Revista Colombiana de Matemáticas
Volumen **29** *(1995), páginas 49–88*

Fermat's Last Theorem: From Fermat to Wiles*

CARLOS JULIO MORENO
City University of New York

> "Es un secreto del oficio que no obedece a las leyes de la inteligencia sino a la magia de los instintos, como sabe la cocinera cuando está la sopa."
> Gabriel García Márquez

1. Introduction

The solution to Fermat's Last Theorem was announced by Wiles (Cambridge, June 23, 1993)[†] as a consequence of his proof of the Shimura-Taniyama-Weil conjecture for semi-stable elliptic curves. Even though Wiles stands as the principal architect of the proof, his methods are built on the ideas of many mathematicians. It is the purpose of this paper to survey some of the principal developments leading up to the final solution.

Even though the course we have charted in this article touches repeatedly on the work of Wiles, it is hoped that the reader will find a demonstration of the continuous influence that Fermat's Last Theorem has had on number theory and of the significant contributions of many other mathematicians among whom we must count Kummer, Herbrand, and Iwasawa.

This article is divided into three parts. **Part I** deals with a brief description of the path leading from Fermat's original formulation to Wiles' final solution. **Part II** describes in greater detail the themes that have directly and indirectly

[*] This expository paper originated from the manuscript of the author's upcoming book *Fermat's Last Theorem: From Fermat to Wiles*, and has been specially adapted from it by invitation of the Managing Editor.

[†] See the appendix to this paper for a timetable of announcements.

contributed to the solution. Lastly, **Part III** will attempt to give a technical outline of the proof in as elementary a way as that is possible.

Part I: The Transition

2. From Fermat to Wiles

Fermat made many *observations on arithmetic* which he wrote down on his copy of Bachet's edition of the *'Arithmetic of Diophantus'*. Among these the most notorious one is his claim that he had found a marvelous proof of the fact that *it is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second into two like powers....* Since then, many mathematicians have tried to verify his claim. Fermat himself must be credited with the discovery of two of the most fundamental principles that have come to play an important role in any discussion of the problem, *e.g.* the controlling effect that unique factorization has on diophantine questions and the method of infinite descent.

After Fermat, Euler was the first mathematician to make new contributions to the problem, particularly in the case of third powers. In this case the equation can be written in the form

$$\begin{aligned} x^3 = z^3 - y^3 &= (z-y)(z^2+xy+y^2) \\ &= (z-y)(z-\rho y)(z-\rho^2 y), \end{aligned}$$

where ρ is a root of the equation $\rho^2+\rho+1=0$. Euler presented during his lifetime two apparently different approaches to the problem. First in 1760 he mentions a proof which is based on the first equation above, namely on the theory of the quadratic form $x^2+xy+y^2 = X^2+3Y^2, \quad X = x+y/2, Y = y/2$. Secondly in 1770 he uses the second equation above which leads to the field $\mathbb{Q}(\sqrt[2]{-3})$. The first approach was undoubtedly the one Fermat had envisaged. In retrospect we see that both of Euler's ideas lead to the properties of factorization in the ring

$$\mathbb{Z}[\rho] := \{a + b\rho : a, b \text{ integers}\}\,.$$

This is perhaps Euler's most lasting contribution to the solution of the problem. The idea that factorization in domains which are extensions of the ordinary integers must play a role in the analysis of the arithmetic properties of possible solutions to the equation. All mathematicians that have worked on the problem have had to contend with the properties of unique factorization, or the lack of it, as it applies to the ring of integers $\mathbb{Z}[\zeta_p]$ (the first instance of this failure occurs, as Kummer discovered, when $p = 23$).

Gauss, who was the first to write down a fairly complete discussion of Euler's case $n = 3$, is reported to have said that he did not want to work on Fermat's problem because of its difficulty. Neglecting the possibility that there may have been other reasons, it is interesting to note that Gauss was the first mathematician to develop a general theory of cyclotomic numbers, which are complex numbers needed for the regular division of the circle (ρ corresponds to trisection). This was done in the seventh chapter of his *'Disquicitiones-Arithmeticae'*. That theory is considered today the foundation of the mathematical edifice on which the whole study of Fermat's equation is based. It is also worth mentioning that Gauss promoted very enthusiastically the number theoretic work of Sophie Germain, most of which dealt with particular criteria to decide the solvability of the equation in certain cases.

Subsequent to the work of Gauss, early in the nineteenth century, other important mathematicians toyed around with the theme of unique factorization, and went beyond Euler to deal successfully with cases $n = 5, 7, 13$. The first significant breakthrough came in the late 1840's when apparently Dirichlet and Kummer thought they had found a proof for the general problem. They soon realized their failure by noticing that unique factorization does not hold in general. It was then left to Kummer to turn a failure into success by creating the theory of *ideal numbers* to deal with the lack of unique factorization.

By the end of the Nineteenth century, many cases of the problem had been solved. In 1893 the German mathematical society had commissioned Hilbert to write a *'Report on Number Theory'*. It was on this occasion that Hilbert presented the synthesis of algebraic number theory that lead him to the formulation of general principles for the study of abelian extensions of number fields. In the final §36 of the *Report*, Hilbert presents his proof that Fermat's Last Theorem holds true in any regular cyclotomic field, thus extending earlier work of Kummer. Very little is said in the *Report* concerning the irregular case, although the French translation did appended a lengthy discussion of Kummer's 1857 Memoir. Hilbert's *Report* was a turning point in the efforts to solve Fermat's problem and quickly lead to new and exciting progress, *e.g.* the sensational result of Wieferich (Crelle, 1909) to the effect that if $x^p + y^p = z^p$ has a solution in the first case ($\gcd(xyz, p) = 1$), then

$$\frac{2^{p-1} - 1}{p} \equiv 0 \pmod{p}.$$

The first twenty years of the twentieth century saw the glorious development of class field theory, culminating in the work of Takagi which in turn paved the way to Artin's formulation. The Dickson-type report prepared by Wahlin and Vandiver (1927), and the class field theory report of Hasse (1928) recapitulated the advances in number theory which had resulted from work related to Fermat's Last Theorem. Herbrand published his results on cyclotomic extensions in 1932 and this was followed by his own report on *'Le développment moderne de la théorie des corps algébriques'*, both published posthumously with the editorial assistance of Chevalley. The latter very clearly showed the benefits of dealing with Fermat's Last Theorem with the assistance of class field theory (see [5], Chap. III, § V). With the hindsight of the recent work, we can now see the pivotal importance that Herbrand's work had on the problem of the construction of unramified extensions of irregular cyclotomic fields. The full significance of Herbrand's work emerged only through the fundamental work of Iwasawa and his school. The rest is recent history and deals with one of the central problems of number theory, namely the nature of the solutions of diophantine equations of degree three and the intimate connection of this topic with the theory of modular forms.

The modern history of Fermat's Last Theorem begins with the work of Taniyama, Shimura and Weil which suggested that the Hasse-Weil zeta function of an elliptic curve defined over the rationals is an automorphic L-function, that is to say, the elliptic curve can be reconstructed from knowledge of the coefficients in the Dirichlet series associated to its L-function. The final solution of Fermat's Last Theorem by Wiles incorporates the results of many people, as well as ideas of his own. It is a vindication of the importance of the work of all the mathematicians that have labored to complete Grothendiecks' reformulation of algebraic geometry as well as Langlands' Program. On the purely arithmetic side one must also realize that certain objects have emerged that play as important a role as that which the class group played in the past, namely the notion of Selmer group. The old theme of the Dirichlet class number formulas has revealed its Protean face and now pervades all the deeper levels of number theory.

We do not wish to enter here more deeply into the history of the many attempts to solve the problem; this will be done in part in other chapters. We do want to invite any prospective student of Fermat's equation to consult the work of Mahony [1], where an interesting description of the mathematical work of Fermat is given. The relevant chapter which discusses the number theory work is entitled "Between Traditions". We must also add that the reader should consult Andre Weil's review of Mahony's book [2]; this is a critical and caustic review which is guaranteed to shake one's confidence in the art of the mathematical historian, unless of course one consults Weil's own attempt to come to terms with Fermat as he does brilliantly in his *'Number Theory: An approach through history'*.

An item of gossip which is worth mentioning is Halmos' note next to Gerd Falting's picture (#579) in his *Photographic Memory* book [3] where it is remarked that an important consequence of Falting's work on the Mordell Conjecture is the fact that for a given n, Fermat's equation can have at most a finite number of solutions. Heath-Brown showed in ([31]) that this implies that the number of exponents n for which Fermat's Last Theorem is true is of density 1; this in the jargon of probability theory ascertains that Fermat's Last Theorem is almost certainly true.

Part II: The Work of Wiles

The purpose of this section is to survey those areas of number theory in which the work of A. Wiles has had an impact. Some of the ideas that we will review have their origen in the work of Kummer and our presentation is intended to exhibit their place in the continuous development of modern number theory. In order of appearance these are

(*i*) The Birch and Swinnerton-Dyer Conjecture.
(*ii*) The Herbrand-Ribet Theorem.

3. The Birch and Swinnerton-Dyer Conjecture

An important insight into the nature of the rational solutions of elliptic curves was obtained by Birch and Swinnerton-Dyer after extensive calculations which were supposed to clarify the validity of the local to global principle for algebraic curves of degree 3. Siegel's work, which dealt exclusively with quadratic forms, had successfully extended the well known class number formulas of Dirichlet so that the local to global principle, as developed earlier by Minkowski and Hasse, became equivalent to the evaluation of a certain L-value, a quantity which subsequently gave rise to the notion of Tamagawa measure. In analogy with these results, Birch and Swinnerton-Dyer proposed a formula which was supposed to relate local information about an elliptic curve at all primes with the arithmetic properties of the curve which are present in such arithmetic objects as the set of all rational points. As a first approximation, their conjecture suggested that the existence of rational points on a cubic curve could be detected by knowledge of the over all local behavior of the curve with respect to all the primes. The proposed formula was to have a structure similar to that of the residue of the Dedekind zeta function of a number field at $s = 1$. In particular, there would be a regulator associated with the rational points of infinite order, there would be a number in the denominator indicating the size of the torsion present and above all, a number which like the class number would serve to measure the extent to which locally trivial classes fail to be globally trivial. Before stating the conjecture precisely, we must introduce several definitions first.

3.1 The Hasse-Weil Zeta Function. We consider an elliptic curve defined over the field of rational numbers by a Weierstrass minimal model

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad a_i \in \mathbb{Z}, \quad i = 1, \dots, 6.$$

We suppose that the conductor of E is N. Recall that the latter contains information about the algebraic structure of the reduction modulo p of E for all primes p. In particular, if the prime p divides N and the reduced curve

$$\tilde{E} : y^2 + \tilde{a}_1xy + \tilde{a}_3y \equiv x^3 + \tilde{a}_2x^2 + \tilde{a}_4x + \tilde{a}_6 \pmod{p}$$

has as a singularity an ordinary double point, then we can associate to E and p a number

$$a_E(p) = \pm 1,$$

depending on whether the slopes of the tangent lines are defined over $\mathbb{F}_p$ or over a quadratic extension. If the singularity is not ordinary, then we define

$$a_E(p) = 0.$$

When the prime p does not divide the conductor N, in which case the reduced curve $\tilde{E}$ has no singularities, we put

$$a_E(p) = p + 1 - \text{ Card } \tilde{E}(\mathbb{F}_p),$$

where the count of the number of $\mathbb{F}_p$-rational points on $\tilde{E}$ includes the point at infinity.

Definition. *With notations as above, the Hasse-Weil zeta function of the elliptic curve E is defined by the Euler product*

$$L(s,E) := (2\pi)^{-s}\Gamma(s)\prod_{q|N}\frac{1}{1-a_E(q)q^{-s}}\prod_{p\nmid N}\frac{1}{1-a_E(p)p^{-s}+p^{1-2s}},$$

where s is a complex variable with real part $> 3/2$.

From the point of view of analytic number theory, it is important to know if the L-function defined above, which can also be expressed formally as a Dirichlet series

$$L(s,E) = (2\pi)^{-s}\Gamma(s)\sum_{n=1}^{\infty}\frac{a_E(n)}{n^s},$$

has a meromorphic continuation to the whole complex s-plane. If this is true, then it makes sense to consider its value at $s = 1$. This is what the Birch and Swinnerton-Dyer conjecture attempts to clarify. Before we explain this connection we give two examples, which are fairly typical.

Example 1. We consider the curve of conductor $N = 11$ defined by the equation

$$E : y^2 - y = x^3 - x^2.$$

A theorem of Eichler and Shimura (see [42], [50]), which generalizes a congruence formula of Kronecker implies that the coefficients $a_E(n)$ which appear in the Dirichlet series representing the L-function $L(s,E)$ are identical with those obtained by formally expanding the product

$$q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 = \sum_{n=1}^{\infty}a_E(n).$$

It is easy to see in this case a hint of how the local to global principle works. In fact the last product provides a tool to prove that for all primes p we have

$$a_E(p) \equiv p + 1 \pmod 5,$$

while it is equally easy to verify that the elliptic curve E has a rational point of order 5, namely $P = (0,0)$. In the following we shall consider q as a local uniformizing parameter at infinity for the Riemann sphere and put

$$q := e^{2\pi i z},$$

where the variable z is supposed to take complex values with positive imaginary part. For notational convenience we we write

$$f_E(z) := \sum_{n=1}^{\infty} a_E(n) q^n.$$

Example 2. We consider the curve of conductor $N = 27$ defined by the equation

$$E : y^2 - y = x^3 - 7,$$

which is isogenous to the Fermat cubic $x^3 + y^3 + z^3 = 0$. Gauss' Last Entry in his Diary essentially determines the coefficients $a_E(p)$, namely

$$a_E(p) = \begin{cases} 0, & \text{if } p \equiv -1 \pmod 3 \\ t_p, & \text{if } p \equiv 1 \pmod 3), \end{cases}$$

where t_p is the unique number which appears in the representation $4p = t_p^2 + 27B^2$ for some integer B, and with t_p normalized so that it is $\equiv -1 \pmod 3$). A theorem of Hecke, generalized by Shimura also establishes that in this case the coefficients in the Dirichlet series representing the L-function $L(s,E)$ can also be obtained by formally expanding the product

$$q \prod_{n=1}^{\infty} (1 - q^{3n})^2 (1 - q^{9n})^2 = \sum_{n=1}^{\infty} a_E(n) q^n.$$

Essential to the demonstration in this case are the properties of the corresponding q-series

$$f_e(z) := \sum_{n=1}^{\infty} a_E(n) q^n$$

when considered as a function of the variable z.

