MODULAR POINTS, MODULAR CURVES, MODULAR SURFACES AND MODULAR FORMS

D. Zagier
University of Maryland
College Park, MD 20742
Max-Planck-Institut für Mathematik,
D-5300 Bonn, FRG

This talk, instead of being a survey, will concentrate on a single example, using it to illustrate two themes, each of which has been a leitmotif of much recent work in number theory and of much of the work reported on at this Arbeitstagung (lectures of Faltings, Manin, Lang, Mazur-Soulé, Harder). These themes are:

i) special values of L-series as reflecting geometrical relationships, and

ii) the analogy and interplay between classical algebraic geometry over $\mathbb{C}$ and algebraic geometry (in one dimension lower) over $\mathbb{Z}$, and more especially between the theory of complex surfaces and the theory of arithmetic surfaces à la Arakelov-Faltings. In particular, we will see that there is an intimate relationship between the positions of modular curves in the homology groups of modular surfaces and the positions of modular points in the Mordell-Weil groups of the Jacobians of modular curves.

The particular example we will treat is the elliptic curve E defined by the diophantine equation

$$y(y - 1) = (x + 1)x(x - 1); \tag{1}$$

most of what we have to say applies in much greater generality, but by concentrating on one example we will be able to simplify or sharpen many statements and make the essential points emerge more clearly.

The exposition has been divided into two parts. In the first (§§1-5), which is entirely expository, we describe various theorems and conjectures on elliptic and modular curves, always centering our discussion on the example (1). In particular, we explain how one can construct infinitely many rational solutions of (1) by a construction due to Heegner and Birch, and how a result of Gross and the author and one of Waldspurger lead one to surmise a relationship between these solutions and the coefficients of a modular form of half-integral weight. The second part (§§6-9) is devoted to a proof of this relationship.

I would like to thank G. van der Geer and B. Gross for useful discussions on some of the material in this talk.

1. The elliptic curve E and its L-series.

Multiplying both sides of (1) by 4 and adding 1 we obtain the Weierstrass form

$$y_1^2 = 4x^3 - 4x + 1 \quad (y_1 = 2y - 1); \tag{2}$$

from this one calculates that the curve E has discriminant $\Delta = 37$ and j-invariant $j = 2^{12}3^3/37$. Of course, (1) and (2) are affine equations and we should really work with the projective equations $y^2z - yz^2 = x^3 - xz^2$ and $y_1^2z = 4x^3 - 4xz^2 + z^3$ whose points are the points of (1) or (2) together with a "point at infinity" (0:1:0). The points of E over any field k form a group with the point at infinity being the origin and the group law defined by $P + Q + R = 0$ if P,Q,R are collinear; the negative of a point (x,y) of (1) or (x,y_1) of (2) is $(x,1-y)$ or $(x,-y_1)$, respectively. In accordance with the philosophy of modern geometry, we try to understand E by looking at the groups $E(k)$ of k-rational points for various fields k.

$k = \mathbb{R}$: The set of real solutions of (1) is easily sketched; it consists of two components, $\alpha \le x \le \beta$ and $\gamma \le x$, where $\alpha = -1.107\ldots$, $\beta = 0.2695\ldots$, $\gamma = 0.8395\ldots$ are the roots of $4x^3 - 4x + 1 = 0$ (the group $E(\mathbb{R})$ is isomorphic to $S^1 \times \mathbb{Z}/2\mathbb{Z}$). We have the real period

$$\omega_1 = \int_{E(\mathbb{R})^0} \frac{dx}{y_1} = 2\int_{\gamma}^{\infty} \frac{dx}{\sqrt{4x^3-4x+1}} = 2.993458644\ldots; \tag{3}$$

the numerical value is obtained by using the formula $\omega_1 = \pi/M(\sqrt{\gamma-\alpha}, \sqrt{\gamma-\beta})$, where $M(a,b)$ denotes the arithmetic-geometric mean of Gauss ($M(a,b) = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$ for $a,b > 0$, where $\{a_0,b_0\} = \{a,b\}$, $\{a_{n+1}, b_{n+1}\} = \{\frac{a_n+b_n}{2}, \sqrt{a_nb_n}\}$).

$k = \mathbb{C}$: As well as the real period we have the imaginary period

$$\omega_2 = 2\int_{\beta}^{\gamma} \frac{dx}{\sqrt{4x^3-4x+1}} = 2.451389381\ldots i \tag{4}$$

(which can be calculated as $i\pi/M(\sqrt{\beta-\alpha}, \sqrt{\gamma-\alpha})$). The set of complex points of the (projective) curve E is isomorphic to the complex torus $\mathbb{C}/\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ via the Weierstrass p-function:

$$\mathbb{C}/\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \xrightarrow{\sim} E(\mathbb{C})$$

$$z \longmapsto (p(z), \frac{p'(z)+1}{2}),$$

$$p(z) = \frac{1}{z^2} + \sum_{m,n}{}' \left(\frac{1}{(z-m\omega_1-n\omega_2)^2} - \frac{1}{(m\omega_1+n\omega_2)^2}\right)$$

($\sum_{m,n}'$ means $\sum_{(m,n)\neq(0,0)}$), which satisfies

$$p'^2 = 4p^3 - g_2 p - p_3,$$

$$g_2 = 60 \sum_{m,n}{}' \frac{1}{(m\omega_1+n\omega_2)^4} = \frac{4\pi^4}{3\omega_2^4}\left(1 + 240 \sum_{n=1}^{\infty} \frac{n^3}{e^{2\pi i n\omega_1/\omega_2}-1}\right) = 4,$$

$$g_3 = 140 \sum_{m,n}{}' \frac{1}{(m\omega_1+n\omega_2)^6} = \frac{8\pi^6}{27\omega_2^6}\left(1 - 504 \sum_{n=1}^{\infty} \frac{n^5}{e^{2\pi i n\omega_1/\omega_2}-1}\right) = -1.$$

<u>$k = \mathbb{Q}$</u>: The Mordell-Weil group $E(\mathbb{Q})$ is infinite cyclic with generator $P_0 = (0,0)$, the first few multiples being

$$2P_0 = (1,0),\ 3P_0 = (-1,1),\ 4P_0 = (2,3),\ 5P_0 = (\tfrac{1}{4},\tfrac{5}{8}),\ 6P_0 = (6,-14)$$

and their negatives $-(x,y) = (x,1-y)$. If we write nP_0 as (x_n,y_n) and x_n as N_n/D_n with $(N_n,D_n) = 1$, then

$$\log \max(|N_n|, |D_n|) \sim cn^2 \qquad (|n| \to \infty)$$

with a certain positive constant c (in other words, the number of solutions of (1) for which x has numerator and denominator less than B is asymptotic to $2c^{-1/2}(\log B)^{1/2}$ as $B \to \infty$). This constant is called the <u>height</u> of P_0 and denoted $h(P_0)$; it can be calculated via an algorithm of Tate (cf. [14], [2]) as

$$h(P_0) = \sum_{i=1}^{\infty} 4^{-i-1} \log(1 + 2t_i^2 - 2t_i^3 + t_i^4),$$

where the t_i $(=1/x_{2^i})$ are defined inductively by

$$t_1 = 1,\ t_{i+1} = (1 + 2t_i^2 - 2t_i^3 + t_i^4)/(4t_i - 4t_i^3 + t_i^4),$$

and we find the numerical value

$$h(P_0) = 0.0511114082\ldots\ . \tag{5}$$

Similarly one can define $h(P)$ for any $P \in E(\mathbb{Q})$; clearly $h(nP_0) =$

$n^2 h(P_0)$.

k = $\mathbb{Z}/p\mathbb{Z}$: Finally, we can look at E over the finite field $k = \mathbb{Z}/p\mathbb{Z}$, $p \neq 37$ prime. Here $E(k)$ is a finite group of order $N(p) + 1$, where

$$N(p) = \#\{x,y \pmod p \mid y^2 - y \equiv x^3 - x \pmod p\}.$$

We combine the information contained in all these numbers into the L-series

$$L_E(s) = \prod_{p \neq 37} \frac{1}{1 + \frac{N(p)-p}{p^s} + \frac{p}{p^{2s}}} \cdot \frac{1}{1 + \frac{1}{37^s}}; \tag{6}$$

the special behavior of 37 is due to the fact that $\Delta \equiv 0 \pmod{37}$, so that the reduction of E over $\mathbb{Z}/37\mathbb{Z}$ is singular. Multiplying out, we obtain $L_E(s)$ as a Dirichlet series

$$L_E(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s}, \tag{7}$$

the first few $a(n)$ being given by

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$a(n)$	1	-2	-3	2	-2	6	-1	0	6	4	-5	-6	-2	2	6

