FERMAT'S LAST THEOREM AFTER 356 YEARS

by William F. Hammond

A Lecture at the Everyone Seminar University at Albany, October 22, 1993

1 Introduction.

The purpose of this expository lecture is to explain the basic ideas underlying the final resolution of "Fermat's Last Theorem" after 356 years as a consequence of the reported establishment by Andrew Wiles of a sufficient portion of the "Shimura-Taniyama-Weil" conjecture. As these notes are being written, the work of Wiles is not available, and the sources of information available to the author are (1) reports by electronic mail, (2) the *AMS Notices* article [16] of K. Ribet, and (3) a preprint [18] by K. Rubin and A. Silverberg based on the June, 1993 lectures of Wiles at the Newton Institute in Cambridge, England. It should be noted that the fact that "Fermat's Last Theorem" is a consequence of sufficient knowledge of the theory of "elliptic curves" has been fully documented in the publications ([14], [15]) of K. Ribet.

"Fermat's Last Theorem" is the statement, having origin with Pierre de Fermat in 1637, that there are no positive integers x, y, z such that $x^n + y^n = z^n$ for any integer exponent $n > 2$. Obviously, if there are no positive integer solutions x, y, z for a particular n, then there are certainly none for exponents that are multiples of n. Since every integer $n > 2$ is divisible either by 4 or by some odd prime p, it follows that "Fermat's Last Theorem" is true if there are no solutions in positive integers of the equation $x^n + y^n = z^n$ when $n = 4$ and when $n = p$ for each prime $p > 2$. The cases $n = 3,\ 4$ are standard fare for textbooks (e.g., see Hardy & Wright [6]) in elementary number theory. Therefore, this discussion will focus on the case $n = p$ where $p > 3$ is prime.

Very briefly, the idea is that we now know enough about the classification of non-degenerate plane cubic curves $F(x, y) = 0$ in two variables, also known as "elliptic curves", with *rational* coefficients to know how to enumerate them in a logical way so that we may conclude that if there were positive integers a, b, c with $a^p + b^p = c^p$, then the curve

$$y^2 = x(x - a^p)(x + b^p),$$

which is an elliptic curve known as the "Frey curve", would fall inside of the enumeration. Because the classification is enumerative, when one is presented with a particular elliptic curve with rational coefficients, one knows where to look for the curve in the classification. The curve just written is not to be found within the classification. As a consequence there cannot be positive integers a, b, c with $a^p + b^p = c^p$.

The enumerative classification of non-degenerate plane cubic curves defined by polynomials with rational coefficients has been entirely conjectural (variously known as the "Taniyama Conjecture", the "Weil Conjecture", the "Taniyama-Shimura Conjecture", ...) until June, 1993. This conjecture, even as a conjecture, has served as an important motivating example for the idea of the "Langlands Program", or perhaps of an extension of that program, that certain kinds of objects in geometry should give rise to certain group representations.

What seems to be believed today[1] is that the portion of the enumerative classification pertaining to "semi-stable" elliptic curves has been proved by Andrew Wiles. That the existence of positive integers a, b, c with $a^p + b^p = c^p$ would violate the enumerative classification of semi-stable elliptic curves was established by 1987 through the work of G. Frey, J.-P. Serre, and K. Ribet.

The primary purpose of this lecture is to explain the enumerative classification of elliptic curves and to give a brief indication of the mathematics involved in showing that the Frey curve violates that classification.

[1]As of the time of this write-up Wiles has stated that a portion of what he announced in June needs further justification and that he expects to be able to complete it. See the appendix.

2 Elliptic curves

A polynomial $f(X,Y)$ of degree d in two variables with coefficients in a field k gives rise to what is called an *affine plane curve* of degree d: for each field K containing k (more generally, for each commutative ring that is a k-algebra) one has the set

$$C_0(K) = \{(x,y) \in K^2 | f(x,y) = 0\},$$

and for each k-linear homomorphism $K \longrightarrow K'$ one has the induced map $C_0(K) \to C_0(K')$. From the polynomial f one obtains a homogeneous polynomial of degree d in three variables with coefficients in k:

$$F(X,Y,Z) = Z^d f(X/Z, Y/Z),$$

and the *projective plane curve* of degree d:

$$C(K) = \{((x,y,z)) \in \mathbf{P}^2(K) | F(x,y,z) = 0\},$$

where $\mathbf{P}^N(K)$ denotes N-dimensional projective space, which is the quotient set of $K^{N+1} - \{0\}$ obtained by identifying points lying on the same line through the origin of K^{N+1}. Since the projective plane $\mathbf{P}^2(K)$ is the disjoint union of the affine plane $\{((x,y,1)) | (x,y) \in K^2\}$ with the "*(projective) line at infinity*" $\{((x,y,0)) | ((x,y)) \in \mathbf{P}^1(K)\}$, it follows that $C(K)$ is the disjoint union of $C_0(K)$ with the finite set of its points lying on the projective line at infinity.

An *elliptic curve defined over* k is the (projective) plane curve E given by a homogeneous polynomial F of degree 3 in three variables with coefficients in k such that (i) F is irreducible over the algebraic closure $\overline{k}$ of k, (ii) the gradient vector ∇F is a non-vanishing vector at points of $\overline{k}^3 - \{0\}$ where F vanishes, and (iii) the set $E(k)$ is non-empty.

If k is any field, then after an isomorphism (see Silverman [27]) one may obtain a given elliptic curve E with an affine equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \ . \tag{1}$$

Then the homogeneous equation for the intersection of $E(K)$ with the line at infinity is

$$x^3 = 0 \ . \tag{2}$$

Thus, in this case, E has a unique point on the line at infinity. If the characteristic of k is different from 2 and 3[†] then one may obtain an equation in "Weierstrass normal form":

$$y^2 = 4x^3 - g_2 x - g_3 \ , \tag{3}$$

which is non-singular if and only if the cubic polynomial in the variable x has distinct roots in $\overline{k}$.

Elliptic curves are the "group objects" in the category of algebraic curves that reside in projective space: for each extension field K of k the set $E(K)$ of "K-valued points" of E is an abelian group. The group law on $E(K)$ is characterized by two conditions:

1. The origin is a given point of $E(k)$.

2. The points obtained by intersecting $E(K)$ with any line in $\mathbf{P}^2(K)$, counted with multiplicities, add up to zero.

When E is given by an equation in the form (1), the origin is usually taken to be the unique point on the line at infinity. If two distinct points of $E(K)$ are given, they determine a line in $\mathbf{P}^2(K)$; the intersection of that line with $E(K)$ is given by a cubic polynomial in a parameter for the line which has two roots in K corresponding to the two given points; hence, there is a third root of that cubic polynomial in K; this root gives rise to a point of $E(K)$, which is the negative of the sum of the two given points. The negative of a given point of $E(K)$ is obtained as the third point in the intersection with $E(K)$ of the line through the given point and the origin.