In both examples given above, the function $f_E(z)$ has a very special property, namely it is a cusp form of weight 2 and Neben-typus (in the sense of Hecke) of level N. More precisely this means that as a function of the variable z in the upper half plane, it satisfies the functional relation

$$f_E\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f_E(z),$$

for all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with integer entries, determinant 1 and with $c \equiv 0 \pmod{N}$. We denote the totality of such matrices by $\Gamma_0(N)$. The suggestion that these two examples are not isolated cases is the content of the very important

Shimura-Taniyama-Weil Conjecture. *If E is an elliptic curve defined by a minimal Weierstrass model over the rational field with conductor N, then its L-function $L(s,E) = (2\pi)^{-s}\Gamma(s)\sum_{n=1}^{\infty} a_E(n)q^n$ is automorphic of level N, i.e., the associated function*

$$f_E(z) = \sum_{n=1}^{\infty} a_E(n)q^n$$

is a cusp form of weight 2 on the modular group $\Gamma_0(N)$.

Remark. The first example given above corresponds to an elliptic curve of smallest possible conductor. It also turns out that the corresponding function $f_E(z)$ is the unique cusp form of level $N = 11$ and weight 2, that is to say, the dimension of the space of cusp forms of weight 2 for the modular group $\Gamma_0(11)$ is one. The fact that this latter property implies those stated above about the L-function $L(s,E)$ results from the Eichler-Shimura congruence relation which ascertains that the Frobenius endomorphism in characteristic p has a canonical lifting to characteristic zero whose representation on the space of cusp forms of weight two coincides with the so called Hecke operator T_p. Exactly the same argument works for all those $\Gamma_0(N)$ for which the space of cusp forms of weight 2 is of dimension 1. This procedure produces only a finite number of elliptic curves whose L-functions are automorphic. The second example corresponds to an elliptic curve that has complex multiplication by the ring of Eisenstein integers $\mathbb{Z}[\rho]$, $\rho^2+\rho+1=0$. (See below for the definition.) In this case a theorem of Deuring shows that the L-function $L(s,E)$ is one of those considered by Hecke with Grossencharacter and therefore has meromorphic continuation. Shimura also proved in this case that if the elliptic curve E is defined over $\mathbb{Q}$ and has complex multiplication by an order in an imaginary quadratic field, then its L-function is automorphic (see Shimura [51]). His idea is to start with the L-function of the elliptic curve with complex multiplications, which is this case happen to be a Hecke L-function with Grossen-character, and show by a slight modification of the Hecke-Weil Converse theorem (see [69]) that the corresponding function $f_E(z)$ is in fact a cusp form whose period lattice generates an elliptic curve which happens to be isogenous to the original elliptic curve. Even though Shimura's argument was more general, the special case needed for the elliptic curves defined over $\mathbb{Q}$ required that the field of the complex multiplications be of class number 1, a fact that greatly limited the supply of elliptic curves whose L-functions were fully well known.

To state the Birch and Swinnerton-Dyer conjecture we make several definitions concerning an elliptic curve. If $\tilde{E}$ denotes the reduction modulo p of a Neron minimal model for the elliptic curve E we put

$$E_0(\mathbb{Q}_p) := \{P \in E(\mathbb{Q}) \mid \tilde{P} \in \tilde{E}_{ns}(\mathbb{F}_p)\}$$
$$E_1(\mathbb{Q}_p) := \{P \in E(\mathbb{Q}_p) \mid \tilde{P} = \tilde{O}\}.$$

Here $\tilde{E}_{ns}$ denotes the non-singular part of $\tilde{E}$, which is itself an algebraic group. An appropriate p-adic analogue of the Weierstrass $\wp$-function shows that the index of the second group with respect to the first is finite. This allows us to define for each prime p the integer

$$c_p := (E_1(\mathbb{Q}_p) : E_0(\mathbb{Q}_p)).$$

This number is none other than the number of components of multiplicity 1 rational over $\mathbb{F}_p$ on the special fiber of Neron's minimal model for E at p, and hence is equal to 1 for all except a finite number of p. If we let ω denote the differential form

$$\omega = \frac{dx}{2y + a_1x + a_3}$$

associated to a global minimal model for E, we put

$$\Omega := \int_{E(\mathbb{R})} | \omega | .$$

This number is either the positive real period of ω or twice that number depending on whether $E(\mathbb{R})$ has one or two connected components. If $P = (x, y)$ is a point on the elliptic curve E, Neron has constructed a canonical height denoted by $h(P)$, which differs from the naive height $h_0(P) := \log\mathrm{Max}(|m|, |n|)$, where $x = m/n$, by at most a bounded function on E. With this canonical height h one associates a biadditive pairing on $E(\mathbb{Q}) \times E(\mathbb{Q})$ given by

$$\langle P, Q\rangle := \frac{1}{2}(H(P + Q) - h(P) - h(Q)).$$

Let

$$\text{Ш} := \ker(H^1(\mathbb{Q}, E) \longrightarrow \oplus_p H^1(\mathbb{Q}_p, E))$$

be the group of classes of principal homogeneous spaces over E defined over the rational field which become trivial over every completion $\mathbb{Q}_p$. It is expected that this is a finite group. We are now ready to state the

Birch and Swinnerton-Dyer Conjecture: *With notations and definitions as above we have:*

(a) *The order of the zero of* $L(s,E)$ *at* $s=1$ *is equal to the rank* g *of the Mordell-Weil group* $E(\mathbb{Q})$.

(b) *Let* $P_1, P_2, \ldots, P_g$ *be* g *linearly independent points on* $E(\mathbb{Q})$ *and let* $B=\sum_{i=1}^{g}\mathbb{Z}P_i$ *be the subgroup of* $E(\mathbb{Q})$ *which they generate. Then*

$$\lim_{s\to 1}\frac{L(s,E)}{(s-1)^g}=\Omega[\text{Ш}]\frac{det\langle P_i,P_j\rangle}{(E(\mathbb{Q});B)^2}\prod_{p|N}c_p,$$

where $[\text{Ш}]$ *is the order of the Tate-Shafarevitch group of* E *over* $\mathbb{Q}$, *where* Ω *is the real period of* E *and* c_p *is the number of components in the reduction of the Neron minimal model modulo* p, *and* $\langle\ ,\ \rangle$ *is the height pairing.*

Elliptic Curves with Complex Multiplications. In this section we shall briefly review that part of the theory of elliptic curves with complex multiplications which is needed for a description of the work of Coates and Wiles on the Birch and Swinnerton-Dyer conjecture.

Throughout this section we assume that K is a complex quadratic number field and let $\mathcal{O}=\mathcal{O}_K$ be its ring of integers. We think of K as equipped with a fixed embedding $K\hookrightarrow\mathbb{C}$. Let $\mathcal{F}$ be an integral ideal in $K\neq\mathcal{O}$. Let $K(\mathcal{F})$ be the ray class field modulo $\mathcal{F}$. Recall that this is the unique abelian extension of K characterized by the property that the prime ideals $\wp$ in K which do not divide $\mathcal{F}$ and which split completely in $K(\mathcal{F})$ are precisely those which are principal, $\wp=(\alpha)$, with a generator α which satisfies $\alpha\equiv 1 \pmod{\mathcal{F}}$. We fix an elliptic curve E, defined over K, with complex multiplication by $\mathcal{O}$. We fix a minimal model

$$E:\quad y^2+a_1xy+a_3y=x^3+a_2x^2+a_4x+a_6,$$

for E defined over $\mathcal{O}_K$ and let $\Delta(E)$ be its discriminant. We recall that the above properties can be interpreted as saying that the ring of endomorphisms of E is isomorphic to $\mathcal{O}_K$, and that there exists a complex number Ω which generates the period lattice of the above model, *i.e.* we have an analytic isomorphism

$$\beta:\mathbb{C}/L\longrightarrow E(\mathbb{C}),$$

where $L=\Omega\cdot\mathcal{O}$, and

$$\beta(z):=(\wp(z)-(a_1^2+4a_2)/12,\wp'(z)/2-a_1\wp(z)/2+(a_1^3+4a_1a_2-12a_3)/24);$$

furthermore, the endomorphism corresponding to an element $\alpha\in\mathcal{O}_K$ is given by $\beta(z)\mapsto\beta(\alpha z)$. The minimality of the model guarantees that Δ is the "smallest" possible discriminant of an equation defining E with the property that E has good reduction at all prime divisors not dividing Δ.

Deuring's theorem concerning the L-function of an elliptic curve with complex multiplications can be described in the present situation as follows. If p

is a prime of good reduction for E, and if p splits in K, in which case E has good ordinary reduction, then the zeta function of E over the finite field $\mathbb{F}_p$ is given by

$$\frac{P_p(T)}{(1-T)(1-pT)},$$

where $P_p(T) = (1-\alpha_p T)(1-\hat{\alpha}_p)$ is a polynomial with integral coefficients and complex conjugate roots such that $\alpha_p\hat{\alpha}_p = p$. The roots α and $\hat{\alpha}$ are in $\mathcal{O}_K$ and generate the prime ideals in K deviding p: $\wp = (\alpha)$, $\hat{\wp} = (\hat{\alpha})$, $\wp\hat{\wp} = p\mathcal{O}$. We define $\psi(\wp) = \alpha_p$ and $\psi(\hat{\wp}) = \hat{\alpha}_p$. If p is a prime of good reduction for E which remains prime in K, in which case we know that E has supersingular reduction, then the numerator of the zeta function has the form $P_p(T) = 1 + pT^2$ and define $\psi(\wp) = -p$ for $\wp = p\mathcal{O}$. We also know that because E has complex multiplications in K, the reduction of E at a prime which divides the discriminant is of additive type and this implies that the numerator of its zeta function is $P_p(T) = 1$. For such primes we put $\psi(\wp) = 0$ for any prime ideal $\wp$ dividing such a p. By multiplicativity we extend the definition of $\psi(I)$ to all ideals I in $\mathcal{O}_K$. ψ is a Hecke character of type A_0 for K of infinity type (1,0). We denote the conductor of ψ by $\mathcal{F}$; if we assume that the class number of K is 1, then we can write $\mathcal{F} = (f)$ for some number f in K. In the following we fix the choice of such a generator f. During's theorem states that the Hasse-Weil zeta function $L(s, E/\mathbb{Q})$ coincides with the Hecke L-function associated to the character ψ:

$$L(s,\psi) := \begin{cases} (2\pi)^{-s}\Gamma(s)\prod_{\wp}\frac{1}{1-\psi(\wp)N\wp^{-s}} \\ (2\pi)^{-s}\Gamma(s)\sum_{I}\frac{1}{NI^{s}}. \end{cases}$$

Remark. Shimura's proof that the function $L(s, E/\mathbb{Q})$ is automorphic consists in examining the functional equations of $L(s,\psi)$ and

$$L(s,\psi,\chi) := (2\pi)^{-s}\Gamma(s)\sum_{I}\psi(I)\chi(N_{K/\mathbb{Q}}(I))N_{K/\mathbb{Q}}(I)^{-s}$$

with primitive characters χ of $(\mathbb{Z}/p\mathbb{Z})^{\times}$ for all rational primes p not deviding the level $M = D_K N_{K/\mathbb{Q}}(\mathcal{F})$, where D_K is the discriminant of K, and applying Weil's generalization of the converse theorem of Hecke (see [69], p. 203).

The main result of Coates and Wiles is the following.

Theorem. (Coates-Wiles) *Suppose that the elliptic curve E has complex multiplications by the ring of integers of an imaginary quadratic field K with class number 1, and that E is defined over K or $\mathbb{Q}$. If F is the field K or $\mathbb{Q}$ and the Mordell-Weil group $E(F)$ has points of infinite order, then the Hasse-Weil zeta function $L(s, E/F)$ vanishes at $s = 1$.*

As for the proof, we note that the existence of a rational point P in $E(\mathbb{Q})$ implies that the fields obtained by adjoining a point Q in $E(\bar{K})$ such that $\pi^{n+1}Q = P$ to the field $F_n = K(E_{\pi^{n+1}})$ provide an extension $F_n(Q)/F_n$ which remains unramified outside $\wp_n$, the unique prime ideal in F_n above $\wp$. By a descent argument modeled on Iwasawa's theory of p-extensions it is then shown that this could only happen if the component $(U_0'/\bar{C}_0(\chi)) \neq \{0\}$. The criterion recalled above then implies that the value

$$(2\pi\Omega)^{-1}L_{\mathcal{F}}(1,\bar{\psi}) \equiv 0 \pmod{\wp}.$$

By varying the prime p subject to the conditions imposed for the validity of the Kummer-Logarithm formula, and noting that $L_{\mathcal{F}}(1,\bar{\psi}) = L(1,\psi) = L(1,E/F)$, Coates and Wiles deduce that the last congruence holds for an infinite number of primes $\wp$, a situation which can only occur if the actual value $L(1,E/F) = 0$.

The requirement that the class number of the field K be 1 was subsequently removed by Arthaud (see [22]), who also showed that the field F could be taken to be an abelian extension of K satisfying a mild restriction.

