Elliptic Curves, Modular Forms, and Fermat's Last Theorem

second edition

edited by

John H. Coates and Shing-Tung Yau

International Press
www.intlpress.com

Elliptic Curves, Modular Forms, and Fermat's Last Theorem, 2nd edition

Editors:
John H. Coates, University of Cambridge
Shing-Tung Yau, Harvard University

ISBN 978-1-57146-185-8

Paperback reissue 2010. Previously published in 1997 under ISBN 1-57146-049-7 (clothbound).

Typeset using the LaTeX system.

Table of Contents

Conference on Elliptic Curves and Modular Forms
Hong Kong, December 18-21, 1993

Foreword

A conference, on the general theme of "Elliptic curves and modular forms", was held in the Mathematics Department and the Institute of Mathematical Sciences of The Chinese University of Hong Kong from December 18–21, 1993. The impetus for organizing the conference arose from Andrew Wiles' deep and spectacular work on the celebrated conjecture that every elliptic curve over Q is modular, although only some of the lectures at the conference were specifically related to this theme. At the time of the conference, the difficulties in the last hurdle in Wiles' work (the proof of the conjectural upper bound for the order of the Selmer group attached to the symmetric square of a modular form) had still not been overcome. However, the optimism shared by all at the conference that is was only a matter of time until the proof would become complete has happily been borne out by subsequent events. It is now history that Wiles himself, assisted by R. Taylor, found a beautiful proof of the desired upper bound. As a result, we now know today the remarkable fact that every semi-stable elliptic curve over Q is modular. Not only is this result revolutionary in its own right for the study of the arithmetic of these elliptic curves, but it has the added bonus that it provides at last a proof of Fermat's last theorem, thanks to the earlier work of Frey, Ribet and others. We are grateful to H. Darmon, F. Diamond and R. Taylor for their kind permission to publish their beautiful survey article on Wiles' work as the first article in this second edition. During the conference itself, lectures were given by John Coates, Noam Elkies, Matthias Flach, Jean-Marc Fontaine, Gerhard Frey, Dick Gross, Victor Kolyvagin, Ken Ribet, Karl Rubin, Jean-Pierre Serre, John Tate, Richard Taylor, and Don Zagier. The present short volume is a mixture of the texts of some of these lectures, together with a number of recent articles related to the general theme of the conference. Finally, the organizers of the conference wish to express their warmest thanks to Professor Charles Kao, Vice-Chancellor, and to Professor S.Y. Cheng, Chairman of the Mathematics Department, and Dr. K.W. Leung, of The Chinese University of Hong Kong for their support and assistance throughout the preparation and running of the conference. Finally, and most importantly, the organizers wish to express their gratitude to The Chinese University of Hong Kong, the Ho Sin Hang Education Endowment Fund, the Lee Hysan Foundation Limited, the Pei Hua Education Foundation Limited, and the Sing Ho Yu Foundation for their generous financial support for the conference.

John Coates, Cambridge University
Shing-Tung Yau, Director
Institute of Mathematical Sciences
The Chinese University of Hong Kong

Fermat's Last Theorem

HENRI DARMON
Department of Mathematics, McGill University, Montreal, QC
Canada H3A 2K6
FRED DIAMOND
Departament of Mathematics, MIT, 77 Massachusetts Avenue,
Camridge, MA 02139, USA
RICHARD TAYLOR
Departament of Mathematics, Harvard University, 1 Oxford Street,
Cambridge, MA 02138, USA

The authors would like to give special thanks to N. Boston, K. Buzzard, and B. Conrad for providing so much valuable feedback on earlier versions of this paper. They are also grateful to A. Agboola, M. Bertolini, B. Edixhoven, J. Fearnley, B.H. Gross, R. Gross, L. Guo, F. Jarvis, H. Kisilevsky, E. Liverance, J. Manoharmayum, K. Ribet, D. Rohrlich, M. Rosen, R. Schoof, J.-P. Serre, C. Skinner, D. Thakur, J. Tilouine, J. Tunnell, A. Van der Poorten, and L. Washington for their helpful comments.