For a given field k the set of homogeneous cubic polynomials in three variables is a vector space over k having the set of "monomials" of degree three in three variables as basis. Thus, the dimension of the space of homogeneous cubics is 10. The linear group $\mathrm{GL}_3(k)$ acts on the space of cubics, and two cubic curves in $\mathbf{P}^2$ that are related by this action are isomorphic. Since $\mathrm{GL}_3(k)$ is 9-dimensional, one is led to think of the *family* of isomorphism classes of elliptic curves as 1-dimensional since "non-singularity" is an "open" condition.

[†]Thus, one sees that the primes 2 and 3 play a special role in the theory of elliptic curves.

3 Elliptic curves over C

When k is the field $\mathbf{C}$ of complex numbers, one knows (see, e.g., Ahlfors [1]) that for each lattice Λ in $\mathbf{C}$ the set of Λ-periodic meromorphic functions on the complex line $\mathbf{C}$ is the field $\mathbf{C}(\wp, \wp')$, which is a quadratic extension of the rational function field $\mathbf{C}(\wp)$, where $\wp$ is the $\wp$-function of Weierstrass. Moreover, $\wp$ satisfies the famous Weierstrass differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda); \tag{4}$$

thus, the formula $z \mapsto (\wp(z), \wp'(z))$ defines a holomorphic map from the punctured complex torus $\mathbf{C}/\Lambda - \{0\}$ to the affine cubic curve

$$y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda); \tag{5}$$

it should hardly be necessary to point out that this map extends to a holomorphic map from the torus $\mathbf{C}/\Lambda$ to the corresponding (projective) elliptic curve by sending the origin of the torus to the unique point of the elliptic curve on the line at infinity. The classical theory of theta functions (see, e.g., Igusa [7] or Siegel [26]) leads to a direct demonstration that this map is a homomorphism from the group law on the complex torus to the group law previously described for an elliptic curve. It is not difficult to see that the analytic manifold given by any elliptic curve defined over $\mathbf{C}$ arises from some complex torus. Indeed each non-singular cubic curve E in $\mathbf{P}^2(\mathbf{C})$ determines a compact connected complex-analytic group. Its universal cover is given by a holomorphic homomorphism $\mathbf{C} \to E$ which has some lattice as kernel.

Any two lattices in $\mathbf{C}$ are related by a change of real basis for $\mathbf{C}$, i.e., by a matrix in $\mathrm{GL}_2(\mathbf{R})$. Consequently, there is only one real-analytic isomorphism class for the complex torus $\mathbf{C}/\Lambda$ as Λ varies. The tori corresponding to two lattices are complex-analytically isomorphic if and only if the corresponding real-linear isomorphism of $\mathbf{R}^2$ satisfies the Cauchy-Riemann partial differential equations, i.e., if and only if the $\mathbf{R}$-linear isomorphism is $\mathbf{C}$-linear.

A lattice Λ may be represented concretely by an ordered basis $\{\omega_1, \omega_2\}$. If $\tau = \omega_2/\omega_1$, then τ is not real, and after permuting the basis members, if necessary, one may assume that τ is in the "upper-half plane"[3] H of $\mathbf{C}$. Observing that Λ is the image under the $\mathbf{C}$-linear map $z \mapsto \omega_1 z$ of the lattice

[3]The fact that the half-plane is a model of non-Euclidean geometry led a popular columnist in November, 1993 to question the validity of the work being discussed here.

with ordered basis $\{1, \tau\}$, one may assume that Λ is this latter lattice. Let $E(\tau)$ be the complex torus $\mathbf{C}/\Lambda$. Allowing for change of basis subject to these assumptions on the basis, one sees that there is an isomorphism of complex-analytic groups $E(\tau') \cong E(\tau)$ if

$$\tau' = \frac{a\tau + b}{c\tau + d}, \tag{6}$$

for some matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}).$$

Conversely, the monodromy principle may be used to show that every complex-analytic isomorphism among the complex tori $E(\tau)$ arises in this way.

The coefficients g_2 and g_3 in the Weierstrass normal form (5) have very explicit constructions as infinite series (see, e.g., Ahlfors [1] or Serre [20]) determined by the given lattice; from this it is straightforward to see that g_w is a *modular form of weight* $2w$: if τ and τ' are related by (6), then

$$g_w(\tau') = (\lambda)^{2w} g_w(\tau), \quad \lambda = c\tau + d \ .$$

Consequently, the map

$$(x, y) \mapsto (\lambda^2 x, \lambda^3 y)$$

carries the curve given by (5) for τ isomorphically to the curve given by (5) for τ'. The discriminant of the cubic polynomial in the Weierstrass normal form (5) is a modular form of weight 12, which up to a multiplicative constant, is:

$$\Delta(\tau) = g_2^3 - 27 g_3^2 \ .$$

Δ is a non-vanishing holomorphic function in H. The *modular invariant* j ([20],[24]) is defined by:

$$j(\tau) = \frac{(12 g_2)^3}{\Delta} \ ;$$

it is a holomorphic function in the upper-half plane H with the property that

$$j(\tau) = j(\tau')$$

if and only if τ and τ' are related by (6). Furthermore, j assumes every value in $\mathbf{C}$ at some point of H. Consequently, the complex-analytic isomorphism classes of complex tori or, equivalently, the isomorphism classes of elliptic curves defined over $\mathbf{C}$, are parameterized via j in a one-to-one manner by the complex numbers.

Since this is an expository discourse, it is hoped that the reader will not feel patronized by having noted the fact that the coincidence of (1) the category of elliptic curves over $\mathbf{C}$ and (2) the category of complex tori is the "genus one" case of the coincidence (see Weyl [33]) of (i) the category of "complete" non-singular algebraic curves over $\mathbf{C}$ and (ii) the category of compact Riemann surfaces (one-dimensional connected complex-analytic manifolds).

Although the classification of elliptic curves over $\mathbf{C}$ via the j-function is a result that is both beautiful and useful, and although two elliptic curves defined over $\mathbf{Q}$ that are isomorphic as curves defined over $\mathbf{Q}$ give rise to elliptic curves defined over $\mathbf{C}$ that have the same j-invariant, it is *not* true that any two elliptic curves defined over $\mathbf{Q}$ having the same j-invariant are isomorphic over $\mathbf{Q}$. Thus, the classification of elliptic curves over $\mathbf{C}$ does not lead directly to the desired enumerative classification of elliptic curves defined over $\mathbf{Q}$, but it does bring to the fore the notion of *modular form*, which is central in the study of elliptic curves defined over $\mathbf{Q}$. What can be said easily is that, according to the Shimura-Taniyama-Weil conjecture, the *isogeny* classes of elliptic curves defined over $\mathbf{Q}$ are parameterized by certain modular forms.

