline{P_0})$. Since any group is generated by the complement of a proper subgroup, it follows that the subgroup generated by $\mathrm{Frob}_L(\overline{P_0})$ is dense in G.

Next we consider two continuous homomorphisms $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \mathbb{E}_{f,\lambda_0}^{\times}$. The first, to be denoted κ_{λ_0}, is obtained by composing the ℓ-adic cyclotomic character $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \mathbb{Z}_\ell^{\times}$ with the inclusion of $\mathbb{Z}_\ell^{\times}$ in $\mathbb{E}_{f,\lambda_0}^{\times}$. For the second character, we compose the canonical surjection

$$\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \mathrm{Gal}(\mathbb{Q}(\mu_N)/\mathbb{Q}) \cong (\mathbb{Z}/N\mathbb{Z})^{\times}$$

with the map

$$\begin{aligned}(\mathbb{Z}/N\mathbb{Z})^{\times} &\longrightarrow \mathbb{T}_f^{\times} \\ d &\longmapsto \langle d\rangle\end{aligned}$$

followed by pr_{λ_0}. This second character will be written $\sigma \mapsto \langle\sigma\rangle_{\lambda_0}$. Note that if p is a prime not dividing $N\ell$ then $\kappa_{\lambda_0}(\sigma_{\mathfrak{p}}) = p$ and $\langle\sigma_{\mathfrak{p}}\rangle_{\lambda_0} = \langle p\rangle_{\lambda_0}$. On the other hand, if p happens to belong to $\overline{P_0}$, then $\det\rho_\lambda(\sigma_{\mathfrak{p}}) = \langle p\rangle_{\lambda_0} p$, because equation (3) holds for $\lambda = \lambda_0$ and $p \in \overline{P_0}$. Therefore, writing $\mathrm{Frob}_{\overline{\mathbb{Q}}}(\overline{P_0})$ for the set of Frobenius elements of prime ideals of $\overline{\mathbb{Q}}$ lying over primes in $\overline{P_0}$, we have

$$\det\rho_{\lambda_0}(\sigma_{\mathfrak{p}}) = \langle\sigma_{\mathfrak{p}}\rangle_{\lambda_0}\kappa_{\lambda_0}(\sigma_{\mathfrak{p}})$$

for $\sigma_{\mathfrak{p}} \in \mathrm{Frob}_{\overline{\mathbb{Q}}}(\overline{P_0})$. Since both sides of this equation are continuous, equality holds on the closure of the subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ generated by $\mathrm{Frob}_{\overline{\mathbb{Q}}}(\overline{P_0})$. Let us consider the image of this subgroup under the natural map $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Gal}(L/\mathbb{Q})$. The image of $\mathrm{Frob}_{\overline{\mathbb{Q}}}(\overline{P_0})$ is $\mathrm{Frob}_L(\overline{P_0})$, and we saw above that the subgroup of $\mathrm{Gal}(L/\mathbb{Q})$ generated by $\mathrm{Frob}_L(\overline{P_0})$ is dense in $\mathrm{Gal}(L/\mathbb{Q})$. Thus the closure of the subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ generated by $\mathrm{Frob}_{\overline{\mathbb{Q}}}(\overline{P_0})$ maps onto $\mathrm{Gal}(L/\mathbb{Q})$. We conclude that if p is any prime not dividing $N\ell$ then $\det\rho_{\lambda_0}(\sigma_{\mathfrak{p}}) = \langle p\rangle_{\lambda_0} p$.

We can now prove that equation (3) holds for $\lambda = \lambda_0$ and all $p \in P_0$. Indeed if B is any nonzero 2×2 matrix over a field which is annihilated by a monic polynomial of degree 2, and if the constant term of that polynomial is the determinant of B, then the polynomial is the characteristic polynomial of B. This completes the proof.