Darmon thanks the members of CICMA and of the Quebec-Vermont Number Theory Seminar for many stimulating conversations on the topics of this paper, particularly in the Spring of 1995. For the same reason Diamond is grateful to the participants in an informal seminar at Columbia University in 1993-94, and Taylor thanks those attending the Oxford Number Theory Seminar in the Fall of 1995.

Parts of this paper were written while the authors held positions at other institutions: Darmon at Princeton, Diamond at Cambridge and the Institute for Advanced Study, and Taylor at Cambridge and Oxford. During some of the period, Diamond enjoyed the hospitality of Princeton University, and Taylor that of MIT. The writing of this paper was also supported by research grants from NSERC (Darmon), EPSRC # GR/J9461 and NSF # DMS 9304580 (Diamond) and by an advanced fellowship from EPSRC (Taylor).

This article owes everything to the ideas of Wiles, and the arguments presented here are fundamentally his [W3], though they include both the work [TW] and several simplifications to the original arguments, most notably that of Faltings. In the hope of increasing clarity, we have not always stated theorems in the greatest known generality, concentrating instead on what is needed for the proof of the Shimura-Taniyama conjecture for semi-stable elliptic curves. This article can serve as an introduction to the fundamental papers [W3] and [TW], which the reader is encouraged to consult for a different, and often more in-depth, perspective on the topics considered. Another useful more advanced reference is the article [Di2] which strengthens the methods of [W3] and [TW] to prove that every elliptic curve that is semistable at 3 and 5 is modular.

For the reprinting of this paper the only changes we have made are to correct a few typos. We have not updated the material to take account of recent developments in the subject, most notably:

- F. Diamond, *The Taylor-Wiles construction and multiplicity one*, Invent. Math., **128** (1997), 379-391 gives a simplification to the arguments which we would have incorporated into this exposition if we were writing it today.

- Some progress has been made on conjecture 3.13 for small primes. See N. Shepherd-Barron and R. Taylor, *Mod 2 and mod 5 icosahedral representations*, J. Amer. Math. Soc., **10** (1997), 283-298.

- The Shimura-Taniyama conjecture is now known for elliptic curves with conductor not divisible by 27. See B. Conrad, F. Diamond, R. Taylor, *Modularity of certain potentially crystalline Galois representations*, in preparation.

Contents

Introduction

Fermat's Last Theorem

Fermat's Last Theorem states that the equation

$$x^n + y^n = z^n, \quad xyz \neq 0$$

has no integer solutions when n is greater than or equal to 3. Around 1630, Pierre de Fermat claimed that he had found a "truly wonderful" proof of this theorem, but that the margin of his copy of Diophantus' *Arithmetica* was too small to contain it:

> "Cubum autem in duos cubos, aut quadrato quadratum in duos quadrato quadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos ejusdem nominis fas est dividere; cujus rei demonstrationem mirabile sane detexi. Hanc marginis exiguitas non caperet."

Among the many challenges that Fermat left for posterity, this was to prove the most vexing. A tantalizingly simple problem about whole numbers, it stood unsolved for more than 350 years, until in 1994 Andrew Wiles finally laid it to rest.

Prehistory: The only case of Fermat's Last Theorem for which Fermat actually wrote down a proof is for the case $n = 4$. To do this, Fermat introduced the idea of *infinite descent* which is still one the main tools in the study of Diophantine equations, and was to play a central role in the proof of Fermat's Last Theorem 350 years later. To prove his Last Theorem for exponent 4, Fermat showed something slightly stronger, namely that the equation $x^4 + y^4 = z^2$ has no solutions in relatively prime integers with $xyz \neq 0$. Solutions to such an equation correspond to rational points on the elliptic curve $v^2 = u^3 - 4u$. Since every integer $n \geq 3$ is divisible either by an odd prime or by 4, the result of Fermat allowed one to reduce the study of Fermat's equation to the case where $n = \ell$ is an *odd prime*.