4 Modular forms

The group $\mathrm{SL}_2(\mathbf{Z})$ is an infinite group that is generated by the two elements

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

which have orders 4 and 6 respectively. The action of $\mathrm{SL}_2(\mathbf{Z})$ on the upper-half plane H by linear fractional transformations has kernel

$$\left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

and the quotient of $\mathrm{SL}_2(\mathbf{Z})$ by this kernel is the group $\mathrm{PSL}_2(\mathbf{Z})$. It is not difficult to see that the set

$$\{\tau \in H \mid -1/2 \leq Re(\tau) \leq 1/2, \quad |\tau| \geq 1\}$$

is a "fundamental domain" for the action of $\mathrm{PSL}_2(\mathbf{Z})$ on H. More precisely, this set meets each orbit, and the only redundancies are the boundary identifications arising from the maps $\tau \mapsto \tau + 1$ and $\tau \mapsto -1/\tau$. The quotient $H/\mathrm{PSL}_2(\mathbf{Z})$ is not compact since the fundamental domain is "open at the top". Beyond that the modular invariant j induces a bicontinuous biholomorphic isomorphism of the quotient $H/\mathrm{PSL}_2(\mathbf{Z})$ with the affine line over $\mathbf{C}$. Since $j(\tau + 1) = j(\tau)$, and since for $q = e^{2\pi i \tau}$ one has $|q| < 1$ for $\tau \in H$, there is a holomorphic function $\tilde{j}$ in the punctured unit disk such that $\tilde{j}(q) = j(\tau)$. Likewise Δ may be regarded as function of q, and one may use the calculus of residues to show that Δ has a simple zero at $q = 0$; hence, $\tilde{j}$ has a simple pole at $q = 0$, or, equivalently, j has a simple pole at ∞ (the "missing top" of the fundamental domain). Thus, j gives rise to a bicontinuous biholomorphic isomorphism

$$H/\mathrm{PSL}_2(\mathbf{Z}) \cup \{\infty\} \longrightarrow \mathbf{P}^1(\mathbf{C}) \ .$$

A non-trivial element of $\mathrm{PSL}_2(\mathbf{Z})$ has a fixed point in H if and only if it has finite order, and one's explicit knowledge of the fundamental domain makes it possible to see that the only elements of finite order are of order 2 or 3[§]. A *congruence subgroup* of $\mathrm{SL}_2(\mathbf{Z})$ is a subgroup Γ that contains one of the principal congruence subgroups; the *principal congruence subgroup* $\Gamma(N)$ of *level* N is the set of all elements γ of $\mathrm{SL}_2(\mathbf{Z})$ that are congruent mod$N$ to the identity matrix. The group $\Gamma_0(N)$ is the congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$ consisting of all elements

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

for which $c \equiv 0 \bmod N$. It is obvious that each congruence group Γ has finite index in $\mathrm{SL}_2(\mathbf{Z})$, and, consequently the quotient H/Γ is a non-compact Riemann surface. Observe that for each level N the group $\Gamma_0(N)$ contains the parabolic element

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} ,$$

[§]Thus, one sees that the primes 2 and 3 play a special role in the study of the group $\mathrm{SL}_2(\mathbf{Z})$.

which gives rise to the holomorphic map $\tau \mapsto \tau + 1$ that fixes the point ∞.

A modular form[5] of *weight* w for Γ is a holomorphic function f in H that satisfies the functional equation

$$f(\gamma \cdot \tau) = (c\tau + d)^w f(\tau), \ \gamma \in \Gamma \tag{7}$$

and that is holomorphic at each *cusp* of Γ. The role of *cusps* for Γ is to provide a slightly larger set H^* than H,

$$H^* = H \cup \{cusps\} \ ,$$

where Γ acts such that H^*/Γ is a compact Riemann surface containing H/Γ as the open complement of a finite set of points arising from cusps. The cusps of Γ are the points of the closure of the boundary of H in $\mathbf{P}^1(\mathbf{C}) = \mathbf{C} \cup \{\infty\}$ that are fixed by some non-trivial parabolic element of Γ. When $\Gamma = \mathrm{SL}_2(\mathbf{Z})$, the set of cusps is $\mathbf{Q} \cup \{\infty\}$. In view of (7) applied to the case $\gamma = T$ one sees that a modular form f of any weight for the group $\Gamma_0(N)$ satisfies

$$f(\tau + 1) = f(\tau) \ , \tag{8}$$

and, therefore, f defines a holomorphic function in the variable $q = e^{2\pi i \tau}$ for $0 < q < 1$. The condition in the definition of *modular form* that f should be *holomorphic at* ∞ means that f as a function of q is holomorphic at $q = 0$. Consequently, f admits an absolutely convergent Fourier expansion

$$f(\tau) = \sum_{m=0}^{\infty} c_m e^{2\pi i m \tau} \ , \tag{9}$$

which is a Taylor series in q.

For any cusp ρ of a congruence group Γ one may define the notion *holomorphic* at ρ for a modular form f by an analogous procedure using an arbitrary parabolic element of Γ that fixes ρ instead of T. For a given congruence group Γ two cusps ρ and σ are *equivalent* if there is some element γ in Γ such that $\sigma = \gamma \cdot \rho$. A modular form f is holomorphic at any cusp that is equivalent to another where it is holomorphic. The modular form f is a *cuspform* if, in addition to being holomorphic at each cusp, f vanishes at each cusp. For a given congruence group Γ a modular form vanishes at

[5]Details concerning the discussion in this section may be found in Shimura's book [24].

any cusp that is equivalent to another where it vanishes. The set of modular forms of given weight w forms a finite-dimensional vector space over $\mathbf{C}$ in which the set of cuspforms is a linear subspace of codimension bounded by the number of equivalence classes of cusps. In fact, using "Eisenstein series" one may show that the codimension of the space of cuspforms in the space of modular forms is often equal to the number of equivalence classes of cusps. For example, with the group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ there are no modular forms of odd weight, there is an Eisenstein series of every even weight greater than 2 that is not a cuspform, and every cusp is equivalent to ∞. Furthermore, since ∞ is the only zero of the cusp form Δ (of the preceding section) in the quotient $H^*/\Gamma(1)$ and since ∞ is a simple zero of Δ, every cuspform for $\Gamma(1)$ is divisible by Δ. Thus, in this case, there are no cuspforms of weight less than 12.