3.8. Modular abelian varieties. Let us now complete the train of thought initiated in Theorem 4. Let $f \in \mathrm{S}_2(N,\chi)$ be a primitive cusp form of level N with Fourier expansion

$$f(z) = \sum_{n\geq 1} a(n)e^{2\pi inz}.$$

For a prime p dividing N we define

$$P_p^*(A_f,t) = \prod_{\sigma}(1 - a(p)^{\sigma}t).$$

We also put $g = [\mathbb{E}_f : \mathbb{Q}]$, $N^*(A_f) = N^g$, and $W^*(A_f) = \prod_\sigma W(f^\sigma)$, so that

$$L^*(A_f, s) = \prod_\sigma L(f^\sigma, s)$$

and

$$\Lambda^*(A_f, s) = (N^*(A_f))^{s/2}((2\pi)^{-s}\Gamma(s))^g L^*(A_f, s).$$

Then

$$\Lambda^*(A_f, s) = \prod_\sigma \Lambda(f^\sigma, s).$$

Now according to Proposition 18, each $\Lambda(f^\sigma, s)$ has an analytic continuation to an entire function of order one satisfying the functional equation $\Lambda(f^\sigma, s) = W(f^\sigma)\Lambda(f^{\sigma\rho}, 2-s)$, where $\rho \in \mathrm{Aut}(\mathbb{C})$ denotes complex conjugation. Since composition with ρ merely permutes the distinct embeddings of $\mathbb{E}_f$ in $\mathbb{C}$, we deduce that $\Lambda^*(A_f, s)$ has an analytic continuation to an entire function of order one satisfying the functional equation

$$\Lambda^*(A_f, s) = W^*(A_f)\Lambda^*(A_f, 2-s).$$

Consequently A_f satisfies Conjecture 1*. However, it follows from a theorem of Carayol [3] (completing work of Deligne [5], Ihara [11], and Langlands [12]) that $P_p(A_f, t) = \prod_\sigma (1 - a(p)^\sigma t)$ for p dividing N, and furthermore that $N(A_f) = N^g$ and $W(A_f) = \prod_\sigma W(f^\sigma)$. Thus a stronger assertion holds:

Theorem 5. *For $f \in \mathrm{Prim}_2(N)$ and $g = [\mathbb{E}_f : \mathbb{Q}]$ the invariants*

$$L(A_f, s), \quad N(A_f), \quad \text{and} \quad W(A_f)$$

coincide with

$$\prod_\sigma L(f^\sigma, s), \quad N^g, \quad \text{and} \quad \prod_\sigma W(f^\sigma)$$

respectively, where σ runs over the distinct embeddings of $\mathbb{E}_f$ in $\mathbb{C}$. Consequently A_f satisfies Conjecture 1.

Let Prim_2 denote the union of the sets $\mathrm{Prim}_2(N)$ over all positive integers N, and let A be an abelian variety over $\mathbb{Q}$. If A is isogenous over $\mathbb{Q}$ to a product of abelian varieties of the form A_f with $f \in \mathrm{Prim}_2$, then we call A a *modular abelian variety*, or in the case of dimension one, a *modular elliptic curve*. Since the L-function, conductor, and root number of A depend on A only up to isogeny over $\mathbb{Q}$, and since all three of these invariants respect products, we deduce:

Corollary. *If A is a modular abelian variety then A satisfies Conjecture 1.*

In the remaining paragraphs we discuss a partial converse to the corollary in the case of dimension one, the converse being contingent on a suitable strengthening of Conjecture 1.

3.9. Conjecture 1 with twists. As we have already mentioned, Conjecture 1 is a special case of a more general hypothesis about L-functions of motives. We shall now state a slight extension of Conjecture 1 (still far from the general case) in which we allow twists of the motives in Conjecture 1 by Artin motives. For the application we have in mind it would suffice to consider Artin motives corresponding to Dirichlet characters, but specializing the context in this way does not seem to simplify the formulation.