. (8)

Since clearly $N(p) \leq 2p$, the product (6) and the sum (7) converge absolutely for $\mathrm{Re}(s) > 2$; in fact, $|N(p)-p|$ is less than $2\sqrt{p}$ (Hasse's theorem), so we have absolute convergence for $\mathrm{Re}(s) > 3/2$. We will see in §3 that $L_E(s)$ extends to an entire function of s and satisfies the functional equation

$$\tilde{L}_E(s) := (2\pi)^{-s} 37^{s/2} \Gamma(s) L_E(s) = -\tilde{L}_E(2-s); \tag{9}$$

in particular, $L_E(s)$ vanishes at $s = 1$. The Birch-Swinnerton-Dyer conjecture relates the invariants of E over $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{Z}/p\mathbb{Z}$ by predicting that

$$\mathrm{ord}_{s=1} L_E(s) = \mathrm{rk}\, E(\mathbb{Q}) = 1$$

and that

$$\frac{d}{ds} L_E(s)\Big|_{s=1} = 2h(P_0) \cdot \omega_1 \cdot S \tag{10}$$

with a certain positive integer S which is supposed to be the order of the mysterious Shafarevich-Tate group Ш. Since the finiteness of

Ш is not known (for E or any other elliptic curve), this last statement cannot be checked. However, $L_E'(1)$ can be computed numerically (cf. §3), and its value 0.3059997738... strongly suggests (cf. (3), (5)) the equation

$$L_E'(1) = 2h(P_0)\omega_1, \tag{11}$$

i.e. (10) with S = 1; the truth of this equation follows from equation (18) below.

2. Twists of L_E; the numbers A(d).

Let p be a prime congruent to 3 (mod 4) which is a quadratic residue of 37 and consider the "twisted" L-series

$$L_{E,p}(s) = \sum_{n=1}^{\infty} \left(\frac{n}{p}\right) \frac{a(n)}{n^s} \tag{12}$$

($\left(\frac{\cdot}{p}\right)$ = Legendre symbol). The proof of the analytic continuation of L_E will also show that each $L_{E,p}$ continues analytically and has a functional equation under $s \to 2-s$. Now, however, the sign of the functional equation is +, so we can consider the value (rather than the derivative) of $L_{E,p}$ at s = 1, and here one can show that

$$L_{E,p}(1) = \frac{2\omega_2}{i\sqrt{p}} A(p)$$

with ω_2 as in (4) and some integer A(p). The value $L_{E,p}(1)$ can be calculated numerically by the rapidly convergent series $L_{E,p}(1) = 2\sum_{n=1}^{\infty} \left(\frac{n}{p}\right) \frac{a(n)}{n} e^{-2\pi n/p\sqrt{37}}$ (cf. §4), so we can compute A(p) for small p. The first few values turn out to be

p	3	7	11	47	67	71	83	107	127	139	151	211	223
A(p)	1	1	1	1	36	1	1	0	1	0	4	9	9

(13)

More generally, $L_{E,d}(s)$ can be defined for all d satisfying $\left(\frac{d}{37}\right)=1$, -d = discriminant of an imaginary quadratic field K (just replace $\left(\frac{\cdot}{p}\right)$ in (12) by $\left(\frac{-d}{\cdot}\right)$, the Dirichlet character associated to $K/\mathbb{Q}$), and we still have

$$L_{E,d}(1) = \frac{2\omega_2}{i\sqrt{d}} A(d) \tag{14}$$

for some $A(d) \in \mathbb{Z}$; the first few values not in (13) are

d	4	40	84	95	104	111	115	120	123	136	148
A(d)	1	4	1	0	0	1	36	4	9	16	9

. (15)

The most striking thing about the values in (13) and (15) is that they are all squares. This is easily understood from the Birch-Swinnerton-Dyer conjecture: the Dirichlet series $L_{E,d}$ is just the L-series of the "twisted" elliptic curve

$$E\langle d\rangle: -dy^2 = 4x^3 - 4x + 1, \tag{16}$$

so A(d) should be either 0 (if $E\langle d\rangle$ has a rational point of infinite order) or (if $E\langle d\rangle(\mathbb{Q})$ is finite) the order of the Shafarevich-Tate group of $E\langle d\rangle$ and hence a perfect square (since this group, if finite, has a non-degenerate $(\mathbb{Q}/\mathbb{Z})$-valued alternating form). Surprisingly, even though we are far from knowing the Birch-Swinnerton-Dyer conjecture or the finiteness of $Ш(E\langle d\rangle)$, we can prove that A(d) is a square for all d, and in fact prove it in two different ways: On the one hand, a theorem of Waldspurger leads to the formula

$$A(d) = c(d)^2, \tag{17}$$

where $c(d)$ $(\in \mathbb{Z})$ is the d^{th} Fourier coefficient of a certain modular form of weight $\frac{3}{2}$. On the other hand, a theorem of Gross and myself gives the formula

$$L_E'(1)L_{E,d}(1) = \frac{4\omega_1\omega_2}{i\sqrt{d}}\, h(P_d) \tag{18}$$

for a certain explicitly constructed point ("Heegner point") P_d in $E(\mathbb{Q})$; writing P_d as b(d) times the generator P_0 of $E(\mathbb{Q})$ and comparing equation (18) with (14) and (11), we obtain

$$A(d) = b(d)^2. \tag{19}$$

We thus have two canonically given square roots b(d) and c(d) of the integer A(d), and the question arises whether they are equal. The object of this paper is to give a geometrical proof of the fact that this is so. First, however, we must define b(d) and c(d) more precisely, and for this we need the modular description of the elliptic curve E, to which we now turn.

3. The modular curve E.

The essential fact about the elliptic curve E is that it is a modular curve. More precisely, let Γ be the subgroup of $SL_2(\mathbb{R})$

generated by the group

$$\Gamma_0(37) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z}) \mid c \equiv 0 \pmod{37}\}$$

and the matrix $w_{37} = \left(\begin{smallmatrix} 0 & -1/\sqrt{37} \\ \sqrt{37} & 0 \end{smallmatrix}\right)$. This group acts on the upper half-plane $\mathcal{H}$ in the usual way and the quotient $\mathcal{H}/\Gamma$ can be compactified by the addition of a single cusp ∞ to give a smooth complex curve of genus 1. We claim that this curve is isomorphic to $E(\mathbb{C})$; more precisely, there is a (unique) isomorphism

$$\mathcal{H}/\Gamma \cup \{\infty\} \xrightarrow{\sim} E(\mathbb{C}) \tag{20}$$

sending ∞ to $0 \in E(\mathbb{C})$ and such that the pull-back of the canonical differential $\frac{dx}{2y-1} = \frac{dx}{y_1}$ is $-2\pi i f(\tau)d\tau$, where

$$f(\tau) = q-2q^2-3q^3+2q^4-2q^5+6q^6-q^7+6q^9+\ldots \quad (q = e^{2\pi i\tau}) \tag{21}$$

is the unique normalized cusp form of weight 2 on Γ, i.e. the unique holomorphic function f on $\mathcal{H}$ satisfying

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^2 f(\tau) \qquad (\tau \in \mathcal{H}, \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma) \tag{22}$$

and $f(\tau) = q + O(q^2)$ as $\mathrm{Im}(\tau) \to \infty$. This claim is simply the assertion of the Weil-Taniyama conjecture for the elliptic curve under consideration, and it is well-known to specialists that the Weil-Taniyama conjecture can be checked by a finite computation for any given elliptic curve; moreover, the particular curve E was treated in detail by Mazur and Swinnerton-Dyer in [11]. Nevertheless, for the benefit of the reader who has never seen an example of a modular parametrization worked out, we will give the details of the proof of (20); our treatment is somewhat different from that in [11] and may make it clearer that the algorithm used would apply equally well to any elliptic curve. The reader who is acquainted with the construction or who is willing to take (20) on faith can skip the rest of this section.