It is not difficult to see that the cuspforms of weight 2 for a congruence group Γ correspond to holomorphic differential 1-forms (differentials of the first kind) on the compact Riemann surface $X = H^*/\Gamma$. Thus, the dimension of the space of cuspforms of weight 2 is the *genus* of X. The fact that there are no cuspforms of weight 2 for the group $\Gamma(1)$ matches the previously mentioned fact that X is $\mathbf{P}^1$. It is certain of the cuspforms of weight two for the groups $\Gamma_0(N)$ that, according to the Shimura-Taniyama-Weil conjecture, parameterize the isogeny classes of elliptic curves defined over $\mathbf{Q}$.

5 Euler products

It will be recalled that the infinite series

$$\sum_{n=1}^{\infty} \frac{1}{n^s}$$

converges for $Re(s) > 1$ and gives rise by analytic continuation to a meromorphic function $\zeta(s)$ in $\mathbf{C}$. For $Re(s) > 1$ $\zeta(s)$ admits the absolutely convergent infinite product expansion

$$\prod_p \frac{1}{1-p^{-s}} \ ,$$

taken over the set of primes. This "Euler product" may be regarded as an analtyic formulation of the principle of unique factorization in the ring $\mathbf{Z}$

of integers. It is, as well, the product taken over all the non-archimedean completions of the rational field $\mathbf{Q}$ (which completions $\mathbf{Q}_p$ are indexed by the set of primes) of the "Mellin transform"[6] in $\mathbf{Q}_p$

$$\xi_p(s) = \frac{1}{1 - p^{-s}} \ ,$$

of the canonical "Gaussian density"

$$\Phi_p(x) = \begin{cases} 1 & \text{if } x \in \text{ closure of } \mathbf{Z} \text{ in } \mathbf{Q}_p \\ 0 & \text{otherwise} \end{cases} \ ,$$

which Gaussian density is equal to its own Fourier transform. For the archimedean completion $\mathbf{Q}_\infty = \mathbf{R}$ of the rational field $\mathbf{Q}$ one forms the classical Mellin transform

$$\xi_\infty(s) = \pi^{-(s/2)}\Gamma(s/2)$$

of the classical Gaussian density

$$\Phi_\infty(x) = e^{-\pi x^2} \ ,$$

(which also is equal to its own Fourier transform). Then the function

$$\xi(s) = \xi_\infty(s)\zeta(s) = \prod_{p \leq \infty} \xi_p(s)$$

is meromorphic in $\mathbf{C}$, and satisfies the functional equation

$$\xi(1-s) = \xi(s) \ . \tag{10}$$

The connection of Riemann's ζ-function with the subject of modular forms begins with the observation that $\zeta(2s)$ is essentially the Mellin transform of $\theta_I(x) = \theta(ix) - 1$, where θ, which is a modular form of weight $1/2$ and level 8, is defined in the upper-half plane H by the formula

$$\theta(\tau) = \sum_{m \in \mathbf{Z}} exp(\pi i \tau m^2) \ .$$

[6] The Mellin transform is, more or less, Fourier transform on the multiplicative group. Classically, the Mellin transform φ of f is given formally by $\varphi(s) = \int_0^\infty f(x)x^s(dx/x)$.

In fact, one of the classical proofs of the functional equation (10) is given by applying the Poisson summation formula[7] to the function $x \mapsto exp(\pi i \tau x^2)$, while observing that the substitution $s \mapsto (1/2) - s$ for $\zeta(2s)$ corresponds in the upper-half plane to the substitution $\tau \mapsto -1/\tau$ for the theta series.

If f is a cuspform for a congruence group Γ containing

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} ,$$

and so, consequently, $f(\tau+1) = f(\tau)$, then, as previously explained, one has the Fourier expansion (9)

$$f(\tau) = \sum_{m=1}^{\infty} c_m e^{2\pi i m \tau} .$$

The Mellin transform $\varphi(s)$ of f_I leads to the Dirichlet series

$$\varphi(s) = \sum_{m=1}^{\infty} c_m m^{-s} , \tag{11}$$

which may be seen to have a positive abscissa of convergence. One is led to the questions:

1. For which cuspforms f does the associated Dirichlet series $\varphi(s)$ admit an analytic continuation with functional equation?

2. For which cuspforms f does the associated Dirichlet series $\varphi(s)$ have an Euler product expansion?

For the "modular group" $\Gamma(1)$ the Dirichlet series associated to every cuspform of weight w admits an analtyic continution with functional equation under the substitution $s \mapsto w - s$. Since $\Gamma(1)$ is generated by the two matrices T and

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

and since the functional equation of a modular form f relative to T is reflected in the formation of the Fourier series (9), the condition that an absolutely

[7] On the other hand, (10) may be regarded directly as a *divergent* model of the Poisson summation formula.

convergent series (9) is a modular form for $\Gamma(1)$ is the functional equation for a modular form relative solely to W. This is equivalent to the (properly formulated) functional equation for the associated Dirichlet series φ together with a "growth condition". For the group $\Gamma_0(N)$, with $N > 1$, the question of a functional equation is more complicated since, although T is available, there is no reason for a cuspform to satisfy a law of transformation relative to W. But note that for any Γ the set of cuspforms of given weight for which the associated Dirichlet series have analytic continuations satisfying a given finite set of functional equations is a vector space. On the other hand, there is no reason to believe, even for level 1, that the cuspforms admitting an Euler product expansion form a vector space.

In a nutshell the cuspforms admitting Euler products are those which arise as eigenforms for an arithmetically defined commutative algebra of semi-simple operators on the space of cuspforms of a given weight introduced by E. Hecke. The theory of Hecke operators is reasonably simple for level 1 but somewhat more complicated in general (see, e.g., Shimura's book [24]).

Observing that the formula

$$ds^2 = \frac{dx^2 + dy^2}{y^2} \ , \ \text{for } \tau = x + iy \in H \ ,$$

gives a (the hyperbolic) $\mathrm{SL}_2(\mathbf{R})$-invariant metric in H with associated invariant measure

$$d\mu = \frac{dx dy}{y^2} \ ,$$

one introduces the Petersson (Hermitian) inner product in the space of cuspforms of weight w for Γ with the definition:

$$\langle f, g \rangle = \int_{H/\Gamma} f(\tau)\, \overline{g(\tau)}\, Im(\tau)^w \, d\mu(\tau) \ . \tag{12}$$

(Integration over the quotient H/Γ makes sense since the integrand

$$f(\tau)\, \overline{g(\tau)}\, y^w$$

is Γ-invariant.)