Increased understanding of the structure of the Birch and Swinnerton-Dyer conjecture came from two other directions. On the one hand, Gross and Zagier demonstrated how to use Heegner points to evaluate the derivative of the Hasse-Weil zeta function at $s = 1$ and on the other hand the Work of Rubin and Kolyvagin established for the first time that for a large class of elliptic curves including those with complex multiplications, the Tate-Shafarevitch group Ш is finite. The latter work depended on the results of many people, but in one essential way it was motivated by Thaine's discovery of a new way of generating units in cyclotomic units while controlling the action of Galois, an idea which he attributes to his close reading of certain paper of Kummer.

4. The Herbrand-Ribet Theorem (Construction of Unramified Extensions)

The logarithmic and exponential functions play a fundamental role in the isomorphism between the multiplicative group of positive real numbers and the additive group. One of Kummer's most important contributions to the solution of Fermat's Last Theorem was his discovery that certain aspects of the problem created by the lack of unique factorization in the ring $\mathbb{Z}(\sqrt[p]{1})$, a situation which is inherently of a multiplicative nature, can be linearized by the introduction of suitable logarithms. One is faced here with a manifestation of one of the most rudimentary parts of class field theory as seen from the optic of the representation theory of groups in the simplest case, *i.e.* that of the algebraic group $GL(1)$ (see [71], vol. III, p. 301).

Kummer developed his ideas in the 1857 memoir in which he proved his irregularity criterion. The main idea behind his argument was based on the interaction between two seemingly unrelated events: on the one hand he supposes that in so far as p-th powers is concerned, the ideals in $\mathbb{Q}(\cos(\frac{2\pi}{p}))$ behave as if they were principal, and on the other hand, the divisibility of some Bernoulli number would imply that some fundamental units in the ring $\mathbb{Z}[\cos(\frac{2\pi}{p})]$ cannot be p-th powers. Looking in retrospect, Kummer was constructing for the first time unramified extensions of the cyclotomic field $\mathbb{Q}(\sqrt[p]{1})$ by extracting the p-th roots of certain units. We shall describe presently how these ideas were transformed in the hands of Herbrand and Ribet, leading up to the work of Wiles which involves the representation theory of the linear algebraic group $GL(2)$.

We let $\zeta = \sqrt[p]{1}$ be a fixed p-th root of unity, $\zeta \neq 1$. Let $\mathcal{C}$ be the p-th Sylow subgroup of the ideal class group of the ring of integers of $\mathbb{Q}(\zeta)$. Let $\mathbb{Z}_p$ be the ring of p-adic integers and $\mathbb{Z}_p^\times$ its group of units. Since $\mathrm{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q})$ is cyclic of order $p-1$, there is a canonical cyclotomic character

$$\chi : \mathrm{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q}) \longrightarrow GL_1(\mathbb{Z}_p),$$

defined by $g \cdot \zeta = \zeta^{\chi(g)}$, $g \in \mathrm{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q})$. Since the Galois group acts as a group of automorphisms of $\mathcal{C}$, and since the latter has a natural structure of $\mathbb{Z}_p$-module, we have a decomposition of $\mathcal{C}$ under the action of $\mathrm{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q})$ into eigenspaces

$$\mathcal{C} = \oplus_{k \pmod{p-1}} \mathcal{C}(\chi^{1-k}),$$

where

$$\mathcal{C}(\chi^{1-k}) := \{c \in \mathcal{C} : g \cdot c = \chi(g)^{1-k} c, \quad \text{for } g \in \mathrm{Gal}\,(\mathbb{Q}(\zeta)/\mathbb{Q})\}.$$

Recall that the k-th Bernoulli number is defined by the series expansion

$$\frac{x}{e^x - 1} := \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}.$$

Actually for Kummer's work it is more convenient to define the Bernoulli numbers B_k as those rational coefficients which appear in the expansion

$$\log\left(\frac{\varepsilon(e^u)}{\varepsilon(1)}\right) = \sum_{k>1} (g^k - 1)\frac{B_k}{k} \cdot \frac{u^k}{k!},$$

where g is a fixed positive primitive root modulo p, and $\varepsilon(x)$ is the algebraic function defined by

$$\varepsilon(x) = \sum_{|i| \leq \frac{g-1}{2}} x^i,$$

whose values for x a primitive p-th root of unity are the so-called cyclotomic units. Kummer's great insight was that divisibility properties of the Bernoulli numbers in the range $1 < 2\nu < p-1$ manifested themselves in the existence of fundamental units which were not p-th powers and therefore could be used to construct unramified extensions of $\mathbb{Q}(\zeta)$, which would in turn imply that some component $\mathcal{C}(\chi^{1-k})$ in the eigenspace decomposition of the p-Sylow subgroup of the class group would necessarily have to be non-trivial.

In the wake of new conceptual advances in class field theory, and following the work of Tagaki, Herbrand (1931) was able to simplify the demonstrations of Kummer's results from his 1857 memoir and at the same time obtain more precise statements. Herbrand was able to prove that if $\mathcal{C}(\chi^{1-k}) \neq \{0\}$ for an even k in the range $1 < k < p-1$, then the numerator of the Bernoulli number B_k would be divisible by p. Herbrand could also prove the converse on the assumption that the class number of the field $\mathbb{Q}(\cos(\frac{2\pi}{p}))$ was not divisible by p. Ribet (1976) was then able to show that in fact the latter improved assumption could be dispensed with. The Herbrand-Ribet theorem would then say that for even k in the range $1 < k < p-1$

$$\mathcal{C}(\chi^{1-k}) \neq \{0\} \quad \text{if and only if} \quad p \mid B_k.$$

Ribet's contribution was a significant one because it brought into the picture the theory of modular and pointed out the importance of constructing unramified extensions of $\mathbb{Q}(\zeta)$ by adjoining certain torsion points in the Jacobian variety of the modular curves. In a subsequent chapter we analyse in detail Ribet's steps leading up to his construction of unramified extensions. For the time being suffices to say that the Herbrand-Ribet result can be stated in the following suggestive way which hints at the role played by the theory of modular forms.

Recall that the Eisenstein series of weight k and level 1 are defined by

$$E_k := \frac{B_k}{2k} + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n,$$

where $\sigma_{k-1}(n)$ is the sum of the $(k-1)$-th divisors of n. The Herbrand-Ribet theorem can then be stated in the form: for even k in the range $1 < k < p-1$ we have

$$\mathcal{C}(\chi^{1-k}) \neq \{0\} \quad \text{if and only if} \quad E_k \in \mathbb{Z}_p[[q]]^{\times},$$

that is to say E_k is a unit in the ring of formal power series with coefficients in $\mathbb{Z}_p$ in the indeterminate q.

To state the results of Wiles (see [17]), we first recall the definition of generalized Bernoulli number. Let ψ be a $\mathbb{Z}$-valued character and put

$$B_{1,\psi} := \frac{1}{p} \sum_{a=1}^{p-1} \psi(a) a.$$

It had been conjectured for some time that for any odd i, $1 < i < p-1$ the order of the component $\mathcal{C}(\chi^i)$ was given by p^n, where $n = ord_p(B_{1,\chi^{-i}})$. It was also known that if the even part of $\mathcal{C}$ was trivial, *i.e.* $\oplus_{i \text{ even}} \mathcal{C}(\chi^i) = \{0\}$, then Stickelberger's theorem and the analytic class number formula would imply the conjecture. Wiles, building up on the ideas of Ribet [48] and of Mazur [63] ,was able to prove the following conditional result:

Theorem. *For i an odd integer in the range $1 < i < p-1$, if $\mathcal{C}(\chi^i)$ is cyclic, then its order is*

$$Card\,\mathcal{C}(\chi^i) = p^n,$$

where $n = ord_p(B_{1,\chi^{-i}})$.

Remark. In joint work with B. Mazur (see [18]), Wiles subsequently removed the assumption of cyclicity in the above theorem.

It is not difficult to imagine, although the actual details as presented in (Wiles [17]) are quite formidable, that as suggested by the statement of the Herbrand-Ribet theorem given above, the group of modular units generated by the Eisenstein series E_k in the closure of the ring $\mathbb{Z}_p[[q]]$ inside the function field of the modular curve $X_1(p)$ which parametrizes elliptic curves together with points of order p would have some influence in the structure of the p-primary part of the ideal class group of $\mathbb{Q}(\zeta)$. The approach taken by Wiles uses the Kubert-Lang characterization of the p-Sylow subgroup of the Manin-Drienfeld group inside the jacobian of $X_1(p)_{/\mathbb{Q}}$ generated by the cusps over the point at infinity. The end result is the construction of a particular representation

$$\rho : \mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{Z}/p^n\mathbb{Z}),$$

which satisfies the following properties:

(i) *ρ is unramified outside p,*

(ii) $\rho \equiv \begin{pmatrix} \chi^i & * \\ 0 & \varepsilon \end{pmatrix}$ (mod p^n), *where ε : $Gal(\bar{\mathbb{Q}}/\mathbb{Q})) \longrightarrow GL_1(\mathbb{Z}_p)$ is the cyclotomic character,*

(iii) ρ (mod p) *is not diagonalizable, i.e. its image contains a subgroup of order p^n, and*

(iv) *the restriction of ρ to $Gal(\mathbb{Q}(\zeta_{p^n})/\mathbb{Q})$ is unramified at p, where ζ_{p^n} is a primitive p^n-th root of unity.*

The integer n that appears above is defined as $n = \mathrm{ord}_p(B_{1,\chi^{-1-i}})$, where

$$B_{1,\chi^{-1-i}} := p \sum_{a\in(\mathbb{Z}/p\mathbb{Z})^\times/\pm 1} B_2\Big(\Big\langle \frac{a}{p} \Big\rangle\Big)\chi(a)^{-1-i};$$

is the generalized Bernoulli number of order 2, and

$$B_2(x) = x^2 - x + \frac{1}{6}$$

is the second Bernoulli polynomial.

The results of Wiles described in this section concerning the structure of the odd part of the p-Sylow subgroup of the ideal class group $\mathcal{C}$, together with those of Mazur and Wiles, have made it clear that in order to understand some of the questions suggested by Fermat's Last Theorem, and the related questions of class field theory, one must also consider certain aspects of the representation theory of the algebraic group $GL(2)$, a situation which is decidedly non-abelian. These themes will reappear in the last section of this chapter in a more decisive way, when we describe how another fundamental contribution by Ribet, completing an insight by Serre which itself was suggested by work of Frey, provided the first hints that the solution of Fermat's Last Theorem via the Shimura-Taniyama-Weil conjecture was consistent with the expected over-all structure of the theory of modular forms.

Part III: Technical Aspects

5. An Outline of Wiles' Proof

This section outlines the main technical steps, starting with Frey's observations, following up with Ribet's theorem, and leading up to Wiles' proof of the Shimura-Taniyama-Weil conjecture for semi-stable elliptic curves.

5.1 The Frey-Hellegouarch Connection. Starting with a remark of Helleguarch, Frey was lead to consider certain elliptic curves whose 2-division points were related to possible solutions of Fermat's equation. More precisely, he considers non-zero integers A, B, C which are relatively prime in pairs and satisfy

$$A + B + C = 0.$$

Three numbers x_1, x_2, x_3 are chosen so that

$$x_1 - x_2 = A, \quad x_2 - x_3 = B, \quad x_3 - x_1 = C.$$

The elliptic curve is easily seen to be independent of the choice of the x_i. For applications to Fermat's equation the choice of interest is $x_1 = A, \quad x_2 = 0, \quad x_3 = -B$. The corresponding elliptic curve is defined by the equation

$$E_{A,B,C} : y^2 = x(x - A)(x + B).$$

Note that a permutation σ of $\{A, B, C\}$ changes the curve $E_{A,B,C}$ at most by a twist over the field $\mathbb{Q}(\sqrt[2]{\operatorname{sgn}(\sigma)})$. Frey obtained the following description of the bad reduction of E:

Bad reduction at $\ell \neq 2$. Let ℓ be an odd prime. It is obvious that if ℓ divides A or B, then $x \equiv 0$ is a double zero of the cubic polynomial $x(x-A)(x+B)$ over the finite field $\mathbb{F}_\ell$; similarly if ℓ divides C, in which case $A \equiv -B \pmod{\ell}$, the cubic polynomial will also have a double zero. Therefore the odd primes of bad reduction are those which divide the product ABC. Furthermore, in this case the singular point is a double point with distinct tangents and the reduction is what one calls of multiplicative type. Following Tate's algorithm it is not difficult to verify that the equation defining $E_{A,B,C}$ is already a Weierstrass minimal model (see Tate [44]).

Bad Reduction at $\ell = 2$. For the case that arises from applications to Fermat's equation, we make the following assumption:

$$A \equiv -1 \pmod 4 \quad \text{and} \quad B \equiv 0 \pmod{32}.$$

Carrying out the substitution $x = 4X$, $y = 8Y + 4X$, produces for E a new equation with integer coefficients:

$$Y^2 + XY = X^3 + \frac{B - A - 1}{4}X^2 - \frac{AB}{16}X.$$

Over the binary field $\mathbb{F}_2$ it reduces to one of the two equations

$$Y^2 + XY = X^3, \quad Y^2 + XY = X^3 + X^2$$

depending on whether $A \equiv 7 \pmod 8$ or $A \equiv 3 \pmod 8$. In both cases the singular point $(0,0)$ is a double point whose tangents are rational over $\mathbb{F}_2$ in the first case and defined over a quadratic extension of $\mathbb{F}_2$ in the second case.