We have two quite different descriptions of the isomorphism (20), depending whether we use the algebraic model (1) or the analytic model $\mathbb{C}/\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ for $E(\mathbb{C})$. We start with the algebraic model. The problem is then to show the existence of two Γ-invariant and holomorphic functions $\xi(\tau)$ and $\eta(\tau)$ satisfying

$$\eta(\tau)^2 - \eta(\tau) = \xi(\tau)^3 - \xi(\tau), \quad -2\pi i f(\tau) = \frac{\xi'(\tau)}{2\eta(\tau)-1} \tag{23}$$

(this gives a map as in (20) with the right pull-back of $\frac{dx}{2y-1}$; that

it is an isomorphism is then easily checked). Equations (23) imply that ξ and η have poles of order 2 and 3, respectively, at ∞, and recursively determine all coefficients of their Laurent expansions. Calculating out to 9 terms, we see that these expansions must begin

$$\begin{aligned} \xi(\tau) &= q^{-2}+2q^{-1}+5+9q+18q^2+29q^3+51q^4+82q^5+131q^6+\dots , \\ \eta(\tau) &= q^{-3}+3q^{-2}+9q^{-1}+21+46q+92q^2+180q^3+329q^4+593q^5+\dots . \end{aligned} \tag{24}$$

So far we have not used the fact that f is a modular form on Γ; we could have taken any power series $f(\tau) = q+\dots$ and uniquely solved (23) to get Laurent series $\xi(\tau) = q^{-2}+\dots$, $\eta(\tau) = q^{-3}+\dots$. However, since ξ and η are supposed to be Γ-invariant functions with no poles in $\mathcal{H}$, and since f is a modular form of weight 2, the two functions $f_4 = f^2\xi$ and $f_6 = f^3\xi$ must be holomorphic modular forms on Γ of weight 4 and 6, respectively. But the space $M_k(\Gamma)$ of modular forms of weight k on Γ is finite-dimensional for any k and one can obtain a basis for it by an algorithmic procedure (e.g., using the Eichler-Selberg trace formulas, but we will find a shortcut here), so we can identify f_4 and f_6 from the beginnings of their Fourier expansions. Once one has candidates f_4 and f_6, one <u>defines</u> $\xi = f_4/f^2$ and $\eta = f_6/f^3$; these are then automatically modular functions on Γ, and the verification of (23) reduces to the verification of the two formulae

$$f_6^2 - f_6 f^3 = f_4^3 - f_4 f^4, \quad f(2f_6 - f^3) = \frac{1}{2\pi i}(2f_4 f' - f f_4') , \tag{25}$$

which are identities between modular forms on Γ (of weights 12 and 8, respectively) and hence can be proved by checking finitely many terms of the Fourier expansions. In our case the dimension of $M_k(\Gamma)$ equals $[\frac{5k}{6}] + 3[\frac{k}{4}]$ for $k > 0$, k even, so $M_2(\Gamma)$ is generated by f while $M_4(\Gamma)$ and $M_6(\Gamma)$ have dimension 6 and 8, respectively. However, we will be able to identify f_4 and f_6 without calculating bases for these spaces. The space $M_2(\Gamma_0(37))$ is the direct sum of $M_2(\Gamma) = \mathbb{C}f$ and the 2-dimensional space of modular forms F of weight 2 on $\Gamma_0(37)$ satisfying $F(-\frac{1}{37\tau}) = -37\tau^2 F(\tau)$. As a basis of this latter space we can choose the theta-series

$$\Theta(\tau) = \sum_{a,b,c,d\in\mathbb{Z}} q^{Q(a,b,c,d)} = 1+2q+2q^2+4q^3+2q^4+4q^5+8q^6+4q^7+10q^8+\dots ,$$

$$\begin{aligned} Q(a,b,c,d) &= \tfrac{1}{8}(4b+c-2d)^2 + \tfrac{1}{4}(2a+c+d)^2 + \tfrac{37}{8}c^2 + \tfrac{37}{4}d^2 \\ &= a^2 + 2b^2 + 5c^2 + 10d^2 + ac + ad + bc - 2bd \end{aligned}$$

and the cusp form

$$h(\tau) = \frac{3}{4}\Theta(\tau) - \frac{1}{2}E_2^-(\tau) = q + q^3 - 2q^4 - q^7 + \dots ,$$

where $E_2^-(\tau) = \frac{3}{2} + \sum_{\substack{d,n>0 \\ 37\nmid d}} dq^{nd}$ is an Eisenstein series. The four functions f^2, Θ^2, Θh and h^2 lie in the space

$$U = \{F \in M_4(\Gamma) \mid \mathrm{ord}_{\tau=A}(F) \geq 2,\ \mathrm{ord}_{\tau=B}(F) \geq 4\},$$

where A and B are the fixed points in $\mathcal{H}/\Gamma_0(37)$ of order 2 and 3, respectively, because any function in $M_2(\Gamma_0(37))$ must vanish at A and vanish doubly at B. For the same reason, $f_4 = f^2\xi$ lies in U (recall that ξ has no poles in $\mathcal{H}$); and since U has codimension 2 in $M_4(\Gamma)$ (a general function in $M_4(\Gamma)$ satisfies $\mathrm{ord}_A F = 2r$, $\mathrm{ord}_B F = 3s+1$ for some $r,s \geq 0$), these five functions must be linearly dependent. Looking at the first few Fourier coefficients, we find that f_4 must be given by

$$f_4(\tau) = (\Theta(\tau) - 3h(\tau))^2 - \frac{37}{4}h(\tau)^2 - \frac{1}{4}f(\tau)^2.$$

As to f_6, we observe that the function

$$\psi(\tau) = \eta(\tau)^2/\eta(37\tau)^2 + 37\eta(37\tau)^2/\eta(\tau)^2$$

is Γ-invariant and holomorphic in $\mathcal{H}$ and has a triple pole at ∞, so must be a linear combination of ξ,η and 1; looking at the first few Fourier coefficients we find that $\psi = \eta - 5\xi + 6$, so f_6 must be $\psi f^3 + 5f_4 f - 6f^3$. As explained above, once we have our candidates f_4 and f_6 it is a finite computation to check (25) and thus establish that $\tau \mapsto (f_4(\tau)f(\tau) : f_6(\tau) : f(\tau)^3)$ maps $\mathcal{H}/\Gamma \cup \{\infty\}$ to $E(\mathbb{C}) \subset \mathbb{P}^2(\mathbb{C})$ as claimed.

For the second description of the map (20), we define a function $\phi:\mathcal{H} \to \mathbb{C}$ by

$$\phi(\tau) = 2\pi i \int_\tau^{i\infty} f(\tau')d\tau' = -q + q^2 + q^3 - \frac{1}{2}q^4 + \frac{2}{5}q^5 - \dots . \qquad (26)$$

From $\phi' = -2\pi i f$ and (22) it follows that the difference $\phi(\frac{a\tau+b}{c\tau+d}) - \phi(\tau)$ is a constant for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Call this constant $C(\gamma)$; clearly $C:\Gamma \to \mathbb{C}$ is a homomorphism. The theory of Eichler-Shimura implies that the image $\Lambda = C(\Gamma)$ is a lattice in $\mathbb{C}$ with $g_2(\Lambda)$ and $g_3(\Lambda)$ rational integers. Since we can calculate $\phi(\tau)$ and hence $C(\gamma)$

numerically (the series in (26) converges rapidly), we can calculate a basis of Λ numerically and get g_2 and g_3 exactly. The result $g_2 = 4$, $g_3 = -1$ shows that Λ is the lattice $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ of §1, and the identity $\phi(\gamma\tau) - \phi(\tau) = C(\gamma)$ shows that $\mathcal{H} \xrightarrow{\phi} \mathbb{C} \to \mathbb{C}/\Lambda$ factors through Γ. We thus obtain a map $\mathcal{H}/\Gamma \xrightarrow{\phi} E(\mathbb{C}) = \mathbb{C}/\Lambda$ such that the pull-back $\phi^*(dz)$ equals $-2\pi i f(\tau)d\tau$, as asserted. In practice, it is easier to calculate the image in E of a particular point $\tau \in \mathcal{H}$ by using (26) and reducing modulo Λ than by using the first description of the map (20).