For the modular group $\Gamma(1)$ the n^{th} Hecke operator $T(n) = T_w(n)$ is the linear endomorphism of the space of cuspforms of weight w arising from the following considerations. Let S_n be the set of $2{\times}2$ matrices in $\mathbf{Z}$ with determinant n. For

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in S_n$$

and for a function f in H one defines

$$(M \cdot_w f)(\tau) \;=\; \det(M)^{w-1}(c\tau + d)^{-w} f(\tau)\;, \tag{13}$$

and then, observing that $\Gamma(1)$ under $\cdot_w$ acts trivially on the modular forms of weight w, one may define the Hecke operator $T_w(n)$ by

$$T_w(n) = \sum_{M \in S_n/\Gamma(1)} (M \cdot_w f)(\tau)\;, \tag{14}$$

where the quotient $S_n/\Gamma(1)$ refers to the action of $\Gamma(1)$ by left multiplication on the set S_n. One finds for m, n coprime that

$$T(mn) = T(m)T(n)\;,$$

and furthermore one has

$$T(p^{e+1}) = T(p^e)T(p) - p^{w-1}T(p^{e-1})\;.$$

Consequently, the operators $T(n)$ commute with each other, and, therefore, generate a commutative algebra of endomorphisms of the space of cusp forms of weight w for $\Gamma(1)$. It is not difficult to see that the Hecke operators are self-adjoint for the Petersson inner product on the space of cuspforms. Consequently, the space of cuspforms of weight w admits a basis of simultaneous eigenforms for the Hecke algebra. A "Hecke eigencuspform" is said to be *normalized* if its Fourier coefficient $c_1 = 1$. If f is a normalized Hecke eigencuspform, then

- The Fourier coefficient c_m of f is the eigenvalue of f for $T(m)$.
- The Fourier coefficients $c(m) = c_m$ of f satisfy

$$c(mn) = c(m)c(n) \;\text{ for } m{,}n \text{ coprime, and}$$

$$c(p^{e+1}) = c(p^e)c(p) - p^{w-1}c(p^{e-1}) \;\text{ for } p \text{ prime.}$$

Consequently, the Dirichlet series associated with a simultaneous Hecke eigencuspform of level 1 and weight w admits an Euler product

$$\varphi(s) = \prod_p \frac{1}{1 - c_p p^{-s} + p^{w-1-2s}} \ . \tag{15}$$

For example, when f is the unique normalized cuspform Δ of level 1 and weight 12, one has

$$\varphi(s) = \prod_p \frac{1}{1 - \tau(p) p^{-s} + p^{11-2s}} \ ,$$

where $c_p = \tau(p)$ is the function τ of Ramanujan.

For the congruence group $\Gamma_0(N)$ a Hecke eigencuspform of weight w gives rise to a Dirichlet series $\varphi(s)$ that admits an Euler product expansion whose factors at primes p coprime to N resemble those given by (15). In order for $\varphi(s)$ to satisfy a functional equation under the substitution $s \mapsto w - s$, one needs to require that the eigencuspform f admits a functional equation not only with respect to each element of the group $\Gamma_0(N)$ but also with respect to the substitution in the upper-half plane H given by the matrix

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

A. Weil ([31]) showed that the cuspforms of weight 2 for the group $\Gamma_0(N)$ satisfying the appropriate functional equation under the mapping of H given by W_N correspond precisely to Dirichlet series with certain growth conditions that admit analytic continuations as meromorphic functions in $\mathbf{C}$ satisfying a finite number of "twisted" functional equations.

The reader will have noticed that it is not extremely easy to characterize the cuspforms of weight 2 that conjecturally (Shimura-Taniyama-Weil) parameterize the isogeny classes of elliptic curves defined over the rational field $\mathbf{Q}$. The Euler product is an extremely important part of the characterization since the Dirichlet series given by such an elliptic curve, as will be made explicit in the next section, is, by its very nature, an Euler product. Weil conjectures explicitly that the Dirichlet series with Euler product given by each elliptic curve defined over $\mathbf{Q}$ satisfies these conditions, i.e., is the Dirichlet series associated to some W_N-compatible Hecke eigencuspform for the group $\Gamma_0(N)$, where N is the "conductor" of E. This has led to efforts, related to the "Langlands program" to understand the W_N-compatible Hecke eigencuspforms in a more intrinsic way as objects of representation theory over $\mathbf{Q}$ (see, e.g., the survey of Gelbart [4]).

6 Elliptic curves over the rational field Q

Let E be an elliptic curve defined over $\mathbf{Q}$. One may clear denominators from its cubic equation, if necessary, in order to arrive at an equation with integer coefficients having no common factor. While the Weierstrass normal form (3) is available to represent the isomorphism class of any elliptic curve over a field of characteristic different from 2 and 3, one needs the generalized Weierstrass form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{16}$$

over an arbitrary field, and, moreover, for each elliptic curve E defined over $\mathbf{Q}$ there is a "best possible" equation (e.g., see Silverman [27]) of the form (16) with integer coefficients called the *Neron model* of E. With an abuse of notation E will denote the Neron model, which may be regarded as a "curve over $\mathbf{Z}$". (One would want to call it an "elliptic curve over $\mathbf{Z}$" if it were "smooth over $\mathbf{Z}$", i.e., if it had good reduction at each prime p; the fact that every Neron model has bad reduction at least once corresponds under the "dictionary" to the fact that there are no cuspforms of weight two and level 1.) It then may be observed that for each prime p the Neron model gives rise to a cubic equation over the finite field $\mathbf{F}_p$. For all but a finite number of p the equation over $\mathbf{F}_p$ is non-singular over $\overline{\mathbf{F}_p}$, i.e., determines an elliptic curve E_p defined over $\mathbf{F}_p$. One says in this case that E has "good reduction" mod p. Following Tate ([30]) one introduces

$$b_2 = a_1^2 + 4a_2 \ ,$$

$$b_4 = a_1a_3 + 2a_4 \ ,$$

$$b_6 = a_3^2 + 4a_6 \ , \text{ and}$$

$$b_8 = b_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2 \ .$$

Then one has

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \ .$$

The non-vanishing of Δ mod p is necessary and sufficient for E to have good reduction mod p. It follows that a prime p divides Δ if and only if E does not have good reduction mod p. If p is a prime for which E has "bad reduction", then there is a single singular point of the reduced curve E_p, and either (a) E_p has distinct tangent lines at the singular point (*semi-stable* reduction) or (b) E_p has a single tangent line occurring with multiplicity 2. E is called *semi-stable* if it has either good or semi-stable reduction at each prime. The *conductor* of E is the integer N defined by

$$N = \prod_p p^{\nu_p} \ ,$$

where

$$\nu_p = \begin{cases} 0 & \text{if } E \text{ has good reduction at } p \ . \\ 1 & \text{if } E \text{ has semi-stable reduction at } p \ . \\ 2 + \lambda_p \geq 2 & \text{otherwise.} \end{cases}$$

The non-negative integer λ_p cannot be positive unless p is 2 or 3. Tautologically, E is semi-stable if and only if its conductor N is square-free.