In 1753, Leonhard Euler wrote down a proof of Fermat's Last Theorem for the exponent $\ell = 3$, by performing what in modern language we would call a 3-descent on the curve $x^3 + y^3 = 1$ which is also an elliptic curve. Euler's argument (which seems to have contained a gap) is explained in [Edw], ch. 2, and [Dic1], p. 545.

It took mathematicians almost 100 years after Euler's achievement to handle the case $\ell = 5$; this was settled, more or less simultaneously, by Gustav Peter Lejeune Dirichlet [Dir] and Adrien Marie Legendre [Leg] in 1825. Their elementary arguments are quite involved. (Cf. [Edw], sec. 3.3.)

In 1839, Fermat's equation for exponent 7 also yielded to elementary methods, through the heroic efforts of Gabriel Lamé. Lamé's proof was even more intricate than the proof for exponent 5, and suggested that to go further, new theoretical insights would be needed.

The work of Sophie Germain: Around 1820, in a letter to Gauss, Sophie Germain proved that if ℓ is a prime and $q = 2\ell + 1$ is also prime, then Fermat's equation $x^\ell + y^\ell = z^\ell$ with exponent ℓ has no solutions (x, y, z) with $xyz \neq 0$

(mod ℓ). Germain's theorem was the first really general proposition on Fermat's Last Theorem, unlike the previous results which considered the Fermat equation one exponent at a time.

The case where the solution (x, y, z) to $x^\ell + y^\ell = z^\ell$ satisfies $xyz \not\equiv 0$ (mod ℓ) was called the ***first case*** of Fermat's Last Theorem, and the case where ℓ divides xyz, the ***second case***. It was realized at that time that the first case was generally easier to handle: Germain's theorem was extended, using similar ideas, to cases where $k\ell + 1$ is prime and k is small, and this led to a proof that there were no first case solutions to Fermat's equation with prime exponents $\ell \leq 100$, which in 1830 represented a significant advance. The division between first and second case remained fundamental in much of the later work on the subject. In 1977, Terjanian [Te] proved that if the equation $x^{2\ell} + y^{2\ell} = z^{2\ell}$ has a solution (x, y, z), then 2ℓ divides either x or y, i.e., "the first case of Fermat's Last Theorem is true for even exponents". His simple and elegant proof used only techniques that were available to Germain and her contemporaries.

The work of Kummer: The work of Ernst Eduard Kummer marked the beginning of a new era in the study of Fermat's Last Theorem. For the first time, sophisticated concepts of algebraic number theory and the theory of L-functions were brought to bear on a question that had until then been addressed only with elementary methods. While he fell short of providing a complete solution, Kummer made substantial progress. He showed how Fermat's Last Theorem is intimately tied to deep questions on class numbers of cyclotomic fields which are still an active subject of research. Kummer's approach relied on the factorization

$$(x + y)(x + \zeta_\ell y) \cdots (x + \zeta_\ell^{\ell-1} y) = z^\ell$$

of Fermat's equation over the ring $\mathbb{Z}[\zeta_\ell]$ generated by the ℓth roots of unity. One observes that the greatest common divisor of any two factors in the product on the left divides the element $(1 - \zeta_\ell)$, which is an element of norm ℓ. Since the product of these numbers is a perfect ℓ-th power, one is tempted to conclude that $(x+y), \ldots, (x+\zeta_\ell^{\ell-1} y)$ are each ℓ-th powers in the ring $\mathbb{Z}[\zeta_\ell]$ up to units in this ring, and up to powers of $(1 - \zeta_\ell)$. Such an inference would be valid if one were to replace $\mathbb{Z}[\zeta_\ell]$ by $\mathbb{Z}$, and is a direct consequence of *unique factorization* of integers into products of primes. We say that a ring R has property UF if every non-zero element of R is uniquely a product of primes, up to units. Mathematicians such as Lamé made attempts at proving Fermat's Last Theorem based on the mistaken assumption that the rings $\mathbb{Z}[\zeta_\ell]$ had property UF. Legend even has it that Kummer fell into this trap, although this story now has been discredited; see for example [Edw], sec. 4.1. In fact, property UF is far from being satisfied in general: one now knows that the rings $\mathbb{Z}[\zeta_\ell]$ have property UF only for $\ell < 23$ (cf. [Wa], ch. 1).