4. Modular forms attached to E

The most important consequence of the modular description of the elliptic curve E is that the L-series of E equals the L-series of the modular form f, i.e. that the numbers $a(n)$ in (7) are precisely the Fourier coefficients in (21). This follows from the Eichler-Shimura theory (cf. [13]). As a consequence, the function $\tilde{L}_E$ defined in (9) has the integral representation

$$L_E(s) = \int_0^\infty f\left(\frac{it}{\sqrt{37}}\right) t^{s-1} dt = \int_1^\infty f\left(\frac{it}{\sqrt{37}}\right)(t^{s-1} - t^{1-s})dt,$$

from which the analytic continuation and functional equation are obvious. Differentiating and setting $s = 1$ we find

$$\frac{\sqrt{37}}{2\pi} L_E'(1) = \tilde{L}_E'(1) = 2\int_1^\infty f\left(\frac{it}{\sqrt{37}}\right)\log t\, dt = 2\sum_{n=1}^\infty a(n) G\left(\frac{2\pi n}{\sqrt{37}}\right),$$

with

$$G(x) = \int_1^\infty e^{-xt}\log t\, dt = \frac{1}{x}\int_x^\infty e^{iu}\frac{du}{u},$$

and since there are well-known expansions for $G(x)$, this can be used to calculate $L_E'(1) = 0.30599...$ to any desired degree of accuracy, as mentioned in §1. Similarly, if $-d$ is the discriminant of an imaginary quadratic field in which 37 splits, then the "twisted" form $f^*(\tau) = \sum\left(\frac{-d}{n}\right)a(n)q^n$ is a cusp form of weight 2 and level $37d^2$ satisfying $f^*(-1/37d^2\tau) = -37d^2\tau^2 f^*(\tau)$, so

$$\tilde{L}_{E,d}(s) := (2\pi)^{-s}37^{s/2}d^s\Gamma(s)L_{E,d}(s) = \int_1^\infty f^*\left(\frac{it}{d\sqrt{37}}\right)(t^{s-1} + t^{1-s})dt,$$

from which we deduce the functional equation $\tilde{L}_{E,d}(s) = \tilde{L}_{E,d}(2-s)$ and

the formula $L_{E,d}(1) = 2 \sum_{n=1}^{\infty} (\frac{-d}{n}) \frac{a(n)}{n} e^{-2\pi n/d\sqrt{37}}$ mentioned in §2. In particular, we can calculate the numbers $A(d)$ defined by (14) approximately and hence, since they are integers, exactly.

The other modular form which will be important to us is the form of weight 3/2 associated to f under Shimura's correspondence. Around ten years ago, Shimura [12] discovered a relationship between modular forms of arbitrary even weight $2k$ and modular forms of half-integral weight $k + 1/2$. This was studied subsequently by many other authors. In particular, Kohnen (in [8] for forms of level 1 and in [9] for forms of odd squarefree level) showed how Shimura's theory could be refined by imposing congruence conditions modulo 4 on the Fourier expansion so as to get a perfect correspondence between appropriate spaces of forms of weights $2k$ and $k + 1/2$. The result in the case $k = 1$ and prime level is the following ([9], Theorem 2):

Theorem 1 (Shimura; Kohnen). *For N prime and $\varepsilon \in \{\pm 1\}$ let $S^{\varepsilon}_{3/2}(N)$ denote the space of all functions $g(\tau)$ satisfying*

i) $g(\tau)/\theta(\tau)^3$, where $\theta(\tau)$ is the standard theta-series $\sum_{n\in\mathbb{Z}} q^{n^2}$, is invariant under $\Gamma_0(4N)$;

ii) $g(\tau)$ has a Fourier development $\sum_{d>0} c(d)q^d$ with $c(d) = 0$ if $-d \equiv 2$ or $3 \pmod 4$ or $(\frac{-d}{N}) = -\varepsilon$.

Let $S^{\varepsilon}_2(\Gamma_0(N))$ denote the space of cusp forms of weight 2 on $\Gamma_0(N)$ satisfying $f(-1/N\tau) = \varepsilon N\tau^2 f(\tau)$. Then $\dim S^{\varepsilon}_{3/2}(N) = \dim S^{\varepsilon}_2(\Gamma_0(N))$, and for each Hecke eigenform $f = \sum a(n)q^n \in S^{\varepsilon}_2(\Gamma_0(N))$ there is a 1-dimensional space of $g \in S^{\varepsilon}_{3/2}(N)$ whose Fourier coefficients are related to those of f by

$$a(n)c(d) = \sum_{\substack{r\mid n \\ r>0}} (\frac{-d}{r}) c(\frac{n^2}{r^2}d) \qquad (n\in\mathbb{N},\ -d \text{ a fundamental discriminant}). \tag{27}$$

In our case $N = 37$, $\varepsilon = +1$ and the space $S^{+}_2(\Gamma_0(37))$ is one-dimensional, spanned by the function f of (21). Theorem 1 therefore asserts that there is a unique function

$$g(\tau) = \sum_{\substack{d>0 \\ -d \equiv 0 \text{ or } 1 \pmod 4 \\ (-d/37) = 0 \text{ or } 1}} c(d)e^{2\pi i d\tau}$$

such that $g(\tau)/\theta(\tau)^3$ is $\Gamma_0(148)$-invariant and the Fourier coefficients $c(d)$ (normalized, say, by $c(3) = 1$) satisfy (27). It is not an entirely trivial matter to calculate these coefficients; a method

for doing so, and a table up to $d = 250$, were given in [3, pp. 118-120, 145] in connection with the theory of "Jacobi forms." We give a short table:

d	3	4	7	11	12	16	27	28	36	40	44	47	48	63	64	67	71	75	83	...	148
$c(d)$	1	1	-1	1	-1	-2	-3	3	-2	2	-1	-1	0	2	2	6	1	-1	-1	...	-3

(28)

We now come to the theorem of Waldspurger [15], mentioned in §2, which relates these coefficients to the values at $s = 1$ of the twisted L-series $L_{E,d}(s)$. Again we need a refinement due to Kohnen [10, Theorem 3, Cor. 1] which gives a precise and simple identity in the situation of Theorem 1:

Theorem 2 (Waldspurger; Kohnen). *Let* $f = \sum a(n)q^n \in S_2^{\varepsilon}(\Gamma_0(N))$, $g = \sum c(d)q^d \in S_{3/2}^{\varepsilon}(N)$ *correspond as in Theorem 1. Let* $-d$ *be a fundamental discriminant with* $(\frac{-d}{N}) = 0$ *or* ε *and let* $L_{f,d}(s)$ *be the associated convolution L-series* $\sum (\frac{-d}{N})a(n)n^{-s}$. *Then*

$$L_{f,d}(1) = 3\pi \frac{\|f\|^2}{\|g\|^2} \frac{|c(d)|^2}{\sqrt{d}} \tag{29}$$

where

$$\|f\|^2 = \int_{\mathcal{H}/\Gamma_0(N)} |f(\tau)|^2 du\, dv, \qquad \|g\|^2 = \int_{\mathcal{H}/\Gamma_0(4N)} |g(\tau)|^2 v^{-1/2} du\, dv \quad (\tau = u + iv) \tag{30}$$

are the norms of f *and* g *in the Petersson metric.* (Note that the identity is independent of the choice of g, since replacing g by λg ($\lambda \in \mathbb{C}^*$) multiplies both $\|g\|^*$ and $|c(d)|^2$ by $|\lambda|^2$.)

Actually, the exact coefficient in (29) is not too relevant to us, for knowing that $L_{f,d}(1)$ is a fixed multiple of $c(d)^2/\sqrt{d}$ implies that the numbers $A(d)$ defined by (14) are proportional to $c(d)^2$, and calculating $A(3) = c(3)^2 = 1$ we deduce (17). Then going back and substituting (17) and (14) into (29) we deduce $3\pi\|f\|^2/\|g\|^2 = 2\omega_2/i$. We now show (since the result will be needed later) that

$$\|f\|^2 = \omega_1\omega_2/2\pi^2 i, \tag{31}$$

it then follows that $\|g\|^2 = 3\omega_1/4\pi$. To prove (31), we recall from §3 that there is an isomorphism ϕ from $\mathcal{H}/\Gamma \cup \{\infty\}$ to $E(\mathbb{C}) = \mathbb{C}/\Lambda$ with $\phi^*(dz) = 2\pi i f(\tau)d\tau$. Since $[\Gamma:\Gamma_0(37)] = 2$ we have

$$2\pi^2\|f\|^2 = 4\pi^2 \int_{\mathcal{H}/\Gamma} |f(\tau)|^2 du\,dv = \int_{\mathcal{H}/\Gamma} |-2\pi i f(\tau)|^2 du\,dv$$

$$= \int_{\mathbb{C}/\Lambda} dx\,dy = \omega_1\omega_2/i$$

as claimed.