One defines the "L-series" of E by

$$L(E,s) \ = \ \prod_{p|N} \frac{1}{1 - c_p p^{-s}} \ \prod_{p \nmid N} \frac{1}{1 - c_p p^{-s} + p^{1-2s}} \ , \tag{17}$$

where c_p is defined when E has good reduction mod p by the formula

$$c_p = p + 1 - |E(\mathbf{F}_p)| \ ,$$

and c_p is defined when E has bad reduction mod p by

$$c_p = \begin{cases} 1 & \text{if } \nu_p = 1 \text{ and the tangents are defined over } \mathbf{F}_p \ . \\ -1 & \text{if } \nu_p = 1 \text{ with "irrational" tangents} \ . \\ 0 & \text{if } \nu_p > 1 \ . \end{cases}$$

One observes readily that the L-function of E codifies information about the number of points of E in the finite field $\mathbf{F}_p$. Quite generally for an algebraic variety defined over $\mathbf{Q}$ the analogous codification of information obtained by counting points in the various reductions mod p of the variety yields the "Hasse-Weil zeta function", which reflects "cohomological" information about E. The L-series of E is the essential part, corresponding to cohomology in dimension 1, of the Hasse-Weil zeta function of E. The Hasse-Weil zeta function is a special case of the general notion (Serre [19]) of "zeta function" for a *scheme of finite type* over $\mathbf{Z}$.

One observes that $L(E,s)$ resembles, at least insofar as one considers its Euler factors for primes p corresponding to good reductions of E, the Dirichlet series associated to a cuspform of weight 2 that admits an Euler product expansion. The observation of this resemblance is the beginning of an appreciation of the Shimura-Taniyama-Weil conjecture. One is led to ask to what extent the two classes of Dirichlet series with Euler products coincide. The conjecture states that the L-function of an elliptic curve defined over $\mathbf{Q}$ with conductor N arises from a cuspform for the group $\Gamma_0(N)$ that is compatible with the substitution in the upper-half plane H given by W_N. Isogenous elliptic curves have the same L-function, and, conversely (cf. Tate [29] and Faltings [3]) two elliptic curves with the same L-function must be isogenous. Thus, the idea of the conjecture is that the isogeny classes of elliptic curves defined over $\mathbf{Q}$ with conductor N are in bijective correspondence with the set of Hecke eigencuspforms for the group $\Gamma_0(N)$ of level N, compatible with the extension of that group by the substitution arising from W_N, having rational Fourier coefficients and not arising from levels dividing N.

7 The Shimura map

Shimura ([23], [24], [25]) showed for a given W_N-compatible Hecke eigencuspform f of weight 2 for the group $\Gamma_0(N)$ with rational Fourier coefficients how to construct how to construct an elliptic curve E_f defined over $\mathbf{Q}$ such that the Dirichlet series $\varphi(s)$ associated with f is the same as the L-function $L(E_f,s)$. Thus, the Shimura-Taniyama-Weil conjecture becomes the statement that Shimura's map from the set of such cuspforms to the set of elliptic curves defined over $\mathbf{Q}$ is surjective up to isogeny. A rough description of the Shimura map follows.

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$, and let $X(\Gamma)$ denote the compact Riemann surface H^*/Γ. The inclusion of Γ in $\Gamma(1)$ induces a "branched covering"

$$X(\Gamma) \longrightarrow X(1) \cong \mathbf{P}^1 \ .$$

One may use the elementary Riemann-Hurwitz formula from combinatorial topology to determine the Euler number, and consequently the genus, of $X(\Gamma)$. The genus is the dimension of the space of cuspforms of weight 2. Even when the genus is zero one obtains embeddings of $X(\Gamma)$ in projective spaces $\mathbf{P}^r$ through holomorphic maps

$$\tau \longmapsto (f_0(\tau), f_1(\tau), \ldots, f_r(\tau)) \ ,$$

where $f_0, f_1, \ldots, f_r$ is a basis of the space of modular forms of weight w with w sufficiently large. For example, any $w \geq 12$ will suffice for $\Gamma(1)$. For $\Gamma_0(N)$ (but not for arbitrary Γ) one may find a basis of the space of modular forms of weight w having rational Fourier coefficients. Using the corresponding projective embedding one finds a *model* for $X_0(N) = X(\Gamma_0(N))$ over $\mathbf{Q}$, i.e., an algebraic curve defined over $\mathbf{Q}$ in projective space that is isomorphic as a compact Riemann surface to $X_0(N)$.

Associated with any "complete non-singular" algebraic curve (i.e., after Weyl [33], any compact Riemann surface) X of genus g is a complex torus, the "Jacobian" $J(X)$ of X, that is the quotient of g-dimensional complex vector space $\mathbf{C}^g$ by the lattice Ω generated by the "period matrix", which is the $g \times 2g$ matrix in $\mathbf{C}$ obtained by integrating each of the g members ω_i of a basis of the space of holomorphic differentials over each of the $2g$ loops in X representing the members of a homology basis in dimension 1. Furthermore, if one picks a base point z_0 in X, then for any z in X, the path integral from z_0 to z of each of the g holomorphic differentials is well-defined modulo the periods of the differential. One obtains a holomorphic map $X \rightarrow J(X)$ from the formula

$$z \longmapsto (\int_{z_0}^{z} \omega_1, \ldots, \int_{z_0}^{z} \omega_g) \ mod \ \Omega \ .$$

This map is, in fact, universal for pointed holomorphic maps from X to complex tori. Furthermore, the Jacobian $J(X)$ is an algebraic variety that admits definition over any field of definition for X and z_0, and the universal map also admits definition over any such field. The complex tori that admit embeddings in projective space are the abelian group objects in the category of projective varieties. They are called *abelian varieties*. Every abelian variety is isogenous to the product of "simple" abelian varieties: abelian varieties having no abelian subvarieties. Shimura showed that one of the simple isogeny factors of $J(X_0(N))$ is an elliptic curve E_f defined over $\mathbf{Q}$ characterized by the fact that its one-dimensional space of holomorphic differentials induces on $X_0(N)$ via the composition of the universal map with projection on E_f the one-dimensional space of differentials on $X_0(N)$ determined by the cuspform f. He showed further that $L(E_f, s)$ is the Dirichlet series $\varphi(s)$ with Euler product given by f. An elliptic curve E defined over $\mathbf{Q}$ is said to be *modular* if it is isogenous to E_f for some W_N-compatible Hecke eigencuspform of weight 2 for $\Gamma_0(N)$. Equivalently E is modular if and only if $L(E, s)$ is the Dirichlet series given by such a cuspform. The Shimura-Taniyama-Weil Conjecture states that every elliptic curve defined over $\mathbf{Q}$ is modular. Shimura [23] showed that this conjecture is true in the special case where the $\mathbf{Z}$-module rank of the ring of endomorphisms of E is greater than one. In this case the point τ (notation of section 3) of the upper-half plane corresponding to $E(\mathbf{C})$ is a quadratic imaginary number, and $L(E, s)$ is a number-theoretic L-function associated with the corresponding imaginary quadratic number field.