The corresponding Weierstrass equation is $\eta^2 = \xi^3 - (c_4/48)\xi - (c_6/864)$, where the associated invariants are

$$c_4 = -(AB + AC + BC)$$
$$c_6 = \frac{(B - A)(C - B)(A - C)}{2}$$
$$\Delta = \left(\frac{ABC}{16}\right)^2.$$

The fact that c_4 and Δ do not have any factor in common, implies that the last equation above is a minimal model for E and that E is a semi-stable elliptic curve. *i.e.*, E has good reduction at all primes ℓ which do not divide ABC, and bad reduction of multiplicative type at all the prime divisors of ABC (see Tate [44]). The primes of bad reduction for a semi-stable elliptic curve occur in the conductor only to the first power, and therefore the conductor of $E_{A,B,C}$ is simply

$$N = \operatorname{rad}\frac{ABC}{16},$$

where $\operatorname{rad}(X)$ denotes the product of the prime numbers that divide X. (See Weil [69], p. 156 (Case (II) with $a = 1$). The modular invariant $j = j_E$ of the elliptic curve $E = E_{A,B,C}$ is

$$j_E = 2^8(C^2 - AB)^3/A^2B^2C^2.$$

If ℓ divides ABC, we see immediately that

$$\operatorname{ord}_\ell(j_E) = -\operatorname{ord}_\ell(\Delta) = \begin{cases} -2\operatorname{ord}_\ell(ABC) & \text{if } \ell \neq 2 \\ 8 - 2\operatorname{ord}_\ell(ABC) & \text{if } \ell = 2. \end{cases}$$

Remark. If in the above calculations we do not assume that the triple (A, B, C) or any of its cyclic permutations satisfy $A \equiv -1 \pmod 4$, and $B \equiv 0 \pmod{16}$, then the corresponding elliptic curve $E_{A,B,C}$ may still be minimal but its reduction at the prime $\ell = 2$ will not be semi-stable. Many such examples arise from Pythagorean triples. For instance the elliptic curve $y^2 = x(x + 3^2)(x - 4^2)$ is semistable, and in fact isomorphic to the modular curve $X_0(15)$ which parametrizes elliptic curves with a cyclic group of order 15, while the elliptic curve $y^2 = x(x + 4^2)(x - 3^2)$ is not semi-stable, *i.e.*, modulo $\ell = 2$ the curve has a cusp.

For the application to Fermat's equation we take an exponent $p \geq 5$ and a triple of non-zero integers a, b, c which satisfy the equation

$$a^p + b^p + c^p = 0,$$

and from it we manufacture a Frey elliptic curve with $A = a^p, B = b^p$:

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

where the a, b, c are ordered so that b is even and a is congruent to $-1 \pmod 4$. As we have indicated above, the discriminant of E is $\frac{(abc)^{2p}}{256}$ and its conductor is the product of the primes dividing abc.

The insight provided by Serre's conjecture ([20], 3.3.1?) [‡] can be interpreted as trying to reconcile two opposite events: on the one hand the assumption that the exponent $p \geq 5$ implies by Mazur's theory of the Eisenstein ideal that the mod p representation resulting from the action of Galois on the field of p-division points is as large as it can possibly be (the full group of automorphisms of $E[p]$ is isomorphic to $GL_2(\mathbb{F}_p)$) and hence highly non-abelian. On the other hand, the miraculous shape of the modular invariant j_E, which is almost a p-th power, implies that the image of inertia I is semi-simple (see Serre [8], p. 277, Corollaire, case b.); this in turn has the tendency to pull the representation toward the abelian side, that is to say, the field of p-division points has a structural similarity with the unramified p-extensions of the cyclotomic field $\mathbb{Q}(\sqrt[p]{1})$. To actually carry out this comparison and obtain a contradiction requires the input of two major results: The Ribet-Mazur Theorem, and Wiles' proof of the Shimura-Taniyama-Weil conjecture for semi-stable elliptic curves. In the following sections we describe how these results enter the picture.

5.2 The Shimura-Taniyama-Weil Conjecture. In section 3.1 we have already described the Shimura-Taniyama-Weil conjecture. We want to add here a few remarks concerning the significance of Wiles' most fundamental theorem.

[‡] A situation which to us is reminiscent of the *"demostrationen mirabilem"* as in *Observatio Domini Petri de Fermat.*

Theorem. (A. Wiles) *If E is a semi-stable elliptic curve defined over $\mathbb{Q}$ and of conductor N, then its Hasse-Weil zeta function*

$$L(s,\pi_E) = (2\pi)^{-s}\Gamma(s)\prod_{q|N}\frac{1}{1-a(q)q^{-s}}\prod_{p\nmid N}\frac{1}{1-a(p)p^{-s}+p^{1+2s}},$$

is automorphic.

Recall that the coefficients $a_E(p)$ are defined as follows: if p does not divide the conductor, then

$$\operatorname{Card} E(\mathbb{F}_p) = p+1-a_E(p),$$

where $E(\mathbb{F}_p)$ is the set of points on E defined over the finite field $\mathbb{F}_p$, including the point at infinity (∞,∞). For a prime q which divides the conductor N, we put $a_E(q)=1$ if the double point has two tangents which are defined over $\mathbb{F}_p$, otherwise we set $a_E(q)=-1$ if the tangents are defined over a quadratic extension of $\mathbb{F}_p$. (See Weil [69] p. 171). If we expand the Euler products that appear in the definition of $L(s,\pi_E)$ we obtain a Dirichlet series

$$L(s,\pi_E) = (2\pi)^{-s}\Gamma(s)\sum_{n=1}^{\infty}\frac{a_E(n)}{n^s}.$$

Note that the automorphic property of the L-function $L(s,\pi_E)$ is equivalent to the union of the following conditions:

- *As a function of the complex variable s, $L(s,\pi_E)$ represents an entire function, which is bounded in vertical strips and satisfies the functional equation*
$$L(s,\pi_E) = W_E N^{1-s} L(2-s,\pi_E),$$
where $W_E=\pm 1$.
- *For all primitive characters χ of conductor m taken from a finite subset of the set $\{4,3,5,7,11,\dots\}$ which is bounded by a polynomial in N, each of the Dirichlet series*
$$L(s,\pi\otimes\chi) = (2\pi)^{-s}\Gamma(s)\sum_{n=1}^{\infty}\frac{a_E(n)\chi(n)}{n^{-s}},$$
initially defined only on a half-plane, has an analytic continuation to an entire function of the complex variable s, bounded in vertical strips and satisfying the functional equation
$$L(s,\pi_E\otimes\chi) = W_E(m^2N)^{1-s}\frac{g(\chi)}{g(\bar{\chi})}\chi(-N)L(2-s,\pi_E\otimes\chi^{-1}),$$
where $g(\chi)$ is the Gaussian sum defined by
$$g(\chi) = \sum_{x \bmod m}\chi(x)e^{2\pi i\frac{x}{m}}.$$

The above is essentially Weil's elaboration of Hecke's converse theorem. As already stated by Weil in his fundamental paper (see [69], p. 172), The above characterization together with important results of Shimura, implies not only that the differential

$$f_E(z)dq = \sum_{n=1}^{\infty} a_E(n)q^n$$

cuts an abelian subvariety on the Jacobian of the modular curve $X_0(N)$, but that it actually corresponds to an elliptic curve E' defined over $\mathbb{Q}$ which is isogenous to the original elliptic curve E.

To say that $L(s, E)$ is automorphic can also be interpreted as saying that the Mellin Transform

$$f_E(z) = \sum_{n=1}^{\infty} a_n q^n,\ q = e^{2\pi i z},$$

is a modular cusp form of weight 2 on the matrix group

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : c \equiv 0 \bmod N \right\}.$$

In other words $f_E(z)$ satisfies the functional equation for $\sigma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$,

$$f_E\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f_E(z).$$

It also guarantees that $f_E(z)$ is an eigen form for all the Hecke operators. In the following section we shall explain the meaning of this last condition.

5.3 The Hecke Ring T. The fundamental properties of the Hecke operators and the associated Hecke rings which are needed for the work of Ribet and Wiles depend on a careful study, first initiated by Mazur, of these operators from the point of view of endomorphisms of the modular scheme over $\mathbb{Z}$. Building up on the well known results of Deligne and Rapoport, and complementing these with fundamental results due to Raynaud and Drienfield, Wiles and Mazur have given a fairly complete exposition of the main properties of the Hecke operators in the §5 of chapter 2. of their paper on *Class fields of abelian extensions of* $\mathbb{Q}$ (See [18], p. 234). For the purpose of this introductory chapter we shall fall back on a coarse description of the Hecke operators which is perhaps more familiar from a classical point of view, as in Shimura's monograph *Introduction To The Arithmetic Theory of Automorphic Functions*, and suffices for most purposes. We shall follow closely the presentation given by Wiles in his paper [17], §2.

Let $\Gamma_1(N)$ be the group of matrices

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : a \equiv d \equiv 1 \pmod N, \quad c \equiv 0 \pmod N \right\}.$$

There is an affine curve $Y_1(N)_{/\mathbb{Q}}$ which is the coarse moduli scheme associated to the functor

"isomorphism classes of elliptic curves with a point of order N".

The curve $Y_1(N)$ can be compactified by taking the normalization of the projective j-line $\mathbf{P}^1_{/\mathbb{Q}}$ in the field of rational functions of $Y_1(N)$ corresponding to the covering induced by the usual assignment $(E, P) \mapsto j = j_E$. The resulting projective non-singular curve $X_1(N)_{/\mathbb{Q}}$ can be thought of as the coarse moduli scheme associated to the functor

> "isomorphism classes of generalized elliptic curves, together with a section everywhere of order N such that the subgroup scheme it generates meets every irreducible component of each geometric fibre".

In figurative terms which can be made quite precise what the above means is that to $Y_1(N)_{/\mathbb{Q}}$ one must add the cusps which in this case correspond to the orbits in the quotient $\mathbf{P}^1(\mathbb{Q})/\Gamma_1(N)$ and to each such cusp one associates a certain polygon which is to be thought of as a kind of generalized cubic curve of arithmetic genus 1, together with a subgroup of order N in the non-singular fiber which varies coherently as we run through all the cusps. Analogous to the above situation, we have a non-singular curve $X_0(N)_{/\mathbb{Q}}$ which is defined as above with the group $\Gamma_0(N)$ instead of $\Gamma_1(N)$; in particular the corresponding affine curve $Y_0(N)_{/\mathbb{Q}}$ can be thought of as the coarse moduli scheme associated to the functor

"isomorphism classes of elliptic curves with a cyclic subgroup of order N".

Associated to $X_0(N)_{/\mathbb{Q}}$ and $X_1(N)_{/\mathbb{Q}}$ there is a natural projection

$$\pi : X_1(N)_{/\mathbb{Q}} \longrightarrow X_0(N)_{/\mathbb{Q}}$$

which is obtained by mapping the pair (E, P) consisting of an elliptic curve E and a point P of order N into the pair (E, C_N), C_N is the cyclic group generated by P.

The Hecke operators can now be described as correspondences on $X_1(N)_{/\mathbb{Q}}$.

The Hecke operator T_ℓ for $\ell \nmid N$. Let ℓ be a prime which does not divide N and let $X_1(N;\ell))_{/\mathbb{Q}}$ be the curve constructed in the same way as $X_1(N)_{/\mathbb{Q}}$ but with the group $\Gamma_1(N)$ replaced by the subgroup $\Gamma_1(N) \cap \Gamma_0(\ell)$, so that $X_1(N;\ell)_{/\mathbb{Q}}$ can be interpreted as the coarse moduli scheme associated to the functor whose points are isomorphism classes of triples (E, P, C_ℓ), where E is a generalized elliptic curve, P is a point of order N, and C_ℓ is a subgroup of order ℓ. We have a projection map

$$\begin{tikzcd}
X_1(N;\ell)_{/\mathbb{Q}} \arrow[d, "\pi_1 \times \pi_2"] \\
X_1(N)_{/\mathbb{Q}} \times X_1(N)_{/\mathbb{Q}}
\end{tikzcd}$$

where the maps π_1, π_2 from $X_1(N;\ell)_{/\mathbb{Q}}$ to $X_1(N)_{/\mathbb{Q}}$ are given on points by

$$\pi_1 : (E, P, C_\ell) \mapsto (E, P) \qquad \pi_2 : (E, P, C_\ell) \mapsto (E/C_\ell, (P + C_\ell)/C_\ell).$$

These projections give rise to a curve on the surface $X_1(N)_{/\mathbb{Q}} \times X_1(N)_{/\mathbb{Q}}$ which is to be thought of as a correspondence and which we denote by T_ℓ. As a correspondence which induces an endomorphism on the Jacobian variety of $X_1(N)_{/\mathbb{Q}}$ we can also describe T_ℓ in terms of its action on the "points" as follows:

For a prime ℓ which does not divide N, we set

$$T_\ell : (E, P) \mapsto \sum_B (E/B, (P + B)/B),$$

where B runs through all the subgroups of order ℓ in E.

The Hecke operator T_p for $p \mid N$. For these a description can be given similar to that above, but now using the curve $X(p)_{/\mathbb{Q}}$ instead of $X_1(N, \ell)$. These are the so-called Atkin-Lehner operators, and on a point (E, P) defined over an algebraically closed field its action is

$$U_p : (E, P) \mapsto \sum_{C_p} (E/C_p, \text{image of } P \text{ in } E/C_p),$$

where C_p runs through all subgroups of order p in E which are not contained in the subgroup of E generated by P.

Remark. For questions related to reduction modulo p, it is important to treat separately the cases of a prime q which divides N and $q = p$ from that with $q \neq p$. For a thorough discussion of this technical matter the reader should consult the paper of Wiles-Mazur ([18], p. 237).