5. Heegner points on E

In this section we describe a construction which associates to each integer $d > 0$ a point $P_d \in E(\mathbb{Q})$. These are the "modular points" of the title, since their construction depends on the modular description of E given in §3.

We assume first that $-d$ is a fundamental discriminant, i.e. the discriminant of an imaginary quadratic field K. We consider points $\tau \in \mathcal{H}$ of the form $\tau = \frac{b+i\sqrt{d}}{2a}$ with

$$a,b \in \mathbb{Z},\ a > 0,\ 37|a,\quad b^2 \equiv -d \pmod{4a}. \tag{32}$$

If $(\frac{-d}{37}) = -1$, there are no such τ and we set $P_d = 0$; otherwise the set of τ is invariant under Γ and there are h distinct points $\tau_1,\dots,\tau_h$ modulo the action of Γ, where $h = h(-d)$ is the class number of K. The theory of complex multiplication shows that these points are individually defined over a finite extension H of $\mathbb{Q}$ (the Hilbert class field of K) and collectively over $\mathbb{Q}$ (i.e. their images in $\mathcal{H}/\Gamma$ are permuted by the action of the Galois group of H over $\mathbb{Q}$). Hence the sum $\phi(\tau_1) + \dots + \phi(\tau_h)$, where $\phi: \mathcal{H}/\Gamma \to E(\mathbb{C})$ is the map constructed in §3, is in $E(\mathbb{Q})$. Moreover this sum is divisible by u, where u is $\frac{1}{2}$ the number of units of K (= 1,2 or 3) if $37\nmid d$ and $u = 2$ if $37|d$; this is because each point $\tau_j \in \mathcal{H}$ is the fixed point of an element of Γ of order u. We define $P_d \in E(\mathbb{Q})$ by

$$uP_d = \sum_{j=1}^{h} \phi(\tau_j); \tag{33}$$

this is well-defined because $E(\mathbb{Q})$ is torsion-free. If d is not fundamental, we define P_d^0 the same way but with the extra condition $(a,b,\frac{b^2+d}{4a}) = 1$ in (32) (now $h(-d)$ is the class number of a certain non-maximal order of K, and the points $\tau_1,\dots,\tau_h \in \mathcal{H}/\Gamma$ are defined over the corresponding ring class field), and then set $P_d = \sum_{e^2|d} P^0_{d/e^2}$.

The definition of P_d just given is a special case of a construc-

tion due to Heegner and Birch (cf. [1]) and in general would yield rational points in the Jacobian of $X_0(N)/w_N$ ($X_0(N) = \mathcal{H}\cup(\text{cusps})/\Gamma_0(N)$). From a modular point of view, a point $\tau \in \mathcal{H}/\Gamma_0(N)/w_N$ classifies isomorphism classes of unordered pairs of N-isogenous elliptic curves $\{E_1,E_2\}$ over $\mathbb{C}$ (namely $E_1 = \mathbb{C}/\mathbb{Z}+\mathbb{Z}\tau$, $E_2 = \mathbb{C}/\mathbb{Z}+N\mathbb{Z}\tau$, with the isogenies $E_1 \to E_2$, $E_2 \to E_1$ induced by $N\cdot id_{\mathbb{C}}$ and $id_{\mathbb{C}}$, respectively), and the points $\tau_1,\dots,\tau_h$ correspond to those with complex multiplication by an order $\mathcal{O}$ of $\mathbb{Q}(\sqrt{-d})$ (namely $E_1 = \mathbb{C}/\mathfrak{a}$, $E_2 = \mathbb{C}/\mathfrak{n}\mathfrak{a}$, where $\mathfrak{a} = \mathbb{Z}+\mathbb{Z}\tau$ is a fractional $\mathcal{O}$-ideal and $\mathfrak{n}$ an intergral $\mathcal{O}$-ideal of norm N). A general formula for the heights of these "Heegner points" was proved recently by B. Gross and myself [4]; the result in our special case becomes

Theorem 3(Gross-Zagier): *Suppose* $-d$ *is a fundamental discriminant with* $(\frac{-d}{37}) = 1$ *and let* $P_d \in E(\mathbb{Q})$ *be the point defined by (32). Then the height of* P_d *is given by*

$$h(P_d) = \frac{\sqrt{d}}{8\pi^2\|f\|^2} L_E'(1)L_{E,d}(1).$$

(To get this statement from [4], take $\chi = 1$ in Theorem 2 there, noting that $v_{f,1} = uP_d$ and $L(f,1,s) = L_E(s)L_{E,d}(s)$; the height in [4] is one-half that on E because it is calculated on $X_0(37)$ which is a double cover of E.)

In view of equation (31), Theorem 3 is equivalent to the formula (18) given in §2. As explained there, this formula gives both equation (10) for $L_E'(1)$ and the relationship (19) between $A(d)$ and the integers $b(d)$ defined by $P_d = b(d)P_0$. The equality $b(d)^2 = c(d)^2$ suggested comparing the values of $b(d)$ and $c(d)$. Note that the numbers $b(d)$ are numerically calculable: one finds the h Γ-inequivalent solutions of (32) by reduction theory, computes the corresponding values $\phi(\frac{b+i\sqrt{d}}{2a})$ by (26), adds the resulting complex numbers; modulo $\Lambda = \mathbb{Z}\omega_1+\mathbb{Z}\omega_2$, the result must be a multiple of the point $P_0 = -.92959\dots + \frac{1}{2}\omega_2$. Thus for $d = 67$ we have $h = 1$ and

$$P_{67} = \phi(\frac{9+i\sqrt{67}}{74}) = .40936\dots \equiv 6P_0 (\text{mod } \Lambda),$$

so $b(67) = 6$; for $d = 83$ we have $h = 3$,

$$P_{83} = \phi(\frac{55+i\sqrt{83}}{222}) + \phi(\frac{19+i\sqrt{83}}{74}) + \phi(\frac{55+i\sqrt{83}}{74}) = (.541\dots + 1.225\dots i)$$

$$+ (.194\dots - .570\dots i) + (.194\dots + .570\dots i) \equiv -P_0 (\text{mod } \Lambda),$$

so $b(83) = -1$; for $d = 148$ we have $h = 2$,

$$2P_{148} = \phi(\frac{i}{\sqrt{37}}) + \phi(\frac{1}{2} + \frac{i}{\sqrt{37}}) = .19189\ldots - .60125\ldots \equiv -6P_0 \pmod{\Lambda},$$

so $b(148) = -3$. In this way one can make a table of the multiples $b(d)$. Such a table (up to $d = 150$) was computed by B. Gross and J. Buhler, while I was independently computing the Fourier coefficients $c(d)$ by the method mentioned in §4; the letter with their data arrived in Germany on the very morning that I had completed my computations and drafted a letter to them, and the perfect agreement of the two tables gave ample reason to conjecture the following:

Theorem 4. $b(d) = c(d)$ *for all* d.

The remainder of this paper is devoted to the proof of this result.

6. Curves on Hilbert modular surfaces

In view of the uniqueness clause in Theorem 1, what we need to do to prove Theorem 4 is simply to show that $\sum b(d)q^d$ belongs to $S^+_{3/2}(37)$, i.e. that the positions of the Heegner points in the Mordell-Weil group of E are the Fourier coefficients of a modular form of weight 3/2. This statement is reminiscent of a theorem of Hirzebruch and the author [7] according to which the positions of certain modular curves in the homology group of a modular surface are the Fourier coefficients of a modular form of weight 2. Since this result is not only very analogous to the one we want, but will actually be used to prove it, we recall the exact statement.