8 The hypothetical Frey curve

Let $p \geq 5$ be a prime. Based on the assumption, which presumably is false, that there are non-zero integers a, b, c such that $a^p + b^p + c^p = 0$, G. Frey observed that the elliptic curve given by the equation

$$y^2 = x(x - a^p)(x + b^p) \ , \tag{18}$$

which is certainly defined over $\mathbf{Q}$, would not be likely to be modular. Thus, if the Shimura-Taniyama-Weil Conjecture were true, then "Fermat's Last Theorem" would also be true. By 1987 it had been shown through the efforts of Frey, Ribet and Serre that the Frey curve (18) is not modular. The proof involves the systematic study of what is known as the "ℓ-adic representation of an elliptic curve defined over $\mathbf{Q}$, which is described in the next section. This same technique is what has been reported to be the basis of the proof of Wiles that every semi-stable elliptic curve defined over $\mathbf{Q}$ is modular. The Frey curve (18) has discriminant $\Delta = (abc)^p$. It is only slightly difficult to see that it is semi-stable, and, therefore, that its conductor N is the square-free integer abc. If the Frey curve is modular, one is led to a cuspform of weight 2 for $\Gamma_0(abc)$. The theory of ℓ-adic representations leads one along a path of reductions of the level N from the initial level abc that enables one to conclude that there is a cuspform of weight 2 for $\Gamma_0(2)$; but the genus of $X_0(2)$ is 0, and, consequently, there is no such cuspform.

9 ℓ-adic representations of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$

Let E be an elliptic curve defined over $\mathbf{Q}$. Inasmuch as the group law $E \times E \to E$ is defined over $\mathbf{Q}$ it follows that for each integer m the group (scheme) $E[m]$ of m-torsion points, i.e., for any field K containing $\mathbf{Q}$ the group $E[m](K)$ consisting of all x in $E(K)$ such that $mx = 0$, is defined by equations with rational coefficients. Consequently, any automorphism of K must carry the group $E[m](K)$ into itself. Since $E(\mathbf{C})$ is the quotient of $\mathbf{C}$ by a lattice, it is clear that $E[m](\mathbf{C})$ is isomorphic to $\mathbf{Z}/m\mathbf{Z} \times \mathbf{Z}/m\mathbf{Z}$; in fact, this latter group is isomorphic to $E[m](K)$ for each algebraically closed field of characteristic 0. There is a unique ring homomorphism $\mathbf{Z}/mn\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z}$ for each integer $n \geq 1$, and the family of these ring homomorphisms gives rise to an inverse system in the category of commutative rings. If one specializes to the case $m = \ell^r$, where ℓ is prime, the projective limit is the ring $\mathbf{Z}_\ell$ of ℓ-adic integers. The groups $E[m]$ form a direct system with respect to the inclusions $E[m] \subseteq E[mn]$, but, corresponding to the inverse system of the groups $\mathbf{Z}/m\mathbf{Z}$, form an inverse system (the Tate system) with respect to the family of homorphisms $E[mn] \to E[m]$ defined by $x \mapsto nx$. If one specializes to the case $m = \ell^r$, where ℓ is prime, one obtains the projective limit

$$T_\ell(E) = \text{proj lim}_{r\to\infty} E[\ell^r](\bar{\mathbf{Q}}) \cong \mathbf{Z}_\ell \times \mathbf{Z}_\ell \ , \tag{19}$$

which is isomorphic to the cohomology module

$$H^1(E, \mathbf{Z}_\ell) \ .$$

The action of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ on the torsion groups $E[m]$ induces an action of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ on the projective limit $T_\ell(E)$. This action gives rise to a representation

$$\rho_\ell : \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}_2(\mathbf{Z}_\ell) \ ,$$

which is called the ℓ-adic representation of E. In considering ρ_ℓ one is reminded of the action of the automorphism group of a manifold M on the cohomology $H^*(M)$ and, more particularly, the action of $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$ on the cohomology of M when M is an algebraic manifold in $\mathbf{P}^n(\mathbf{C})$ defined by equations with *real* coefficients, but one must keep in mind that the transformations of $E(\bar{\mathbf{Q}})$ arising from the elements of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ are not even remotely continuous in the classical topology on $E(\mathbf{C})$. More generally, there is an *algebraic* way of defining the cohomology ring $H^*(M, \mathbf{Z}_\ell)$ (see Tate [28]) when M is an algebraic variety with the property that automorphisms fixing the field of definition act on $H^*(M, \mathbf{Z}_\ell)$. An introduction to the study of ρ_ℓ may be found in Serre's "Montreal Notes" [21].

The canonical ring homomorphism from the ring $\mathbf{Z}_\ell$ of ℓ-adic integers to the field $\mathbf{Z}/\ell\mathbf{Z}$ induces a group homomorphism $A \mapsto \overline{A}$, called *reduction* mod ℓ, from the group $\mathrm{GL}_2(\mathbf{Z}_\ell)$ to the finite group $\mathrm{GL}_2(\mathbf{Z}/\ell\mathbf{Z})$. An ℓ-adic representation ρ of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ is called *modular* if it is isomorphic to the representation ρ_ℓ arising from the elliptic curve E_f that is the image under the Shimura map of a modular form f. A representation

$$\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Z}/\ell\mathbf{Z})$$

is called *modular* if it is isomorphic to $\overline{\rho_\ell}$ for some modular ℓ-adic representation ρ_ℓ. In the extensive detailed study of representations of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ particular attention has been paid to the question of when a representation in $\mathrm{GL}_2(\mathbf{Z}_\ell)$ is modular and also to the question of when a representation of $\mathrm{GL}_2(\mathbf{Z}/\ell\mathbf{Z})$ is modular. Under certain conditions (see Serre [22] and Ribet [14], [15]) one can show that ρ_ℓ is modular if $\overline{\rho_\ell}$ is modular, i.e., ρ_ℓ is modular if it is *congruent* mod ℓ to a modular ℓ-adic representation. Such arguments are central both to the work of Ribet in showing that the Shimura-Taniyama-Weil conjecture implies "Fermat's Last Theorem" and to the reported work of Wiles in proving that semi-stable elliptic curves are modular. In the work of Ribet the basic idea is that the modularity of the Frey curve, which has square-free conductor $N = abc$, implies the existence of a cusp form of weight 2 and level N. By using an argument at the scene of the mod ℓ representations, Ribet shows that one may split each odd prime divisor out of the level N and arrive at the conclusion that there is a cusp form of weight 2 and level 2, which is not possible.