Consider as before an abelian variety A over $\mathbb{Q}$ together with its associated family of ℓ-adic representations $\{\rho_\ell\}$. In addition, let τ be a continuous finite-dimensional complex representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and let $\mathbb{E}_\tau \subset \mathbb{C}$ be a finite extension of $\mathbb{Q}$ such that τ is realizable on an $\mathbb{E}_\tau$-vector space W. If λ is a place of $\mathbb{E}_\tau$ lying over some ℓ and $\mathbb{E}_{\tau,\lambda}$ is the completion of $\mathbb{E}_\tau$ at λ then we obtain a representation $\rho_\ell \otimes \tau$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the $\mathbb{E}_{\tau,\lambda}$-vector space

$$U_\lambda = (\mathbb{E}_{\tau,\lambda} \otimes_{\mathbb{Q}_\ell} V_\ell(A)) \otimes (\mathbb{E}_{\tau,\lambda} \otimes_{\mathbb{E}_\tau} W).$$

Given a prime p, we choose $\ell \neq p$ and put

$$P_p(A,\tau,t) = \det\left(1 - t(\rho_\ell \otimes \tau)(\sigma_\mathfrak{p})|U_\lambda^{I(\mathfrak{p})}\right).$$

As before, the semistable reduction theorem implies that the coefficients of $P_p(A,\tau,t)$ lie in $\mathbb{E}_\tau$ and are independent of ℓ and λ. Furthermore, the complex numbers $\alpha_{i,p}$ in the factorization

$$P_p(A,\tau,t) = \prod_{i=1}^{2g\dim\tau} (1 - \alpha_{i,p}t).$$

still satisfy

$$|\alpha_{i,p}| \leq \sqrt{p} \quad (1 \leq i \leq 2g\dim\tau),$$

so that the Euler product

$$L(A,\tau,s) = \prod_p P_p(A,\tau,p^{-s})^{-1}$$

converges for $\mathrm{Re}(s) > 3/2$. Also, the conductor $N(A,\tau)$ of the compatible family $\{\rho_\ell \otimes \tau\}_\ell$ is defined, as is the root number $W(A,\tau)$, which is a complex number of absolute value 1 (no longer necessarily equal to ± 1 unless τ is equivalent to its contragredient τ^*). If the conductors $N(A)$ and $N(\tau)$ of A and τ are relatively prime, then

$$N(A,\tau) = N(A)^{\dim\tau} N(\tau)^{2g}$$

and

$$W(A,\tau) = \det\tau((-1)^g N(A)) W(A)^{\dim\tau} W(\tau)^{2g},$$

where in the second equation $W(\tau)$ is the root number of τ and $\det\tau$ is thought of as a Dirichlet character.

Conjecture 2. *Put* $\Lambda(A,\tau,s) = N(A,\tau)^{s/2}((2\pi)^{-s}\Gamma(s))^{g\dim\tau}L(A,\tau,s)$. *Then* $\Lambda(A,\tau,s)$ *has an analytic continuation to an entire function of order one satisfying the functional equation*

$$\Lambda(A,\tau,s) = W(A,\tau)\Lambda(A,\tau^*,2-s).$$

For $A = A_f$ and certain τ with solvable image a statement along these lines follows from the Rankin-Selberg method and the theory of base change (cf. [1], [13], [21]). If $A = A_f$ and τ is one-dimensional then Conjecture 2 is subsumed in the results of Carayol [3].