The Diamond operator $\langle a \rangle$. For each element $a \in (\mathbb{Z}/N\mathbb{Z})^\times/(\pm 1)$, which we identify with the Galois group of the totally real field $\mathbb{Q}(\zeta_p)^+$, we associate the automorphism of the curve given by

$$\langle a \rangle : (E, P) \mapsto (E, aP).$$

The Atkin Involution w_ζ. For each $\zeta \in \boldsymbol{\mu}_p$, the group of p-th roots of unity, we define an involution w_ζ of $X_1(N)_{/\mathbb{Q}(\zeta)^+}$. For a point on $X_1(N)_{/\mathbb{Q}}$ defined over $\mathbb{Q}(\zeta)$ we put

$$w_\zeta : (E, P) \mapsto (E/\langle P \rangle, P'),$$

where P' is the point on $E/\langle P \rangle$ which under the Weil pairing $\langle P \rangle \times E/\langle P \rangle \longrightarrow \boldsymbol{\mu}_p$ induced on the p-division points corresponds to the equation $(P, P') = \zeta$. One verifies directly that $w_\zeta = w_{\zeta^{-1}}$ is defined over $\mathbb{Q}(\zeta)^+$.

As correspondences on the Riemann surface $X_1(N)_{/\mathbb{C}}$, which we view as orbits of points z on the upper half plane H, the operators defined above have the following description:

We let σ_a be a matrix in $SL_2(\mathbb{Z})$ such that $\sigma_a \equiv \begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} \pmod{p}$. We then have

$$\begin{aligned} T_\ell &: z \mapsto \sigma_\ell(\ell z) + \sum_{b=0}^{\ell-1} \left(\frac{z+b}{\ell}\right), \\ U_p &: z \mapsto \sum_{b=0}^{p-1} \left(\frac{z+b}{p}\right), \\ \langle a \rangle &: z \mapsto \sigma_a(z), \\ w_\zeta &: z \mapsto -1/pz \quad \text{for } \zeta = e^{2\pi i}. \end{aligned}$$

One last operator that we define is

$$U'_p = w_\zeta^{-1} U_p w_\zeta.$$

By viewing the Jacobian variety $J_1(N)_{/\mathbb{Q}}$ of $X_1(N)_{/\mathbb{Q}}$ as the Albanese variety, *i.e.* as equivalence classes of formal sums of points, we can transport the action of all of the operators defined above, except for w_ζ, to the Jacobian under the map $f \mapsto f_*$ to yield endomorphisms of $J_1(N)_{/\mathbb{Q}}$. w_ζ also induces an endomorphism on $J_1(N)_{/\mathbb{Q}(\zeta)^+}$. Viewed as endomorphisms, the operators defined above satisfy the following relations:

- $\{T_\ell, U_p, U'_p, \langle a \rangle\}$ form a family of commuting operators for all $\ell \nmid N$, and for all $a \in (\mathbb{Z}/N\mathbf{z})^\times/(\pm 1)$.
- $w_\zeta \circ \langle m \rangle \circ w_\zeta = \langle m^{-1} \rangle$.
- $\langle m \rangle \circ w_\zeta = w_{\zeta^m}$.
- $w_\zeta \circ T_\ell \circ w_\zeta = T_\ell \circ \langle \ell^{-1} \rangle$.

We can also view the operators as acting on the space of weight 2 cusp forms on $\Gamma_1(N)$. This leads to the following simple description of the Hecke operators T_ℓ, U_p on the Fourier expansion of a cusp form $f = \sum_{n=1}^{\infty} a(n)q^n$ about the cusp at infinity:

$$T_\ell f = \sum_{n=1}^{\infty} a(\ell n)q^n + \ell \sum_{n=1}^{\infty} a(n)q^{\ell n},$$

$$U_p f = \sum_{n=1}^{\infty} a(np)q^n.$$

Since the Jacobian $J_0(N)_{/\mathbb{Q}}$ of the curve $X_0(N)_{/\mathbb{Q}}$ is a subvariety of the Jacobian $J_1(N)_{/\mathbb{Q}}$ of $X_1(N)_{/\mathbb{Q}}$, we can restrict the action of the endomorphism T_ℓ, T_p to endomorphisms of $J_0(N)_{/\mathbb{Q}}$. In the discussion that follows different types of Hecke rings will be used. The following one suffices for the discussion of the Ribet-Mazur theorem.

Definition. *Let N be a positive integer. Let $\mathbf{T} = \mathbf{T}_N$ be the ring generated by the Hecke operators $T_\ell, \ell \nmid N$ and the operators $U_p, p|N$ acting on the space of weight-2 cusp forms on $\Gamma_0(N)$.*

We shall now describe the correspondence that exists between the eigenfunctions of the Hecke operators and ℓ-adic representations of the absolute Galois group. Let $\underline{m}$ be a maximal ideal of the Hecke ring $\mathbf{T}$. Deligne and Serre have proved the following result:

There is a unique semisimple representation

$$\rho_{\underline{m}} : Gal(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbf{T}/\underline{m}),$$

satisfying

$$trace(\rho_{\underline{m}}(Frob_r)) \equiv T_r \pmod{\underline{m}}, \quad \det(\rho_{\underline{m}}(Frob_r)) \equiv r \pmod{\underline{m}}$$

for almost all primes r. The representation $\rho_{\underline{m}}$ is unramified at all primes r prime to $\underline{m}N$, and the displayed relation holds for all such primes.

In the following we say that $\rho_{\underline{m}}$ is the representation of Gal $(\bar{Q}/\mathbb{Q})$ attached to $\underline{m}$. We describe more precisely how one attaches a mod ℓ representation of Gal $(\bar{\mathbb{Q}}/\mathbb{Q})$ to mod ℓ modular forms which are eigenvectors for the Hecke operators T_n. Let $\mathcal{L}$ be the space of weight-2 cusp forms on $\Gamma_0(N)$ whose q-expansions about the cusp at infinity lie in $\mathbb{Z}[[q]]$. Both $\mathbf{T}$ and $\mathcal{L}$ are free $\mathbb{Z}$-modules with the same rank equal to the genus g of the curve $X_0(N)$. Let $\mathbb{F}$ be the residue class field $\mathbf{T}/\underline{m}$. The q-expansion map

$$\mathcal{L} \longrightarrow \mathbb{Z}[[q]]$$

induces an injective map

$$\mathcal{L} \otimes_{\mathbb{Z}} \mathbb{F} \longrightarrow \mathbb{F}[[q]].$$

It can be verified that the bilinear pairing

$$(\mathcal{L} \otimes_{\mathbb{Z}} \mathbb{F}) \times (\mathbf{T} \otimes_{\mathbb{Z}} \mathbb{F}) \longrightarrow \mathbb{F}$$

which sends the pair (f, T) to the coefficient of q in the q-expansion of $f|T$ induces an isomorphism of $\mathbb{F}$-vector spaces

$$\mathcal{L} \otimes_{\mathbb{Z}} \mathbb{F} \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(\mathbf{T}, \mathbb{F})$$

of dimension g.

Definition. *Suppose that*

$$\rho : Gal(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F})$$

is a continuous homomorphism, where $\mathbb{F}$ is a finite field. Let ℓ be the characteristic of $\mathbb{F}$. We say that the representation ρ is modular of level N if the *determinant of* ρ *is the* mod ℓ *cyclotomic character and if there is a homomorphism*

$$\omega : \mathbf{T} \longrightarrow \bar{\mathbb{F}}$$

such that

$$trace(\rho(Frob_r)) = \omega(T_r)$$

for almost all prime numbers r.

We recall that a representation Gal $(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F})$ is said to be *finite at p* if there is a finite flat $\mathbb{F}$-vector space scheme $\mathcal{V}$ over $\mathbb{Z}_p$ for which the resulting representation of the decomposition group $D_p =$ Gal $(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ coincides with the restriction of ρ to D_p. After these preparations we are ready to state the *Main Theorem* of Ribet.

Theorem. *Let*

$$\rho : Gal(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F})$$

be a mod ℓ *modular representation of level* Mp*, where* p *is not a prime dividing* M*. Assume that* ρ *is finite at* p*. Then* ρ *is modular of level* M *provided that at least one of the following two conditions holds:*

(*i*) *The prime* ℓ *is not a divisor of* M*;*

(*ii*) *We do not have* $p \equiv 1 \pmod{\ell}$*.*

Remark. The proof of the above theorem is given in Ribet [19]. It depends on delicate arithmetic properties of the Jacquet-Langlands correspondence between certain types of cusp forms on the modular curve $X_0(N)$ and cusp forms on Shimura curves. A brief outline of the proof can be found in Oesterle's Bourbaki Report [38]. On the basis of important results of Carayol, Diamond has given a significant improvement of the above result which allows greater flexibility on the primes that can divide the level (see [29]).

5.4 Shimura-Taniyama-Weil implies Fermat. In this subsection we describe how to use Ribet's theorem to deduce Fermat's Last Theorem from the Shimura-Taniyama-Weil conjecture. We assume that (a, b, c) is a triple of non-zero relatively prime integers which satisfy the Fermat equation

$$a^\ell + b^\ell + c^\ell = 0$$

with $\ell \geq 5$. If necessary, we may permute the (a, b, c) so that b is even and $a \equiv 3 \pmod 4$, and the Frey curve

$$E : y^2 = x(x - a^\ell)(x + b^\ell)$$

is semi-stable with bad reduction precisely at the primes that devide the product abc. Let N be the conductor of E, and note that 2 divides N. Because of Wiles' theorem concerning the Shimura-Taniyama-Weil conjecture for semi-stable elliptic curves over $\mathbb{Q}$, there is a cusp eigen form f of weight 2 on $\Gamma_0(N)$, with integral q-expansion coefficients, whose Mellin transform is the Hasse-Weil

L-function of E over $\mathbb{Q}$. The maximal ideal of the Hecke ring associated to f provides a mod ℓ representation ρ of Gal$(\bar{\mathbb{Q}}/\mathbb{Q})$ which is realizable over the $\mathbf{F}_\ell$-vector space

$$V = E[\ell]$$

consisting of the ℓ-division points on E. The assumption that $\ell \geq 5$ guarantees that Gal$\,\bar{\mathbb{Q}}/\mathbb{Q})$ acts irreducibly on $E[\ell]$, as is easily seen from Mazur's theory of the Eisenstein ideal, that is to say reducibility would imply the existence of a large torsion subgroup on $E(\mathbb{Q})$. Furthermore, since E is semistable, the representation is finite at every prime $p \neq 2$ which divides N. This is easily seen from the Tate parametrization. (Serre [8]).

Now the claim is that Ribet's Main Theorem implies that ρ is modular of level 2. To see this we argue as follows. Suppose that N is divisible by ℓ. Then ρ is finite and since $\ell \not\equiv 1 \pmod{\ell}$, ρ is seen to be modular of level N/ℓ. Taking $N_0 = N/\ell$ in this case, and putting $N_0 = N$ if ℓ does not divide N, we derive in all cases that ρ is modular of level N_0 which divides N and is prime to ℓ. The Main Theorem can now be applied inductively to eliminate all odd primes, and obtain that ρ is of level 2. But this is clearly impossible because the modular curve $X_0(2)$ is of genus 0 and hence there are no cusp forms of weight 2 and level 2.

5.5 Galois Representation on the Space of Inflection Points. Let $E/\mathbb{Q}$ be an elliptic curve, and let ℓ be an odd prime. We let $E[\ell]$ be the subgroup of $E(\mathbb{C})$ consisting of the ℓ-division points, *i.e.*, those points with $\ell \cdot P = \mathcal{O}$. As a vector space over $\mathbb{F}_\ell$, $E[\ell]$ is of dimension 2, and with a suitable choice of basis, we get a mod ℓ representation

$$\rho_{E,\ell} : \text{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F}_\ell).$$

The kernel of this representation fixes a finite extension of $\mathbb{Q}$ which we denote by $\mathbb{Q}(E[\ell])$. The latter notation is justified by noting that the coordinates of all the points in $E[\ell] - \mathcal{O}$ are algebraic numbers and their adjuntion to $\mathbb{Q}$ generate a finite extension. The existence of the Weil pairing shows that the determinant of $\rho_{E,\ell}$ correspond to the cyclotomic character

$$\varepsilon : \text{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_1(\mathbb{F}_\ell)$$

which gives the action of Gal$\,(\bar{\mathbb{Q}}/\mathbb{Q})$ on the ℓ-th roots of unity. In particular, $\varepsilon(\text{Frob}_p) \equiv p \pmod{\ell}$ for $p \neq \ell$. Furthermore, if c denotes a complex conjugation, then $\varepsilon(c) = -1$, that is to say $\rho_{E,\ell}$ is an odd representation.

We will be particularly interested in the case when $\rho_{E,\ell}$ is absolutely irreducible; this implies that its image is the full group $GL_2(\mathbb{F}_\ell)$ of automorphisms of $E[\ell]$. As indicated earlier, Serre ([20], p. 201) has shown that if E comes from a Frey curve associated to a rational solution of Fermat's equation, then $\rho_{E,\ell}$ is absolutely irreducible for $\ell \geq 5$. The case that plays a fundamental role in Wiles' argument is $\ell = 3$ which is not covered by Serre's result. To circumvent this difficulty in his original proof, Wiles introduced a deformation technique based on the well known diophantine properties of the icosahedral equation $X(5)_{/\mathbb{Q}}$ and on the Hilbert irreducibility theorem to show that if $\rho_{E,3}$ is not absolutely irreducible, then there is an elliptic curve $E'/\mathbb{Q}$, with isomorphic representations $E[5] \simeq E'[5]$ which has the properties that (i) *the Galois representation $\rho_{E',3}$ has full image*, (ii) *E' has semistable representation at 5.* After verifying that E' is semistable at all primes $p \neq 5$, Wiles applies his main theorem to the curve E'. Thus E' will be modular and so will be the mod 5 representation $\rho_{E',5}$. This implies that $\rho_{E,5}$ is modular and hence by another application of his theorem, Wiles proves the Shimura-Taniyama-Weil conjecture for E.