APPENDIX: Current Status

Andrew Wiles posted the following announcement in the "UseNet" electronic news group called "sci.math":

From: wiles@rugola.Princeton.EDU (Andrew Wiles)
Newsgroups: sci.math
Subject: Fermat status
Message-ID: <1993Dec4.013650.12700@Princeton.EDU>
Date: 4 Dec 93 01:36:50 GMT
Sender: news@Princeton.EDU (USENET News System)
Organization: Princeton University
Lines: 21
Originator: news@nimaster
Nntp-Posting-Host: rugola.princeton.edu

In view of the speculation on the status of my work on the Taniyama-Shimura conjecture and Fermat's Last Theorem I will give a brief account of the situation. During the review process a number of problems emerged, most of which have been resolved, but one in particular I have not yet settled. The key reduction of (most cases of) the Taniyama-Shimura conjecture to the calculation of the Selmer group is correct. However the final calculation of a precise upper bound for the Selmer group in the semistable case (of the symmetric square representation associated to a modular form) is not yet complete as it stands. I believe that I will be able to finish this in the near future using the ideas explained in my Cambridge lectures.

The fact that a lot of work remains to be done on the manuscript makes it still unsuitable for release as a preprint . In my course in Princeton beginning in February I will give a full account of this work.

Andrew Wiles.

References

[1] Lars V. Ahlfors, *Complex Analysis*, 3rd edition, McGraw Hill, 1979.

[2] A. Borel & W. Casselman, eds., *Automorphic Forms, Representations, and L-Functions*, *Proceedings of Symposia in Pure Mathematics*, vol. 33, (bound in two parts), American Mathematical Society, 1979.

[3] G. Cornell & J. Silverman, eds., *Arithmetic Geometry*, Springer-Verlag, 1986.

[4] S. Gelbart, "Elliptic curves and automorphic representations", *Advances in Mathematics*, vol. 21 (1976), pp. 235-292.

[5] ———, "An elementary introduction to the Langlands program", *Bulletin of the American Mathematical Society*, vol. 10 (1984), pp. 177-219.

[6] G. H. Hardy & E. M. Wright, *An Introduction to the Theory of Numbers*, 4th edition, Oxford University Press, 1960.

[7] J.-I. Igusa, *Theta Functions*, *Die Grundlehren der mathematischen Wissenschaften*, vol. 194, Springer-Verlag, 1972.

[8] Y. Ihara, K. Ribet, & J.-P. Serre, eds., *Galois Groups over* $\boldsymbol{Q}$, *Mathematical Sciences Research Institute Publications*, no. 16, Springer-Verlag, 1989.

[9] W. Kuyk et al., eds., *Modular Functions of One Variable*, I, II, ..., VI, *Lecture Notes in Mathematics*, nos. 320, 349, 350, 476, 601, 627, Springer-Verlag, 1973-1977.

[10] R. P. Langlands, "Representation theory and arithmetic", Lecture at the Symposium on the Mathematical Heritage of Hermann Weyl, *Proceedings of Symposia in Pure Mathematics*, vol. 48, pp. 25-33, American Mathematical Society, 1988.

[11] B. Mazur, "Arithmetic on curves", *Bulletin of the American Mathematical Society*, vol. 14 (1986), pp. 207-259.

[12] ———, "Number Theory as Gadfly", *The American Mathematical Monthly*, vol. 98 (1991), pp. 593-610.

[13] A. Ogg, *Modular Forms and Dirichlet Series*, W. A. Benjamin, Inc., 1969.

[14] K. Ribet, "On modular representations of Gal($\bar{\mathbf{Q}}/\mathbf{Q}$) arising from modular forms", *Inventiones Mathematicae*, vol. 100 (1990), pp. 431-476.

[15] ———, "From the Taniyama-Shimura conjecture to Fermat's Last Theorem", *Annales de la Faculté des Sciences de Toulouse*, vol. 11 (1990), pp.116-139.

[16] ———, "Wiles proves Taniyama's conjecture; Fermat's Last Theorem follows", *Notices of the American Mathematical Society*, vol. 40 (1993), pp. 575-576.

[17] K. Ribet, ed., *Current Trends in Arithmetical Algebraic Geometry*, *Contemporary Mathematics*, vol. 67, American Mathematical Society, 1987.

[18] K. Rubin & A. Silverberg, "Wiles' proof of Fermat's Last Theorem", preprint, November, 1993.

[19] J.-P. Serre, "Zeta and L functions", *Arithmetical Algebraic Geometry: Proceedings of a Conference Held at Purdue University, December 5-7, 1963*, Harper & Row, Publishers, 1965.

[20] J.-P. Serre, *A Course in Arithmetic*, *Graduate Texts in Mathematics*, vol. 7, Springer-Verlag, 1973.

[21] ———, *Abelian l-adic Representations and Elliptic Curves*, W. A. Benjamin, Inc., 1968.

[22] ———, "Sur les représentations modulaires de degré 2 de Gal($\bar{\mathbf{Q}}/\mathbf{Q}$)", *Duke Mathematical Journal*, vol. 54 (1987), pp. 179-230.

[23] G. Shimura, "On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields", *Nagoya Mathematical Journal*, vol. 43 (1971), pp. 199-208.

[24] ———, *Introduction to the Arithmetic Theory of Automorphic Forms*, Iwanami Shoten Publishers and Princeton University Press, 1971.

[25] ———, "On the factors of the jacobian variety of a modular function field", *Journal of the Mathematical Society of Japan*, vol. 25 (1973), pp. 523-544.

[26] C. L. Siegel, *Topics in Complex Function Theory*, 3 volumes, Wiley-Interscience, 1969, 1971, 1973.

[27] J. H. Silverman, *The Arithmetic of Elliptic Curves*, *Graduate Texts in Mathematics*, vol. 106, Springer-Verlag, 1986.

[28] J. T. Tate, "Algebraic cycles and poles of zeta functions", *Arithmetical Algebraic Geometry: Proceedings of a Conference Held at Purdue University, December 5-7, 1963*, Harper & Row, Publishers, 1965.

[29] ———, "Endomorphisms of abelian varieties over finite fields", *Inventiones Mathematicae*, vol. 2 (1966), pp. 134-144.

[30] ———, "The arithmetic of elliptic curves", *Inventiones Mathematicae*, vol. 23 (1974), pp. 179-206.

[31] A. Weil, "Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen", *Mathematische Annalen*, vol. 168 (1967), pp. 149-56.

[32] ———, *Dirichlet Series and Automorphic Forms*, *Lecture Notes in Mathematics*, no. 189, Springer-Verlag, 1971.

[33] Hermann Weyl, *The Concept of a Riemann Surface*, (English translation by Gerald R. MacLane), Addison-Wesley, 1964.