3.10. Epilogue: the Shimura-Taniyama conjecture. Let us now consider Conjecture 2 in the special case where $\dim A$ and $\dim\tau$ are both one. Thus A is an elliptic curve and τ can be identified with a primitive Dirichlet character χ. We shall further assume that the integers $N = N(A)$ and $r = N(\chi)$ are relatively prime, whence

$$N(A,\chi) = Nr^2$$

and

$$W(A,\chi) = \chi(-N)W(A)W(\chi)^2.$$

In this setting the assertion of Conjecture 2 has a particularly elementary formulation. To begin with, let us put

$$a(p) = \begin{cases} 1 - |\tilde{A}(\mathbb{F}_p)| + p & \text{if } A \text{ has good reduction at } p \\ 1 & \text{if } A \text{ has split multiplicative reduction at } p \\ -1 & \text{if } A \text{ has nonsplit multiplicative reduction at } p \\ 0 & \text{if } A \text{ has additive reduction at } p. \end{cases}$$

Then the Euler factors of A are determined by the elementary rule

$$P_p(A,t) = \begin{cases} 1 - a(p)t + pt^2 & \text{if } A \text{ has good reduction at } p \\ 1 - a(p)t & \text{if } A \text{ has bad reduction at } p. \end{cases}$$

Therefore

$$L(A,s) = \prod_{p\nmid N(A)} (1 - a(p)p^{-s} + p^{1-2s})^{-1} \cdot \prod_{p|N(A)} (1 - a(p)p^{-s})^{-1}.$$

Furthermore, since we are assuming that r is relatively prime to N, the L-function $L(A,\chi,s)$ coincides with the naive twist of $L(A,s)$ by χ: if we write $L(A,s)$ as a Dirichlet series

$$L(A,s) = \sum_{n\geq 1} a(n)n^{-s},$$

then

$$L(A, \chi, s) = \sum_{n \geq 1} \chi(n) a(n) n^{-s}.$$

Thus in the case at hand Conjecture 2 asserts that the function

$$\Lambda(A, \chi, s) = (Nr^2)^{s/2} (2\pi)^{-s} \Gamma(s) \sum_{n \geq 1} \chi(n) a(n) n^{-s}$$

is entire of order one and satisfies the functional equation

$$\Lambda(A, \chi, s) = \chi(-N) W(A) W(\chi)^2 \Lambda(A, \overline{\chi}, 2 - s).$$

Now compare this assertion to condition (i) of the following result, which is a version of Weil's converse to Hecke theory specialized to the case of weight 2 and trivial character:

Theorem 6. *Let N be a positive integer and $a(1), a(2), a(3), \ldots$ a sequence of complex numbers satisfying the formal identity*

$$\sum_{n \geq 1} a(n) n^{-s} = \prod_{p \nmid N} (1 - a(p) p^{-s} + p^{1-2s})^{-1} \cdot \prod_{p \mid N} (1 - a(p) p^{-s})^{-1}.$$

Suppose furthermore that

$$|a(p)| < \begin{cases} 2p & \text{if } p \nmid N \\ p & \text{if } p \mid N, \end{cases}$$

so that the Dirichlet series and Euler product actually converge for $\mathrm{Re}(s) > 2$. *Put*

$$f(z) = \sum_{n \geq 1} a(n) e^{2\pi i n z}.$$

Then the following are equivalent:

(i) *There exists a complex number $W(f)$ of absolute value 1 such that for every positive integer r prime to N and every primitive Dirichlet character χ modulo r, the function*

$$\Lambda(f, \chi, s) = (Nr^2)^{s/2} (2\pi)^{-s} \Gamma(s) \sum_{n \geq 1} \chi(n) a(n) n^{-s}$$

has an analytic continuation to an entire function of order one satisfying the functional equation

$$\Lambda(f, \chi, s) = \chi(-N) W(f) W(\chi)^2 \Lambda(f, \overline{\chi}, 2 - s).$$