Let p be a prime congruent to 1 (mod 4) and let $\mathcal{O} = \mathbb{Z} + \mathbb{Z}\frac{1+\sqrt{p}}{2}$ be the ring of integers in $\mathbb{Q}(\sqrt{p})$. The group $PSL_2(\mathcal{O})$ (*Hilbert modular group*) acts on $\mathcal{H} \times \mathcal{H}$ by

$$M\circ(\tau_1,\tau_2) = (\frac{a\tau_1+b}{c\tau_1+d}, \frac{a'\tau_2+b'}{c'\tau_2+d'}) \quad (M = \pm\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathcal{O}),\ \tau_1,\tau_2 \in \mathcal{H}),$$

where $'$ denotes conjugation in $\mathbb{Q}(\sqrt{p})/\mathbb{Q}$. The quotient $\mathcal{H} \times \mathcal{H}/SL_2(\mathcal{O})$ can be naturally compactified by the addition of finitely many points ("cusps"), and when the singularities thus introduced are resolved by cyclic configurations of rational curves according to Hirzebruch's recipe [6] the resulting surface $Y = Y_p$ is a nearly smooth compact algebraic surface (it still has quotient singularities coming from the points in $\mathcal{H} \times \mathcal{H}$ with a non-trivial isotropy group in $PSL_2(\mathcal{O})$, so it is a rational homology manifold or "V-manifold"). The middle homology of Y splits as

$$H_2(Y) = H_2^c(Y) \oplus \langle S_1 \rangle \oplus \ldots \oplus \langle S_r \rangle, \tag{34}$$

where $S_1, \ldots, S_r$ are (the homology classes of) the curves used in the resolutions of the cusp singularities and $H_2^c(Y)$ consists of homology classes orthogonal to the S_j; the homology groups in (34) are taken with coefficients in $\mathbb{Q}$.

For each integer $N > 0$ there is an algebraic curve $T_N \subset Y$ defined as follows. Consider all equations

$$A\tau_1\tau_2 + \frac{\lambda}{\sqrt{p}}\tau_1 - \frac{\lambda'}{\sqrt{p}}\tau_2 + B = 0 \tag{35}$$

with $A,B \in \mathbb{Z}$, $\lambda \in \mathbb{Q}$, and $\lambda\lambda' + ABp = N$. Each one defines a curve in $\mathcal{H} \times \mathcal{H}$ isomorphic to $\mathcal{H}$ and the union of these curves is invariant under $SL_2(\mathbb{Q})$; T_N is defined as closure in Y of the image of this union in $\mathcal{H} \times \mathcal{H}/SL_2(\mathbb{Q})$. If $(\frac{N}{p}) = -1$, there are no solutions of $\lambda\lambda' + ABp = N$ and T_N is empty. If $(\frac{N}{p}) = +1$ then T_N is irreducible (all equations (35) are equivalent under $PSL_2(\mathbb{Q})$) and isomorphic to the modular curve $X_0(N)$. The main result of [7] is

Theorem 5 (Hirzebruch-Zagier). Let $[T_N^c]$ denote the projection to $H_2^c(Y)$ of the homology class of T_N in the splitting (34). Then the power series $\sum_{N=1}^{\infty} [T_N^c] e^{2\pi i N\tau}$ is a modular form of weight 2, level p and Nebentypus $(\frac{\cdot}{p})$.

Here "modular form of weight 2, level p and Nebentypus" means a modular form $F(\tau)$ satisfying $F(\frac{a\tau+b}{c\tau+d}) = (\frac{a}{p})(c\tau+d)^2F(\tau)$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(p)$; when we say that a power series with coefficicents in $H_2^c(Y)$ is such a form we mean that each component (with respect to a basis of $H_2^c(Y)$ over $\mathbb{Q}$) is. Alternatively, if $[X]$ is any homology class in $H_2(Y)$, then the power series $\sum (X \circ T_N^c) e^{2\pi i N\tau}$, where $(X \circ T_N^c)$ denotes the intersection pairing of $[X]$ and $[T_N^c]$, is a modular form of the specified type, now with ordinary numerical Fourier coefficients. In particular, this is true for $X = T_M$, one of our special curves on Y. In fact the proof of Theorem 5 in [7] consisted in calculating the intersection numbers $(T_M \circ T_N^c)$ explicitly and showing that they were the Fourier coefficients of a modular form. The formula obtained for $(T_M \circ T_N^c)$, in the case when N and M are coprime, was

$$(T_M \circ T_N^c) = \sum_{\substack{x^2 < 4NM \\ x^2 \equiv 4NM \pmod{p}}} H\left(\frac{4NM - x^2}{p}\right) + I_p(MN) \tag{36}$$

where

$$H(d) = \sum_{e^2 | d} h'(-d/e^2)$$

($h'(-d) = h(-d)$ for $d > 4$, $h'(-3) = 1/3$, $h'(-4) = 1/2$) and $I_p(n)$ is a certain arithmetical function whose definition we do not repeat. The proof of (36) was geometrical: the physical intersection points of T_M and T_N in $\mathcal{H} \times \mathcal{H}/PSL_2(\mathcal{O})$ are in 1:1 correspondence with certain equivalence classes of binary quadratic forms and are counted by the first term in (36), while the term $I_p(MN)$ counts the intersection points of T_M and T_N at infinity and the intersection of T_M with the combination of cusp-resolution curves S_j which was removed from T_N to get T_N^c.

7. Heegner points as intersection points of modular curves on modular surfaces

Now suppose that p is a prime satisfying $p \equiv 1 \pmod 4$, $(\frac{p}{37}) = 1$, and (for later purposes) $p > 2 \cdot 37$, say $p = 101$. As already mentioned, the curve T_{37} on Y_p is in this case isomorphic to $X_0(37) = \mathcal{H} \cup \{\text{cusps}\}/\Gamma_0(37)$. For instance, if $p = 101$ we can get an equation (35) for T_{37} by taking $A = B = 0$ and $\lambda = 21+2\sqrt{101}$, an element of $\mathcal{O}$ of norm 37; then the solution of (35) is given parametrically by $\{(\lambda\tau, \lambda'\tau),\ \tau \in \mathcal{H}\}$ and the matrices $M \in SL_2(\mathcal{O})$ which preserve this set are those of the form $\begin{pmatrix} a & b\lambda \\ c/\lambda & d \end{pmatrix}$ with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(37)$, so we get a degree 1 map $\mathcal{H}/\Gamma_0(37) \to \mathcal{H} \times \mathcal{H}/SL_2(\mathcal{O})$ and hence a map $X_0(37) \to Y_{101}$. On Y_p we have an extra involution ι which is induced by the involution $(\tau_1, \tau_2) \mapsto (\tau_2, \tau_1)$ of $\mathcal{H} \times \mathcal{H}$, and this induces the involution w_{37} on $T_{37} = X_0(37)$, so our curve $E \simeq \mathcal{H} \cup \{\text{cusps}\}/\Gamma$ can be found on the quotient surface Y/ι. However, since all T_N are invariant under ι and there is no difference (except a factor of 2) between the intersection theory of ι-invariant curves on Y or of their images in Y/ι, we will continue to work on the surface Y rather than the quotient surface Y/ι, which has a one-dimensional singular locus.

In §5 we constructed for each $d > 0$ a set of $(1+(\frac{-d}{37}))H(d)$ points in $X_0(37)$, namely the set of roots of quadratic equations $a\tau^2+b\tau+c=0$ with $b^2-4ac=-d$ and $37 | a$. (If d is of the form $3n^2$ or $4n^2$ then $H(d)$ is not an integer and we are using the convention that a fixed point of an element of order u in $\Gamma_0(37)$ is to be counted with multiplicity $\frac{1}{u}$ in $\mathcal{H}/\Gamma_0(37)$; from now on we will ignore this technicality.) Call this set P_d. The point $P_d \in E(\mathcal{O})$ was (one-half of) the sum of the images of the points of P_d in E. If we

worked on $X_0(37)$, or on some other $X_0(M)$ of higher genus, we would have to take the sum in the Jacobian of the curve rather than on the curve itself, i.e. P_d would be the point of $\mathrm{Jac}(X_0(M))$ represented by the divisor $P_d - \deg(P_d)\cdot(\infty)$ of degree 0.