For applications to Fermat's Last Theorem, N. Elkies has given a more direct verification of the irreducibility condition that is needed to apply Wiles' theorem to the Frey curves. The major drawback of Elkies reasoning is that it only yields what is needed to prove Fermat's Last Theorem, but it cannot be used to verify the full Shimura-Taniyama-Weil conjecture for all semistable elliptic curves. The attractiveness of Elkies' idea is that it has the appearance of a simple excercise in Galois theory.

Fix a prime $p > 3$. The results of Frey, Serre and Ribet imply that if (A, B, C) are integers satisfying

(i) $A + B + C = 0$,

(ii) A, B, C each has the form $2^r 3^s m^p$ with either $r = 0$, or $r \geq 4$,

then the curve $E : y^2 = x(x + A)(x - B)$ cannot be modular. His argument uses the idea that if $H = \text{image}(\rho_{E,3})$ is properly contained in the octahedral group $GL_2(\mathbb{F}_3)$, then either

(*a*) H is a 2-Sylow subgroup or is contained in a 2-Sylow subgroup, or

(*b*) it acts reducibly on $E[3]$, and therefore E admits a rational 3-isogeny.

Elkies excludes the possibility (a) by noting that if the order of H is a power of 2, then the j-invariant of E is a rational cube and this would imply that the corresponding elliptic curve has Weierstrass model $y^2 = x^4 + 1$ of conductor 36 and admits complex multiplications and hence cannot be semistable. As for the possibility (b), he shows that if E admits a 3-isogeny, then it is possible to produce another elliptic curve of the Frey type for which $H = \text{image}(\rho_{E,3})$ is maximal and hence modular by Wiles' theorem.

On the assumption that the image of $\rho_{E,3}$ is the full group $GL_2(\mathbb{F}_3)$, one is lead to consider one of its two complex octahedral representations

$$\rho : \text{Gal}\,(\mathbb{Q}(E[3])/\mathbb{Q}) \longrightarrow GL_2(\mathbb{C}).$$

We recall that the octahedral group is the 2-covering of the group of symmetries of a regular octahedron and that it is generated by two elements t and r whose images in the group of rotations correspond respectively to a rotation by 120^o about the midpoints of two opposite faces and an element of order 8 whose image is a rotation by 45^o about the axis passing through two opposite vertices. We may in fact take

$$t = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad r = \begin{pmatrix} \sqrt{-2} & 1 \\ 1 & 0 \end{pmatrix}$$

The mapping from $GL_2(\mathbb{F}_3)$ to $GL_2(\mathbb{C})$ obtained by sending the generators to r and t

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

provides the desired embedding $GL_2(\mathbb{F}_3) \hookrightarrow GL_2(\mathbb{Z}[\sqrt{-2}])$.

The resulting 2-dimensional complex representation ρ has attached to it an Artin L-function

$$\begin{aligned} L(s,\rho) &= (2\pi)^{-s}\Gamma(s) \prod_{q|N_\rho} \frac{1}{1 - b_E(q)q^{-s}} \prod_{p\nmid N_\rho} \frac{1}{1 - b_E(p)p^{-s} + \varepsilon(p)p^{-2s}} \\ &= (2\pi)^{-s}\Gamma(s) \sum_{n=1}^{\infty} \frac{b_E(n)}{n^s}, \end{aligned}$$

where N_ρ is the conductor of ρ.

By using base change theory on $GL(2)$, Langlands and Tunnell have shown that the Mellin transform of $L(s,\rho)$

$$g_E(z) = \sum_{n=1}^{\infty} b_E(n) q^n, \quad q = e^{2\pi i z},$$

is a cusp form of weight 1 on $\Gamma_0(N_\rho)$ with nebentypus $\varepsilon(n) = \left(\frac{-3}{n}\right)$, the Kronecker symbol associated with the quadratic extension $\mathbb{Q}(\sqrt{-3})$. The latter extension has a unique Eisenstein series associated with it

$$E_\varepsilon(z) = 1 + 6 \sum_{n=1}^{\infty} \Big(\sum_{d|n} \varepsilon(d) \Big) q^n.$$

The product

$$f(z) = g_E E_\varepsilon,$$

is a cusp form of weight 2 on $\Gamma_0(N_\rho)$ which modulo 3 is an eigenform of all the Hecke operators. A modification of the Serre-Deligne type (see [52], Theorem 6.7) will produce a cusp eigen form $\tilde{f}$ whose coefficients can be thought as lying in the ring of integers $\mathcal{O}$ of a finite extension K of $\mathbb{Q}_3$, with maximal ideal λ. The fundamental property of the eigen form $\tilde{f}$ is that its associated ℓ-adic representation

$$\rho_{\tilde{f},\lambda} : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{Q}_3),$$

satisfies the congruence

$$\text{trace } \rho_{\tilde{f},\lambda}(\mathrm{Frob}_p) \equiv \text{trace } \rho_{E,3}(\mathrm{Frob}_p) \pmod{\lambda}.$$

In the following section we shall review how these congruences provide a means of classifying all modular eigen forms whose reductions modulo 3 lead to the same modular representation.

5.6 The Universal Modular Deformation. Let ℓ be an odd prime, let $\mathcal{O}$ denote the ring of integers of a finite extension $K/\mathbb{Q}_\ell$, let λ be its maximal ideal and let $\mathbb{F} = \mathcal{O}/\lambda$ be the residue field.

We consider the continuous representation

$$\rho_{E,\ell} : \mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F}),$$

and assume it has the following properties:

- $\rho_{E,\ell}$ is modular, *i.e.*, it is the reduction modulo ℓ of some ℓ-adic representation associated to a modular form by Eichler-Shimura-Deligne procedure.
- The restriction of $\rho_{E,\ell}$ to the group $\mathrm{Gal}\left(\bar{\mathbb{Q}}/\mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell}\right)$ is absolutely irreducible.
- The restriction of $\rho_{E,\ell}$ to the decomposition group $D_\ell = \mathrm{Gal}\,(\bar{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ is either of the form
$$\begin{pmatrix} \psi_1 & * \\ 0 & \psi_2 \end{pmatrix}$$
with ψ_1 and ψ_2 distinct characters with ψ_2 unramified; or is induced from a character χ of the unramified quadratic extension of $\mathbb{Q}_\ell$ whose restriction to the inertia I_ℓ group is the fundamental character of level 2, $I_\ell \twoheadrightarrow \mathbb{F}_{\ell^2}$.
- If $p \neq \ell$ then either $\rho_{E,\ell}\mid_{I_p} \simeq \begin{pmatrix} \chi & 0 \\ 0 & 1 \end{pmatrix}$, or $\rho_{E,\ell}\mid_{I_p} \simeq \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$, or $\rho_{E,\ell}\mid_{D_p}$ is absolutely irreducible and in the case $\rho_{E,\ell}\mid_{D_p}$ is absolutely reducible it is assumed that $p \not\equiv -1 \pmod{\ell}$.

By a deformation data we understand a quadruple $\mathcal{D} = \{t, \Sigma, \mathcal{M}, \mathcal{O}\}$ consisting of

- An adjective $t \in \{$ flat, ordinary $\}$.
- A finite set of primes Σ which includes ℓ and all the primes where the elliptic curve E has bad reduction.
- A subset $\mathcal{M}$ of $\Sigma - \{\ell\}$ with the understanding that at each of its primes q, the restrictions to the decomposition groups D_q of both the mod ℓ representation $\rho_{E,\ell}$ or its parent representation on the Tate module $T_\ell(E)$ are prescribed once and for all.
- $\mathcal{O}$ is the ring of integers of a finite extension $K/\mathbb{Q}_\ell$ which contains all the eigenvalues of the modular lift $\tilde{f}$ constructed in the last section.

Let $N(\rho_{E,\ell})$ be the ℓ-free part of the conductor of $\rho_{E,\ell}$. Suppose $\Sigma = \{q_i\}$ and let $N(\rho_{E,\ell}) = \prod q_i^{s_i}$ with $s_i \geq 0$. For each prime q let $n(q) = \dim_{\mathbb{F}_\ell}(E[\ell]^{I_q})$.

Define integers

$$M_0 = N(\rho_{E,\ell}) \prod_{q \notin \mathcal{M} \cup \{\ell\}} q^{n(q)}, \qquad M = M_0 \ell^{\tau(\rho_{E,\ell})},$$

where $\tau(\rho_{E,\ell}) = 1$ if $\rho_{E,\ell}$ is ordinary and $\tau(\rho_0) = 0$ otherwise.

Let H be the subgroup of $(\mathbb{Z}/M\mathbb{Z})^{\times}$ generated by the ℓ-Sylow subgroup of $(\mathbb{Z}/q_i^{r_i}\mathbb{Z})^{\times}$ for each $q_i \in \mathcal{M}$ as well as by $(\mathbb{Z}/q_i\mathbf{z})^{\times}$ for each $q = q_i \in \mathcal{M}$ for which

$$\rho_{E,\ell}\,|_{D_q} = \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$$

for a suitable basis, with χ_1, χ_2 unramified and with the product $\chi_1\chi_2 = \omega$, the Teichmuller character and $\dim_{\mathbb{F}_\ell}(E[\ell]^{I_q}) = 1$.

Recall that the subgroup $H \subseteq (\mathbb{Z}/M\mathbb{Z})^{\times}$ acts on the modular curve $X_1(M)$ via the diamond operators $\Delta = \{\langle a \rangle : (a, M) = 1\}$. We let $X_H(M) = X_H(M)_{/\mathbb{Q}}$ be the quotient $X_1(M)/H$. Let $J_H(M)$ be the Jacobian of $X_H(M)$. We recall that we have Hecke operators (correspondences):

T_p, for $p \nmid M$, $\quad U_q$, for $q \mid M$, $\quad$ and diamond operators $\langle a \rangle$, for $(a, M) = 1$.

The Hecke ring $\mathbf{T}_H(M)$ is defined as the ring of endomorphisms of $J_H(M)$ which is generated over $\mathbb{Z}$ by the Hecke operators

$$\{T_p = T_{p*} \text{ for } p \nmid M,\ \ U_q = U_{q*} \text{ for } q \mid M,\ \ \langle a \rangle = \langle a \rangle_* \text{ for } (a, M) = 1\}.$$

We now introduce the following **Working Hypothesis**:

Associated with $\rho_{E,\ell}$ there is a unique maximal ideal $\underline{m}$ of $\mathbf{T} = \mathbf{T}_H(M)$ and a continuous semisimple Galois representation $\rho_{\underline{m}}$

$$\rho_{\underline{m}} : Gal(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbf{T}/\underline{m}),$$

which satisfies

$$trace\ \rho_{\underline{m}}(Frob_p) = T_p, \quad det\ \rho_{\underline{m}}(Frob_p) = \langle p \rangle p$$

for each prime $p \nmid M\ell$, with the following properties:

(i) $\rho_{E,\ell} = \rho_{\underline{m}}$,
(ii) $U_p \in \underline{m}$ for $p \in \Sigma - \{\mathcal{M} \cup \ell\}$,
(iii) $U_\ell \notin \underline{m}$ if $\rho_{E,\ell}$ is ordinary.

Remark. If $\ell \nmid M$ (resp. $\ell \mid M$) we say that $\underline{m}$ is ordinary if $T_\ell \notin \underline{m}$ (resp. $U_\ell \notin \underline{m}$).

We now define a Hecke ring $\mathbf{T}_{\mathcal{D}}$ which depends on the deformation data $\mathcal{D}$ attached to the mod ℓ representation $\rho_{E,\ell}$. It is defined differently, depending on whether $\rho_{E,\ell}$ is flat at ℓ or ordinary.

Suppose we are in the flat case, *i.e.*, $\mathcal{D} = \{\text{ flat }, \Sigma, \mathcal{M}, \mathcal{O}\}$. We then put

$$\mathbf{T}_{\mathcal{D}} := \mathbf{T}_H(M)_{\underline{m}} \otimes_{\mathbb{Z}_\ell} \mathcal{O},$$

where $\mathbf{T}_H(M)_{\underline{m}}$ is the completion of $\mathbf{T}_H(M)$ at $\underline{m}$.

In the ordinary situation, the ring we want to define is a special case of certain rings first introduces by Hida (see [6], or Berkeley Math. Congress talk (1986)) in his study of the relation between families of ℓ-adic modular eigen forms and continuous families of ℓ-adic Galois representations. In the case under consideration the maximal ideal $\underline{m}$ is ordinary, *i.e.* $U_\ell \notin \underline{m}$. For each integer $n \geq 1$, let $\mathbf{T}_n = \mathbf{T}_H(M_0\ell^n)_{\underline{m}_n}$, where $\underline{m}_n$ is the inverse image of $\underline{m}$ under the natural maps $\mathbf{T}_n \longrightarrow \mathbf{T}_{n-1}$. For a suitable positive integer m, the p-adic limit

$$e = \varprojlim_r U_\ell^{\ell^r(\ell^m - 1)},$$

exists and defines an idenpotent on the Hecke ring $\mathbf{T}_n$ which cuts out the normalized ordinary eigenforms. With Hida, we define a 2-dimensional Noetherian local Hecke ring by setting

$$e\mathbf{T}_H(M_0\ell^\infty)_{\underline{m}} := \varprojlim e\mathbf{T}_H(M_0\ell^n)_{\underline{m}_n}.$$

In the ordinary case, that is to say when $\mathcal{D} = \{\text{ordinary}, \Sigma, \mathcal{M}, \mathcal{O}\}$, we define

$$\mathbf{T}_{\mathcal{D}} = e\mathbf{T}_H(M_0\ell^\infty)_{\underline{m}} \otimes_{\mathbb{Z}_\ell} \mathcal{O}.$$

Remark. The idea of considering limits of Hecke operators seems to have arisen first in the work of O. Atkin concerning *Congruence Hecke Operators*. The deeper properties of these limits with respect to the concept of ordinary modular forms were studied by many mathematicians among whom we only mention Dwork and Serre. Dwork initiated the study of ℓ-adic Banach space properties of the operator U_ℓ by imposing growth conditions on the coefficients of modular forms and found a way to bound the number of unit zeros of the Fredholm determinant associated to U_ℓ. In another direction, Serre's theory of ℓ-adic modular eigen forms and the finiteness properties of the associated modulo ℓ Galois representations lead to a closer study of the interaction between Iwasawa's theory of cyclotomic extensions (via ℓ-adic L-functions) and the congruence properties between modular. This lead Hida to introduce the rings we defined above in the ordinary case and to study their properties as modules over the Iwasawa algebra $\Lambda = \mathbb{Z}_\ell[[T]]$, where $T = \varprojlim\langle 1 + Np\rangle - 1$. A very interesting aspect of Hida's construction, which is not obvious from the way we defined the Hecke rings above is that, as Shimura had observed first, the study of the ℓ-adic properties of modular forms of arbitrary integral weight k can be reduced to weight 2.