It turns out that the full force of property UF is not really needed in the applications to Fermat's Last Theorem. Say that a ring R has property UF_ℓ if the following inference is valid:

$$ab = z^\ell, \text{ and } \gcd(a, b) = 1 \Rightarrow a \text{ and } b \text{ are } \ell\text{th powers up to units of } R.$$

If a ring R has property UF, then it also has property UF_ℓ, but the converse need not be true. Kummer showed that Fermat's last theorem was true for exponent ℓ if $\mathbb{Z}[\zeta_\ell]$ satisfied the property UF_ℓ (cf. [Wa]). The proof is far from trivial, because of difficulties arising from the units in $\mathbb{Z}[\zeta_\ell]$ as well as from the possible failure of property UF. (A number of Kummer's contemporaries, such as Cauchy and Lamé, seem to have overlooked both of these difficulties in their attempts to prove Fermat's Last Theorem.)

Kummer then launched a systematic study of the property UF_ℓ for the rings $\mathbb{Z}[\zeta_\ell]$. He showed that even if $\mathbb{Z}[\zeta_\ell]$ failed to have unique factorization, it still possessed unique factorization into prime *ideals*. He defined the *ideal class group* as the quotient of the group of fractional ideals by its subgroup consisting of principal ideals, and was able to establish the finiteness of this class group. The order of the class group of $\mathbb{Z}[\zeta_\ell]$, denoted h_ℓ, could be taken as a measure of the failure of the ring $\mathbb{Z}[\zeta_\ell]$ to satisfy UF. It was rather straightforward to show that if ℓ did not divide h_ℓ, then $\mathbb{Z}[\zeta_\ell]$ satisfied the property UF_ℓ. In this case, one called ℓ a *regular prime.* Kummer thus showed that Fermat's last theorem is true for exponent ℓ if ℓ is a regular prime.

He did not stop here. For it remained to give an efficient means of computing h_ℓ, or at least an efficient way of checking when ℓ divides h_ℓ. The class number h_ℓ can be factorized as a product

$$h_\ell = h_\ell^+ h_\ell^-,$$

where h_ℓ^+ is the class number of the real subfield $\mathbb{Q}(\zeta_\ell)^+$, and h_ℓ^- is defined as h_ℓ/h_ℓ^+. Essentially because of the units in $\mathbb{Q}(\zeta_\ell)^+$, the factor h_ℓ^+ is somewhat difficult to compute, while, because the units in $\mathbb{Q}(\zeta_\ell)^+$ generate the group of units in $\mathbb{Q}(\zeta_\ell)$ up to finite index, the term h_ℓ^- can be expressed in a simple closed form. Kummer showed that if ℓ divides h_ℓ^+, then ℓ divides h_ℓ^-. Hence, ℓ divides h_ℓ if and only if ℓ divides h_ℓ^-. This allowed one to avoid the difficulties inherent in the calculation of h_ℓ^+. Kummer then gave an elegant formula for h_ℓ^- by considering the Bernoulli numbers B_n, which are rational numbers defined by the formula

$$\frac{x}{e^x - 1} = \sum \frac{B_n}{n!} x^n.$$

He produced an explicit formula for the class number h_ℓ^-, and concluded that if ℓ does not divide the numerator of B_{2i}, for $1 \leq i \leq (\ell - 3)/2$, then ℓ is regular, and conversely.