The geometric content of (36) is that the intersection points of T_N and T_M in $\mathcal{H} \times \mathcal{H}/\mathrm{PSL}(\mathbb{Z})$ are the points of P_d for certain d, namely those of the form $\frac{148N-x^2}{p}$, i.e.

$$T_{37} \cap T_N = \bigcup_{\substack{|x|<\sqrt{148N} \\ x^2 \equiv 148N \pmod p}} P_{(148N-x^2)/p} \cup D_\infty \tag{37}$$

where D_∞ is contained in the part of Y_p at infinity (resolutions of the cusp singularities); here when we write union we of course mean for the points to be counted with appropriate multiplicities, i.e. we are working with divisors rather than just sets of points. If we simply count the points in (37), i.e. replace each P_d by its degree, we obtain the numbers (36), and Theorem 5 tells us that these are the Fourier coefficients of a modular form of weight 2, level p, and Nebentypus $(\frac{-}{p})$. If instead we add the points in (37) in the Jacobian of T_{37}, i.e. replace each P_d by P_d, then we will deduce from this that the corresponding statement holds:

Proposition: For $N > 0$ define $B(N)$ by

$$B(N) = \sum_{\substack{x^2<148N \\ x^2 \equiv 148N \pmod p}} b\left(\frac{148N-x^2}{p}\right)$$

with $b(d)$ as in §5. Then $\sum B(N)q^N$ is a modular form of weight 2, level p and Nebentypus $(\frac{-}{p})$.

Proof. Let M denote the set of all modular forms of the specified type, so that Theorem 5 asserts

$$\sum (T_N^C \circ X) q^N \in M \quad \text{for all} \quad [X] \in H_2(Y). \tag{38}$$

The space M is finite-dimensional and has a basis consiting of modular forms with rational Fourier coefficients. Hence there is an infinite set R of finite relations over $\mathbb{Z}$ defining M, i.e. a set R whose elements are sequences

$$R = (r_0, r_1, r_2, \ldots), \quad r_N \in \mathbb{Z}, \quad r_N = 0 \quad \text{for all but finitely many } N$$

and such that

$$\sum_{N=0}^{\infty} C(N)q^N \in M \Leftrightarrow \sum_{N=0}^{\infty} r_N C(N) = 0 \qquad (\forall R \in \mathcal{R}). \tag{39}$$

(For instance, one could find integers $N_1,\dots,N_d$ with $d = \dim M$ and such that the N_j^{th} Fourier coefficients of forms in M are linearly independent; then for each N we have a relation $C(N) = \sum_{j=1}^{d} \lambda_j C(N_j)$ with rational numbers $\lambda_1,\dots,\lambda_d$, and we could take for $\mathcal{R}$ the set of these relations, each multiplied by a common denominator.) Equation (38) now implies that

$$\sum_{N=1}^{\infty} r_N (T_N^c \circ X) = 0$$

for all $R \in \mathcal{R}$, and since this holds for all homology classes X, we must have $\sum r_N [T_N^c] = 0$ in $H_2(Y,\mathbb{Q})$. Since T_N^c is a linear combination of T_n and curves S_j coming from the cusp resolutions, this means

$$\sum_{N=1}^{\infty} r_N [T_N] + \sum_{j=1}^{r} s_j [S_j] = 0 \tag{40}$$

in $H_2(Y,\mathbb{Q})$ for some rational numbers $s_1,\dots,s_r$. Multiplying by a further common denominator we can assume that the s_j are also integers and that the relation (40) holds in integral homology. But the Hilbert modular surface Y is known to be simply connected, so the exact sequence

$$0 = H^1(Y,\mathcal{O}) \to H^1(Y,\mathcal{O}^*) \to H^2(Y,\mathbb{Z}) \quad (\mathcal{O} = \text{structure sheaf of } Y)$$

induced from $0 \to \mathbb{Z} \to \mathcal{O} \to \mathcal{O}^* \to 0$ shows that any divisor on Y which is homologous to 0 is linearly equivalent to 0. Hence the relation (39) implies that the divisor $\sum r_N T_N + \sum s_j S_j$ is the divisor of a meromorphic function on Y, i.e. there is a meromorphic function Φ on Y which has a zero or pole of order r_N on each T_N (resp. s_j on each S_j) and no other zeros or poles. If we restrict Φ to T_{37}, then it follows that the zeros and poles of Φ occur at the intersection points of T_{37} with other T_N and at the cusps, and in fact (by (38)) that

$$\text{divisor of } \Phi\Big|_{T_{37}} = \sum_{N \geq 1} r_N \sum_{\substack{x^2 < 148N \\ x^2 \equiv 148N \pmod{p}}} P_{(148N-x^2/p)} + d_\infty$$

where d_∞ is a divisor with support concentrated at the cusps. Take the image in E, observing that the cusps map to 0, and add the points obtained; since the points of a principal divisor sum to zero and the points of P_d sum to $b(d)P_0$, we deduce $\sum r_N B(N) = 0$ with $B(N)$ as in the Proposition. The desired result now follows from equation (39).

8. Completion of the proof

We are now nearly done. For each $N > 0$ define

$$C(N) = \sum_{\substack{x^2<148N \\ x^2\equiv 148N \pmod p}} c\left(\frac{148N-x^2}{p}\right),$$

where $c(d)$ are the Fourier coefficients defined in §4. Then

$$G(z) := \sum_{N>0} C(N)q^N = \sum_{\substack{d>0 \\ x\in\mathbf{Z} \\ pd+x^2\equiv 0 \pmod{148}}} c(d)q^{(pd+x^2)/148}$$

$$= g(pz)\theta(z)\,|\,U_{148},$$

where $\theta = \sum q^{x^2}$ and U_m is the map which picks out every m^{th} coefficient of a Fourier expansion, i.e.

$$\phi(z)\,|\,U_m = \frac{1}{m}\sum_{j \pmod m} \phi\left(\frac{z+j}{m}\right).$$

Since g is a modular form of weight $3/2$ and θ one of weight $1/2$, and since U_m maps modular forms to modular forms of the same weight, it is clear that $G(z)$ is a modular form of weight 2; a routine calculation shows that it has level p and Nebentypus $(\frac{\cdot}{p})$. Hence both $G(z)$ and $F(z) = \sum B(N)q^N$ belong to the finite-dimensional space M. Moreover, since $b(d) = c(d)$ for small d by the calculations mentioned in §5, the first Fourier coefficients of F and G agree, and this suffices to show $F = G$. Specifically, with $p = 101$ the agreement of $c(d)$ and $b(d)$ for $d < 150$ implies the agreement of $B(N)$ and $C(N)$ for $1 \le N \le 100$, and this is more than enough to ensure that $F = G$ (it would suffice to have agreement up to $N = 9$). Hence $B(N) = C(N)$ for all N. We claim that this implies $b(d) = c(d)$ for all d. Indeed, suppose inductively that $b(d') = c(d')$ for all $d' < d$. If $(\frac{-d}{37}) = -1$ or $-d \equiv 2$ or $3 \pmod 4$ then $c(d)$ and $b(d)$ are both zero and there is nothing to prove. Otherwise we can find an

integer n with

$$n^2 \equiv -pd \pmod{148}, \quad |n| \le 37.$$

Take $N = \frac{pd+n^2}{148}$. Then in the equations

$$B(N) = \sum_{\substack{x^2<148N \\ x^2\equiv 148N \pmod p}} b\left(\frac{148N-x^2}{p}\right), \quad C(N) = \sum_{\substack{x^2<148N \\ x^2\equiv 148N \pmod p}} c\left(\frac{148N-x^2}{p}\right)$$

the numbers $\pm n$ occur as values of x and all other values of x are larger in absolute value because $|n| \le 37 < \frac{1}{2}p$ by assumption. Thus $B(N)$ equals 1 or 2 times $b(d)$ plus a certain linear combination of $b(d')$ with $d' < d$, and $C(N)$ equals the same multiple of $c(d)$ plus the same linear combination of lower $c(d')$, so the equality $B(N) = C(N)$ and the inductive assumption $b(d') = c(d')$ imply that $b(d) = c(d)$ as desired.

9. Generalization to other modular curves

Our exposition so far was simplified by several special properties of the elliptic curve E: that it was actually isomorphic to a modular curve rather than just covered by one, that its Mordell-Weil group had rank one and no torsion, etc. We end the paper by discussing to what extent the results proved for E generalize to other curves.