In the next section we shall describe how these rings are related to Mazur's universal deformation rings.

5.7 The Universal Deformation Ring. Start with a modulo ℓ representation $\rho_{E,\ell}$ and consider deformation data $\mathcal{D} = \{t, \Sigma, \mathcal{M}, \mathcal{O}\}$ as in the previous section. With Mazur ([62], p. 387) we consider continuous homomorphisms

$$\text{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(R),$$

where R is a complete Noetherian $\mathcal{O}$-algebra with residue field $\mathbb{F}_\lambda$. One requires that ρ should have the following properties:

(1) The composition of ρ with the map $GL_2(R) \longrightarrow GL_2(\mathbb{F}_\lambda)$ induced by the residue map $R \longrightarrow \mathbb{F}_\lambda$ coincides with the starting representation $\rho_{E,\ell}$.

(2) ρ is unramified outside the set Σ.

(3) In keeping with the fact that $\rho_{E,\ell}$ arises from the reduction modulo ℓ of the ℓ-adic representation of Gal$\,(\bar{\mathbb{Q}}/\mathbb{Q})$ on the Tate module $T_\ell(E)$, we require that the representations ρ under consideration also have the same type of behavior as the latter with respect to the restriction to the decomposition group D_ℓ, namely, ρ should be ordinary if E has ordinary or multiplicative reduction at ℓ or that ρ should be flat (*i.e.* ρ can be realized as a D_ℓ-module arising from the action of D_ℓ on the $\bar{\mathbb{Q}}_\ell$-points of some finite flat group scheme over $\mathbb{Z}_\ell$) if E has supersingular reduction.

(4) det ρ should coincide with the composite map

$$\text{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \xrightarrow{\varepsilon} GL_1(\mathbf{z}_\ell) \longrightarrow GL_1(R),$$

under the structural map $\mathbb{Z}_\ell \hookrightarrow \mathcal{O} \hookrightarrow R$.

Mazur calls the representations ρ which possess the properties described above, *deformations of $\rho_{E,\ell}$ of type Σ*. Two such representations ρ, ρ' are naturally identified if they are conjugate by an element of the congruence subgroup ker$(GL_2(R) \longrightarrow GL_2(\mathbb{F}_\lambda))$. The deformations of $\rho_{E,\ell}$ are then simply the equivalence classes of such representations.

A fundamental theorem of Mazur in the ordinary case ([62], p. 400) and of Ramakrishna ([40], p. 281) in the flat case, states that under the Working Hypothesis of the irreducibility of $\rho_{E,\ell}$ there exists a universal deformation

$$\rho_{\mathcal{D}} : \text{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(R_{\mathcal{D}}),$$

associated with the deformation data $\mathcal{D}$ with the property that an arbitrary deformation ρ : Gal$\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(R)$ of type $\mathcal{D}$ may be obtained from $\rho_{\mathcal{D}}$ by a unique homomorphism $R_{\mathcal{D}} \longrightarrow R$.

In the following we refer to $R_{\mathcal{D}}$ as the *universal deformation ring of type $\mathcal{D}$*. It is also important to note here that the Working Hypothesis holds for the representation $\rho_{E,\ell}$.

5.8 "Coronidis loco" [§] Putting together the results stated in the last two sections, Wiles now obtains a canonical surjective map

$$\phi_{\mathcal{D}} : R_{\mathcal{D}} \longrightarrow \mathbf{T}_{\mathcal{D}}$$

which induces the representation $\rho_{\underline{m}}$ introduced in the previous section. As a generalization of some earlier result of Mazur and Tilouine in the ordinary case, Wiles makes the fundamental observation that

$$\phi_{\mathcal{D}} \text{ is an isomorphism.}$$

In the special case under consideration, which is of relevance to the proof of Fermat's Last Theorem, this would imply that all deformations of $\rho_{E,\ell}$ come from modular deformations, and in particular, the ℓ-adic representation on the Tate module $T_\ell(E)$ is associated to a modular form hence establishing the Shimura-Taniyama-Weil conjecture. Before he can reach this "high point", Wiles must establish some very delicate properties of the ring $\mathbf{T}_{\mathcal{D}}$. This is the most technical part of the whole proof and in its final stages it was achieved in joint work with Richard Taylor (see [45]). We describe in this section some of the key components of the proof leading up to Wiles' most important result

Theorem. *With notations as above, the map*

$$\phi_{\mathcal{D}} : R_{\mathcal{D}} \longrightarrow \mathbf{T}_{\mathcal{D}}$$

is an isomorphism.

We recall first that Mazur had proved the fundamental property that the rings of the type $\mathbf{T}_{\mathcal{D}}$ have the *ubiquitous* property of being Gorenstein rings, which gives us perfect pairings of $\mathcal{O}$-modules

$$\langle \ , \ \rangle : \mathbf{T}_{\mathcal{D}} \times \mathbf{T}_{\mathcal{D}} \longrightarrow \mathcal{O}.$$

Let

$$\pi = \pi_{\mathcal{D}} : \mathbf{T}_{\mathcal{D}} \longrightarrow \mathcal{O}$$

be the homomorphism associated to the modular lifting $\tilde{f}$ discussed in section (5.5) and denote by $\tilde{\pi} : \mathcal{O} \longrightarrow \mathbf{T}_{\mathcal{D}}$, the adjoint of π under the above pairing. Define a principal ideal $(\eta_{\mathcal{D}})$ of $\mathbf{T}_{\mathcal{D}}$ by putting

$$(\eta_{\mathcal{D}}) = (\tilde{\pi}(1)).$$

Using the homomorphisms $R_{\mathcal{D}} \xrightarrow{\phi} \mathbf{T}_{\mathcal{D}} \xrightarrow{\pi_{\mathcal{D}}} \mathcal{O}$ we introduce two prime ideals

$$\wp_R := \ker\left(R_{\mathcal{D}} \longrightarrow \mathcal{O}\right), \qquad \wp_{\mathbf{T}} := \ker\left(\mathbf{T}_{\mathcal{D}} \longrightarrow \mathcal{O}\right).$$

The part of the main theorem which is needed for the proof of Fermat's Last Theorem can be restated in the following form.

[§] See Weil's *Basic Number Theory*, p. 288, for the analogous and equally important case arising with respect to the use of Dirichlet's class number formulas in the case of quadratic number fields.

Theorem. *Suppose that the deformation data $\{t, \Sigma, \mathcal{M}, \mathcal{O}\}$ associated to $\rho_{E,\ell}$ is minimal, i.e., $\Sigma = \mathcal{M} \cup \{\ell\}$ and that $\rho_{E,\ell}$ is irreducible when restricted to*

$$Gal(\mathbb{Q}(E[\ell])/\mathbb{Q}\left(\sqrt{(-1)^{(\ell-1)/2}\ell}\right).$$

Let K be the field of fractions of $\mathcal{O}$. Then with notations as above we have:

(i) $\#Hom_{\mathcal{O}}(\wp_R/\wp_R^2, K/\mathcal{O}) = \#(\wp_{\mathbf{T}}/\wp_{\mathbf{T}^2})^2 \cdot c_\ell/\#(\mathcal{O}/\eta_{\mathcal{D}})$, *where $c_\ell = \#(\mathcal{O}/(U_\ell^2 - \langle\ell\rangle))$ when $\rho_{E,\ell}$ is flat at ℓ and $det\rho_{E,\ell} = \omega$ and $c_\ell = 1$ when $\rho_{E,\ell}$ is ordinary.*

(ii) *$\mathbf{T}_{\mathcal{D}}$ is a local complete intersection over $\mathcal{O}$.*

(iii) *$\phi : R_{\mathcal{D}} \longrightarrow \mathbf{T}_{\mathcal{D}}$ is an isomorphism.*

The proof of (i) depends on Wiles' interpretation of the group on the left hand side of the equality as a Selmer group constructed from the symmetric square $\mathrm{Sym}^2 \rho_{\bar{f},\lambda}$ of the representation associated to the weight 2 modular lifting of $\rho_{E,\ell}$. It is here that the local to global principles appear in a Galois cohomological setting and are controlled by the use of appropriate duality pairings.

The crucial reduction for (iii) is based on a criterion developed by Wiles which can be used to verify when a surjective homomorphism between two complete local Noetherian algebras, in which the target is assumed to be a Gorenstein ring, is in fact an isomorphism. The main idea is to check that the Zariski tangent spaces associated to certain prime ideals are isomorphic and possess certain finiteness properties. Using certain ideas from the theory of Fitting ideals and generalizing a result of Tate, Wiles verifies that these properties hold if the target algebra is a local complete intersection. It is this last property, as it applies to a (minimal) Hecke algebra that completes the final argument. (See Wiles [47], chap. III and Appendix; Taylor-Wiles [45]).

5.9 The Fundamental Principles at Work. We began this chapter by describing how Euler discovered that the field $\mathbb{Q}(\sqrt{-3})$, which is closely related to the trisection of the circle, is intimately connected with Fermat's Last Theorem for cubes. We ended the chapter by outlining Wiles' discovery that the solution to Fermat's Last Theorem for all exponents is closely connected with the field $\mathbb{Q}(E[3])$, which arises from the bi-trisection of the torus. We now attempt in general terms to explicate the principles that form the foundation of Wiles' proof.

First and foremost is the Principle of Factorization. On a naive level we can claim that the fundamental theorem of arithmetic for the ordinary integers, or the lack of it in larger domains, has been a central theme in all the work related to Fermat's problem, as is quite obvious in Kummer's work. At an intermediary level, factorization plays a role in the construction of global objects like zeta an L-functions, which are build using Euler products out of local data. This procedure opens up analytic possibilities for the study of arithmetic objects as in the work of Dirichlet, Riemann and Artin. At a much higher level, where the objects of main study are representations arising from infinite dimensional spaces of analytic functions, the principles based on factorization form an important part of the theory. A simple example of this manifests itself in the pivotal role that the Tychonoff theorem about products of locally compact topological spaces plays in the construction of adels and ideles. A more relevant example is the Jacquet-Langlands theorem about the tensor product decomposition of automorphic representations for $GL(2)$, a result which underlies some of the structural properties of the Hecke rings that appear in the work of Wiles.

The second principle is **Fermat's Method of Infinite Descent**. Anyone who has reviewed the factual evidence in favor of the validity of Fermat's claim prior to the work of Wiles cannot fail to see the central role played by variations of Fermat's method of infinite descent which are very close to its original inception, as in the work of Kummer both in the regular and the irregular case. In Wiles' approach, which evolved through the work of many mathematicians, the method of infinite descent appears in a cohomological dressing first developed by the Norwegian mathematician Selmer. Loosely speaking, Wiles' study of certain Selmer groups allows him to control those cohomology classes which are lifts from positive characteristic by including the latter within the family of cohomology classes where the ramification is tame, *i.e.*, whose reduction modulo p is of "multiplicative type". This is an area which in the future will continue to excercise a tight control on many problems of a diophantine nature. It is the great merit of Fermat to have been the first to discover the original idea.

A third item, which can be thought of as a close derivative of the first and second ones is the **Local to Global Principle**. The possibility of constructing global objects from local data, an abstraction which first arouse from Fermat's use of congruences to solve diophantine problems, *e.g.* $p = x^2 + y^2$, or the analysis of global objects by studying the properties of its local components, *e.g.* Langlands' use of the Selberg trace formula in his base-change proof that octahedral representations are automorphic, are just two examples of the role played by this principle in modern number theory. A more concrete example arises in the work of Kolyvagin and in the influence which it had in Wiles' development of his ideas.

A fourth item is what one may loosely refer to as the **Lefchetz Principle** which states that what happens in characteristic zero also happens in characteristic p. It is difficult to pin down a concrete example of its occurrence in Wiles' proof. In its favor we need only think carefully about the meaning and significance of Serre's great insight that *"modulo ℓ representations that look modular actually come from the reduction of modular representations"*. Indirectly we do observe in Wiles' proof the effect that the representations of the groups $\mathrm{Aut}E[\ell] \simeq GL_2(\mathbb{F}_\ell)$ have; as is well known the latter theory is controlled by the finite subgroups of $GL_2(\mathbb{C})$, a situation which can be traced back (in a weak form) to the rigidity of the symmetry groups of the regular solids. By itself this is not a very profound observation; what is indeed surprising is the fact that only for $\ell = 2$ or 3 do these groups admit complex irreducible representations of dimension 2 whose behavior is very similar both in positive characteristic as well as in characteristic 0. For $\ell > 3$ these groups do not admit such representations, even though here one has a very beautiful theorem of Lusztig which describes the Brauer lift of their natural 2-dimensional representation to one over the field $\mathbb{Q}_\ell$. Also for $\ell > 3$ these groups have large families of representations (the cuspidal ones) whose character values on the non-split classes do look as if they were the eigen values Hecke operators, a tantalizing possibility which may force us to admit in Serre's insight the presence of virtual representations, as in the case of the Artin representation.