(ii) *f is a primitive cusp form of weight 2 for $\Gamma_0(N)$.*

Theorem 6 can be pieced together from Weil [22], Deligne-Serre ([7], p. 515, Lemme 4.9), and the theory of new forms ([2],[4],[14],[15]). It applies in particular to the situation at hand, because if $a(p)$ is the coefficient of p^{-s} in the L-series $L(A,s)$ of an elliptic curve A over $\mathbb{Q}$, then

$$|a(p)| \leq \begin{cases} 2\sqrt{p} & \text{if } p \nmid N \\ 1 & \text{if } p|N, \end{cases}$$

which is a stronger estimate than that required by the hypothesis of the theorem. Thus conditions (i) and (ii) are equivalent for $L(A,s)$, and if we grant Conjecture 2 then it follows that there is a primitive cusp form f for $\Gamma_0(N)$ of weight 2 such that $L(f,s) = L(A,s)$. Now this equation implies in particular that the Fourier coefficients of f are rational, whence $\mathbb{E}_f = \mathbb{Q}$ and A_f is an elliptic curve. Furthermore, Theorem 5 gives $L(A_f,s) = L(A,s)$, and then the isogeny theorem of Faltings implies that A is isogenous over $\mathbb{Q}$ to A_f. Thus A is a modular elliptic curve. To summarize, if we grant Conjecture 2, then we are forced to believe:

Conjecture 3. *Every elliptic curve over* $\mathbb{Q}$ *is modular.*

REFERENCES

1. J. Arthur and L. Clozel, *Simple Algebras, Base Change, and the Advanced Theory of the Trace Formula*, Annals of Math. Studies 120, Princeton Univ. Press, Princeton, 1989.
2. A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134 – 160.
3. H. Carayol, *Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. Ec. Norm. Sup. **19** (1986), 409 – 468.
4. W. Casselman, *On some results of Atkin and Lehner*, Math. Ann. **201** (1973), 301 – 314.
5. P. Deligne, *Formes modulaires et représentations ℓ-adiques*, Séminaire Bourbaki, Lect. Notes in Math. 1799, Springer-Verlag, 1971, pp. 139 – 172.
6. P. Deligne, *Les constantes des équations fonctionelles des fonctions L*, Modular Functions of One Variable, II, Lect. Notes in Math. 349, Springer-Verlag, 1973, pp. 501–595.
7. P. Deligne and J-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. Ec. Norm. Sup. **7** (1974), 507 -530.
8. M. Eichler, *Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion*, Arch. Math. **5** (1954), 355 – 366.
9. A. Grothendieck, *Modèles de Néron et monodromie*, Groupes de Monodromie en Géometrie Algébrique, Lect. Notes in Math. 288, Springer-Verlag, 1971, pp. 313 – 523.
10. J. Igusa, *Kroneckerian model of fields of elliptic modular functions*, Amer. J. Math. **81** (1959).
11. Y. Ihara, *Hecke polynomials as congruence ζ-functions in elliptic modular case*, Ann. Math. **85** (1967).
12. R. P. Langlands, *Modular forms and ℓ-adic representations*, Modular Functions of One Variable, II, Lect. Notes in Math. 349, Springer-Verlag, 1973, pp. 361–500.
13. R. P. Langlands, *Base Change for* GL(2), Annals of Math. Studies 96, Princeton Univ. Press, Princeton, 1980.

14. W. W. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285 – 315.
15. T. Miyake, *On automorphic forms on* GL_2 *and Hecke operators*, Ann. Math. **94** (1971), 174 – 189.
16. K. Ribet, *Torsion points of abelian varieties in cyclotomic extensions*, L'Enseignement Math. **27** (1981), 315 – 319.
17. J-P. Serre, *Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures)*, Séminaire Delange-Poitou-Pisot 1969/70 no. 19.
18. J-P. Serre and J. Tate, *Good reduction of abelian varieties*, Ann. Math. **88** (1968), 492 – 517.
19. G. Shimura, *Correspondances modulaires et les fonctions* ζ *de courbes algébriques*, J. Math. Soc. Japan **10** (1958), 1 – 28.
20. G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Iwanami Shoten and Princeton University Press, Princeton, 1971.
21. J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. AMS **5** (1981), 173 – 175.
22. A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **168** (1967), 149 – 156.