First, we could replace E by an arbitrary elliptic curve whose L-series coincides with the L-series of a modular form f of weight 2 and some (say, prime) level N, with $f(-\frac{1}{N\tau}) = N\tau^2 f(\tau)$. Then we would again have a covering map $\phi: X_0(N)/w_N \to E$, Heegner points $P_d \in E(\mathbb{Q})$ for all $d > 0$ (with $P_d = 0$ if $-d \not\equiv 0 \pmod{4N}$), and a relationship $c(d)^2 \sim h(P_d)$ for the Fourier coefficients $c(d)$ of a modular form in $S^+_{3/2}(N)$ corresponding to f as in Theorem 1. We could then ask whether all the P_d belong to a one-dimensional subspace $\langle P_0 \rangle$ of $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}}$ and, if so, whether the coefficients $b(d)$ defined by $P_d = b(d)P_0 \in E(\mathbb{Q}) \otimes \mathbb{Q}$ are proportional to the Fourier coefficients $c(d)$. More generally, we could forget elliptic curves entirely and simply start with a modular curve $X_0(N)$ or $X_0(N)/w_N$ (still, say, with N prime). The construction of §5 yields Heegner points P_d in the Jacobian of this curve over $\mathbb{Q}$. To avoid torsion we tensor with $\mathbb{Q}$ and write $V = \text{Jac}(X_0(N)/w_N)(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{Q}$. The Hecke algebra acts on V the same way as it acts on cusp forms of weight 2, so $V \otimes \mathbb{R}$ splits as $\oplus_f V_f$, where the f are Hecke eigen-

forms $f = \sum_1^\infty a(n)q^n$ in $M_2^+(\Gamma_0(N))$ (normalized by $a(1) = 1$) and V_f is the subspace of $V \otimes \mathbb{R}$ on which the n^{th} Hecke operator acts as multiplication by $a(n)$. For each f we define $P_{d,f}$ as the component of P_d in V_f. The Fourier coefficients $a(n)$ will be in $\mathbb{Z}$ if f corresponds to an elliptic curve E defined over $\mathbb{Q}$; in that case V_f is isomorphic to $E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{Q}$ and we are back in the situation described before. In general the $a(n)$ will be integers in an algebraic number field $K_f \subset \mathbb{R}$, the Fourier coefficients $c(d)$ of the form in $S_{3/2}^+(N)$ corresponding to f can also be chosen to lie in K_f, and the main theorem of [4] combined with Theorem 2 tells us that $h(P_{d,f})$ is proportional to $c(d)^2$. This suggests that the right generalization of Theorem 4 is:

Theorem 6. *Let f, $c(d)$ be as above. Then $P_{d,f} = c(d)P_0$ for all d and some $P_0 \in V_f$. In particular, the projections $P_{d,f}$ of the Heegner points all lie in a one-dimensional subspace of V_f.*

Theorem 6 is equivalent (because of the uniqueness clause in Theorem 1 and the way the Hecke operators act on Heegner points) to the following apparently weaker theorem:

Theorem 6'. *The power series $\sum_{d>0} P_d q^d$ is a modular form of weight 3/2 and level N.*

(As with Theorem 5, this means that $\sum P_d q^d \in V[[q]]$ belongs to the subspace $V \otimes S_{3/2}^+(N)$ or, in more down-to-earth terms, that each component of this power series, with respect to a fixed basis of V over $\mathbb{Q}$, is a modular form in $S_{3/2}^+(N)$.)

How can we prove these theorems? The argument of §§6-7 permits us to embed our modular curve in the Hilbert modular surface Y_p for any prime $p \equiv 1 \pmod 4$ with $(\frac{N}{p}) = 1$ and to prove that the power series

$$\sum_M \left(\sum_{\substack{x^2 < 4NM \\ x^2 \equiv 4NM \pmod p}} P_{(4NM-x^2)/p} \right) q^M$$

is a modular form (with coefficients in V) of weight 2, level p and Nebentypus. To deduce Theorem 6' we would need the following assertion:

Let $h(\tau)$ be a power series of the form

$$\sum_{\substack{d>0 \\ -d \equiv \text{square} \pmod{4N}}} b(d)q^d$$

with N prime, and suppose that the power series

$$h(p\tau)\theta(\tau)\,|U_N = \sum_{M>0}\left(\sum_{\substack{x^2<4NM \\ x^2\equiv 4NM \pmod{p}}} b\left(\frac{4NM-x^2}{p}\right)\right)q^M$$

is a modular form of weight 2, level p and Nebentypus $\left(\frac{\cdot}{p}\right)$ for every prime $p \equiv 1 \pmod 4$ with $\left(\frac{N}{p}\right) = 1$. Then h belongs to $S^+_{3/2}(N)$.

This assertion is extremely likely to be true. The argument of §8 proves it — even if the hypothesis on $h(p\tau)\theta(\tau)\,|U_N$ is made for only one prime $p > 2N$ — under the additional assumption that one possesses a candidate $g = \sum c(d)q^d \in S^+_{3/2}(N)$ for h with $c(d) = b(d)$ for sufficiently many values of d. Thus the method of proof we used for $N = 37$ can be used for any other fixed value of N if we do a finite amount of computation. To get a general proof of Theorems 6 and 6' along these lines one would need either to prove the assertion above or else to generalize the geometric proof in some way (perhaps by using Hilbert modular surfaces of arbitrary discriminant, for which the intersection theory has been worked out by Hausmann [5]).

In any case, however, we would like to have a proof of Theorem 6 using only intrinsic properties of the modular curve, rather than its geometry as an embedded submanifold of an auxiliary modular surface. Such a proof has been given by B. Gross, W. Kohnen and myself. It is a direct generalization of the main result of [4]: instead of a formula for the height $h(P_d)$ of a Heegner point, we give a formula for the height pairing $(P_d, P_{d'})$ of two Heegner points, where $(\ ,\ ): V \times V \to \mathbb{R}$ is the bilinear form associated to the quadratic form h. The formula implies that $\sum_{d>0} (P_d, P_{d'})q^d$ belongs to $S^+_{3/2}(N)$ for each discriminant d', and Theorem 6' follows.

Finally, we mention that the correct generalization of Theorem 4 to composite levels N should be formulated using the theory of "Jacobi forms" developed in [3] rather than the theory of modular forms of half-integral weight. This, too, will be carried out in the joint work with Kohnen and Gross mentioned above.

Bibliography

[1] B. Birch, Heegner points of elliptic curves, Symp. Mat., Ist. di Alta Mat. 15(1975), 441-445.

[2] J. Buhler, B. Gross and D. Zagier, On the conjecture of Birch and Swinnerton-Dyer for an elliptic curve of rank 3, to appear in Math. Comp. (1985).

[3] M. Eichler and D. Zagier, The Theory of Jacobi Forms, to appear in Progress in Mathematics, Birkhäuser, Boston-Basel-Stuttgart (1985).

[4] B. Gross and D. Zagier, Points de Heegner et dérivées de fonctions L, C.R. Acad. Sci. Paris 297(1983), 85-87.

[5] W. Hausmann, Kurven auf Hilbertschen Modulflächen, Bonner Math. Schriften 123(1980).

[6] F. Hirzebruch, Hilbert modular surfaces, L'Ens. Math. 19(1973), 183-281.

[7] F. Hirzebruch and D. Zagier, Intersection numbers of curves on Hilbert modular surfaces and modular forms of Nebentypus, Inv. Math. 36(1976), 57-113.

[8] W. Kohnen, Modular forms of half-integral weight on $\Gamma_0(4)$, Math Ann. 248(1980), 249-266.

[9] W. Kohnen, New forms of half-integral weight, J. reine Angew. Math. 333(1982), 32-72.

[10] W. Kohnen, Fourier coefficients of modular forms of half-integral weight, to appear in Math. Ann. (1985).

[11] B. Mazur and H.P.F. Swinnerton-Dyer, Arithmetic of Weil curves, Inv. Math. 25(1974), 1-61.

[12] G. Shimura, On modular forms of half-integral weight, Ann. of Math. 97(1973), 440-481.

[13] H.P.F. Swinnerton-Dyer and B. Birch, Elliptic curves and modular functions, in Modular Functions of One Variable IV, Lecture Notes in Math. 476, Springer, Berlin-Heidelberg-New York (1975), 2-32.

[14] J. Tate, Letter to J-P. Serre, Oct. 1, 1979.

[15] J.-L. Waldspurger, Sur les coefficients de Fourier des formes modulaires de poids demi-entier, J. Math. pures et appl. 60 (1981), 375-484.