The singular role played by the group Aut $E[3]$ and its associated Steinberg representation

$$(\det{}^{-1} \otimes \mathrm{Sym}^2)\rho_{E,3} : \mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow g\ell_2(\mathbb{F}_3)$$

is the only visible trace of the Lefchetz principle in the proof. We can only speculate about the limitations that would have arisen, had Wiles not had the prime $\ell = 3$ at his disposal.

The fifth and last principle that should be mentioned is relatively new and stems from the original work of Mazur on the Eisenstein ideal, namely the Gorenstein Property of Hecke Rings. Recall that a local $\mathbb{Z}_\ell$-algebra R, free of finite rank as a module over $\mathbb{Z}_\ell$ is said to be a Gorenstein ring if and only if its $\mathbb{Z}_\ell$-dual $R^* := \mathrm{Hom}_{\mathbb{Z}_\ell}(R, \mathbb{Z}_\ell)$ is free (of rank 1) as a module over R. Mazur demonstrated the deep arithmetic significance of this property by deriving from it a very useful multiplicity one theorem for representations that are realizable on the torsion part of the Jacobian of modular curves. We are at present very far from deriving precise results like Mazur's from the analytic versions of the strong multiplicity one theorem for $GL(n)$, even though conjecturally a theory of motives does suggest strongly analogues of Mazur's results. Wiles' proof goes beyond this property and requires that in fact the completions of the Hecke rings at its proper maximal ideals be Local Complete Intersections. The significance of this new property, together with the tools that Wiles has created to deal with it, merits further closer scrutiny and will very likely be part of the mathematical landscape which deals with diophantine problems.

Appendix: Announcements

1. June 23, 1993: Wiles announces his solution at the end of a three day conference in Cambridge, England.
2. November 1993: Marilyn vos Savant's notorious book is published....!

3. Wiles' Fermat Status(Internet 12/4/93): "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 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". A. W.

4. October 24, 1994: Manuscripts by Wiles and by Taylor-Wiles are released.

5. May 1995: The articles by Wiles and Taylor are published in Annals of Mathematics.

References

1. MICHAEL SEAN MAHONEY, *The Mathematical Career of Pierre De Fermat (1601–1665)*, Princeton University Press, Princeton, New Jersey, 1973.
2. A. WEIL, *A review of Mahoney's book*, Bull. A.M.S. **79** (1973), no. 6, pp. 1138–1149.
3. PAUL HALMOS, *I have a photographic Memory*, American Mathematical Society, Providence, Rhode Island, 1987.
4. J. HERBRAND, *Sur Les Corps Circulaires*, Jour. Pures et Appl., **38** (1932), pp. 417-441.
5. J. HERBRAND, *Le developpment moderne de la theorie des corps algebriques*, Memorial des Sciences Mathematiques **75** (1936), pp. 1-72.
6. H. HIDA, *Iwasawa modules attached to congruences of cusp forms*, Ann. Sci. Ecole Norm. Sup. (4) **19** (1986), pp. 231-273.
7. J. TUNNEL, *Artin Conjecture for 2 dimensional representations of octahedral type*, Bull. A.M.S. **5** (1981), 173–175.
8. J. P. SERRE, *Proprietés Galoisiennes des Points d'ordre fini des coubes elliptiques*, Invent. math. **15** (1972), 259–331.
9. M. FLACH, *A finiteness theorem for the symmetric square of an elliptic curve*, Invent. math. **109** (1992), 307–327.
10. M. FLACH, *A generalization of the Cassels-Tate pairing*, Jour. Reine Ungewand. Math. **412** (1990), pp. 113-127.
11. F. THAINE, *On the ideal class groups of real abelian number fields*, Ann. of Math. (2) **128** (1988), pp. 1–18.
12. K. RUBIN, *The work of Kolyvagin on the arithmetic of elliptic curves, Arithmetic of complex manifolds*, Lecture Notes in Math. vol. 1399, Springer–Verlag, New York, 1989, pp. 128–136.

13. K. Rubin & A. Silverberg, *A report on Wiles's Cambridge Lectures*, Bull. A.M.S., **31** (1994), pp. 15-38.

14. P. Ribenboim, *13 Lectures on Fermat's Last Theorem*, Springer-Verlag, New York, 1993.

15. B. Mazur, *On the passage from local to global in number theory*, Bull. A.M.S. (1) **29** (1993), pp. 14–50.

16. A. Wiles & J. Coates, *On the conjecture of Birch and Swinnerton–Dyer*, Invent. Math. **39** (1977).

17. A. Wiles, *Modular curves and the class group of* $\mathbb{Q}(\zeta_p)$, Invent. Math. **58** (1980), pp. 1–35.

18. B. Mazur & A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Invent. Math. **76** (1984), pp. 179–330.

19. K. Ribet, *On modular representations of* $Gal(\overline{Q}/Q)$ *arising from madular forms*, Invent. Math. **100** (1990), pp. 431-476.

20. J. P. Serre, *Sur les representations modulaires de degrée 2 de* $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, Duke Math. Jour., **54** (1987), pp. 179-230.

21. G. Frey, *Links between stable elliptic curves and certain diophantine equations*, Annales Universitatis Saraviensis **1**, (1986), pp. 1–40.

22. N. Arthaud, *On Birch and Swinnerton-Dyer's Conjecture for elliptic curves with complex multiplication. I*, Compositio Mathematica **37** (1978), pp. 209-232.

23. B.J. Birch, & W. Kwyk, eds., *Modular Functions of One Variable IV*, Lecture Notes in Math., vol. 476, Springer -Verlag New York, 1975.

24. S.Bloch, & K. Kato, *L-Functions and Tamagawa Numbers of Motives.* In: *The Grothendieck Festschrift*, vol. 1, (Progess in Mathematics Series, vol. 86.) Birkhauser, Boston, 1990.

25. P. Berthelot, *Systemes de Honda des Schemas en* $\mathbb{F}_p$*-Vectoriels*, Bull. Soc. Math. France, **105** 1977, pp. 225–239.

26. J. W. S. Cassels, *Lectures on Elliptic Curves*, London Mathematical Society, Student Texts 24, Cambridge University Press, New York, 1991.

27. J. W. S. Cassels, *Diophantine equations with special reference to elliptic curves*, Jour. London Math. Soc., **41** (1966), pp. 193–291.

28. J. Coates & C. Schmidt, *Iwasawa Theory for the Symmetric Square of an Elliptic Curve*, J. Crelle **375/376** (1987), pp. 104–156.

29. F. Diamond, *Congruence primes for cusp forms of weight* $k \geq 2$, Asterique **196-197** (1991), pp. 205-213.

30. N. Elkies, *Wiles minus epsilon implies Fermat*, preprint, June-July, 1993.

31. R. Heath-Brown, *Fermat's Last Theorem holds for almost all exponents*, Bull. London Math. Soc. **17** (1985), pp. 15-16.

32. J.E. Humphreys, *The Steinberg Representation*, Bull. A.M.S., **16**, no. 2 (1987), pp. 247–263.

33. M. Koike, *Congruences Between Cusp Forms and Linear Representations of the Galois Group.* Nagoja Math. Jour. **64** (1976), pp. 63–85.

34. S. Lang, *Cyclotomic Fields Vols. I and II*, Springer-Verlag, New York, 1990.

35. R. Langlands, *Base Change for GL(2)*, Annals of Math. Studies, vol. 96, Princeton Univ. Press., Princeton 1980.

36. L. J. Moredell, *Three Lectures on Fermat's Last Theorem*, Chelsea Publ. Co., New York, 1962.

37. C. J. Moreno, *Algebraic Curves over Finite Fields*, Cambridge Tracts in Math., vol. 97, Cambridge, 1991.

38. J. Oesterle, *Nouvelles approches du "Theoreme de Fermat"*, Sem. Bourbaki, 40e annee, 1987–88, no. 694, 1988.

39. F. Pollaczek, *Uber die irregularen Kreiskorper der* ℓ*-ten und* ℓ^2*-ten Ein-*

heitswurzeln, Math. Zeit. **21** (1924), pp. 1–37.

40. R. RAMAKRISHNA, *On a variation of Mazur's deformation functor*, Compositio Mathematica **87** (1993), pp. 269–286.
41. E.S. SELMER, *The diophantine equation* $ax^3 + by^3 + cz^3 = 0$, Acta Math. **85** (1951), pp. 203–262.
42. G. SHIMURA, *A reciprocity Law in Non-Solvable Extensions,* Jour. Crelle **221** (1966), pp. 209–220.
43. J. SILVERMAN, *The Arithmetic of Elliptic Curves,* Graduate Texts in Math. vol. 106, Springer-Verlag, New York, 1986.
44. J. TATE, *The Arithmetic of Elliptic Curves,* Invent. Math. **23** (1974), pp. 179–206.
45. R. TAYLOR & A. WILES, *Ring Theoretic Properties of Certain Hecke algebras,* Ann. of Math. **141** no. 3 (1995), pp. 553–572.
46. J. TUNNELL, *Artin's Conjecture for Representations of Octahedral Type*, Bull. A.M.S. **5** (1981), pp. 173–175.
47. A. WILES, *Elliptic Modular Curves and Fermat's Last Theorem,* Ann. of Math. **141** no. 3 (1995), pp. 443–551.
48. K. A. RIBET, *A modular construction of unramified p-extensions of* $\mathbb{Q}(\mu_p)$, Inv. Math. **34** (1976), pp. 151–162.
49. P. DELIGNE & M. RAPOPORT, *Schémes des modules de courbes elliptiques,* Lecture Notes in Math., vol. 349, Springer, New York, 1973.
50. G. SHIMURA, *Introduction to the Arithmetic Theory of Modular Forms,* Publ. Math Soc. Japan, vol. 11, Tokyo-Princeton, 1971.
51. G. SHIMURA, *On elliptic curves with complex multiplication as factors of the jacobians of modular function fields,* Nagoya Math. Jour. **43** (1971), pp. 199–208.
52. P. DELIGNE & J.-P. SERRE, *Form Modulaires de poids 1,* Ann. Scient. Ec. Norm.Sup., 4^e série **7** (1074), pp. 507–530.
53. T. MIYAKE, *On automorphic forms on* GL_2 *and Hecke operators,* Ann. of Math. **94** (1971), pp. 174–189.
54. M. RAYNAUD, *Schemes en groupes de type* $(p, ..., p)$, Bull Soc. Math. France **102** (1974), pp. 241–280.
55. A. OGG, *A survey of Modular Functions of one variable,* Lecture Notes in Math., vol. 349, Springer, New York, 1973.
56. J.-P. SERRE, *Formes modulaires et fonctions zeta,* Lecture Notes in Math, vol. 350, Springer-Verlag, New York, 1973.
57. G. FALTINGS, *Endlichkeitssatze fur abelsche Varietaten uber Zahlkorpern,* Invent. Math. **73** (1983), pp. 349–366.
58. R. GRENNBERG, *Iwasawa Theory and p-adic Deformations of Motives,* Proc. Symposia Pure Math. **55** (1994), Part 2, pp. 193–223.
59. LI GUO, *On a generalization of Tate Dualities with application to Iwasawa theory,* Compositio Mathematica **85** (1993), pp. 125–161.
60. B. MAZUR, *Ernst Edward Kummer, Collected Papers, Vols. I and II edited by A. Weil,* Bull. A.M.S. **83** (1977), 976–988.
61. B. MAZUR, *Number Theory as Gadfly,* Amer. Math. Monthly, August-September (1991), pp. 593–610.
62. B. MAZUR, *Deforming Galois Representations,* In *Galois Groups over* $\mathbb{Q}$, Y. Ihara, K. Ribet & J. P. Serre, Eds., Mathematical Sciences Research Institute Publications, vol. 16, Springer-Verlag, New York, 1989, pp. 385–437.
63. B. MAZUR, *Modular Curves and the Eisenstein Ideal,* Publ. IHES **47** (1977), pp. 33–186.
64. B. MAZUR & J. TILOUINE, *Representationnes galoisiennes, differentielles de Kahler et conjectures principales,* Publ. IHES **71** (1990), pp. 65–103.
65. E. E. KUMMER, *Allgemeiner Beweis des Fermat'schen Satzes, dass die Gleichung*

$x^\ell + y^\ell = z^\ell$ *durch ganze Zahlen unlosbar ist, fur alle diejenigen Potenz-Exponenten* λ, *welche ungerade Primzahlen sind und in den Zahlern der ersten* $1/2(\lambda - 3)$ *Bernoulli'schen Zahlen als Factoren nicht vorkommen,* Crelle **40** (1850), pp. 130–138. (Collected Papers, vol. I, pp. 336-344).

66. H. M. EDWARDS, *Fermat's Last Theorem,* Graduate Texts in Mathematics, vol. 50, Springer-Verlag, New York, 1977.

67. H.S. VANDIVER, *Fermat's Last Theorem: Its History and the Nature of the Known Results Concerning it,* Amer. Math. Monthly, December 1946, pp. 555–578.

68. L. WASHINGTON, *Introduction to Cyclotomic Fields,* Springer-Verlag, New York, 1980.

69. A. WEIL, *Uber die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen,* Math. Ann. **168** (1967), pp. 149–156.

70. A. WEIL, *Number Theory: An Approach Through History from Hammurapi to Legendre,* Birkhauser, Boston, 1984.

71. A. WEIL, *Collected Papers,* Springer-Verlag, New York, 1979.

(Recibido en Agosto de 1995)

CARLOS JULIO MORENO
BARUCH COLLEGE & GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK
NEW YORK, NY 10010
e-mail: carlos@newton.baruch.cuny.edu