The conceptual explanation for Kummer's formula for h_ℓ^- lies in the work of Dirichlet on the analytic class number formula, where it is shown that h_ℓ^- can be expressed as a product of special values of certain (abelian) L-series

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$$

associated to odd Dirichlet characters. Such special values in turn can be expressed in terms of certain generalized Bernoulli numbers $B_{1,\chi}$, which are related to the Bernoulli numbers B_i via congruences mod ℓ. (For more details, see [Wa].)

These considerations led Kummer to initiate a deep study relating congruence properties of special values of L-functions and of class numbers, which was to emerge as a central concern of modern algebraic number theory, and was to reappear – in a surprisingly different guise – at the heart of Wiles' strategy for proving the Shimura-Taniyama conjecture.

Later developments: Kummer's work had multiple ramifications, and led to a very active line of enquiry pursued by many people. His formulae relating Bernoulli numbers to class numbers of cyclotomic fields were refined by Kenneth Ribet [R1], Barry Mazur and Andrew Wiles [MW], using new methods from the theory of modular curves which also play a central role in Wiles' more recent work. (Later Francisco Thaine [Th] reproved some of the results of Mazur and Wiles using techniques inspired directly from a reading of Kummer.) In a development more directly related to Fermat's Last Theorem, Wieferich proved that if ℓ^2 does not divide $2^{\ell-1}-1$, then the first case of Fermat's Last Theorem is true for exponent ℓ. (Cf. [Ri], lecture VIII.)

There were many other refinements of similar criteria for Fermat's Last theorem to be true. Computer calculations based on these criteria led to a verification that Fermat's Last theorem is true for all odd prime exponents less than four million [BCEM], and that the first case is true for all $\ell \leq 8.858 \cdot 10^{20}$ [Su].

The condition that ℓ is a regular prime seems to hold heuristically for about 61% of the primes. (See the discussion on p. 63, and also p. 108, of [Wa], for example.) In spite of the convincing numerical evidence, it is still not known if there are infinitely many regular primes. Ironically, it is not too difficult to show that there are infinitely many irregular primes. (Cf. [Wa].)

Thus the methods introduced by Kummer, after leading to very strong results in the direction of Fermat's Last theorem, seemed to become mired in difficulties, and ultimately fell short of solving Fermat's conundrum[1].

Faltings' proof of the Mordell conjecture: In 1985, Gerd Faltings [Fa] proved the very general statement (which had previously been conjectured by Mordell) that any equation in two variables corresponding to a curve of genus strictly greater than one had (at most) finitely many rational solutions. In the context of Fermat's Last Theorem, this led to the proof that for each exponent $n \geq 3$, the Fermat equation $x^n + y^n = z^n$ has at most finitely many integer solutions (up to the obvious rescaling). Andrew Granville [Gra] and Roger Heath-Brown [HB] remarked that Faltings' result implies Fermat's Last Theorem for a set of exponents of density one.

However, Fermat's Last Theorem was still not known to be true for an infinite set of prime exponents. In fact, the theorem of Faltings seemed ill-equipped for dealing with the finer questions raised by Fermat in his margin, namely of finding a complete list of rational points on *all* of the Fermat curves $x^n + y^n = 1$ simultaneously, and showing that there are no solutions on these curves when $n \geq 3$ except the obvious ones.

Mazur's work on Diophantine properties of modular curves: Although it was not realized at the time, the chain of ideas that was to lead to a proof of Fermat's Last theorem had already been set in motion by Barry Mazur in the mid seventies. The modular curves $X_0(\ell)$ and $X_1(\ell)$ introduced in section 1.2 and 1.5 give rise to another naturally occurring infinite family of Diophantine equations. These equations have certain systematic rational solutions corresponding to the cusps that are defined over $\mathbb{Q}$, and are analogous to the

[1] However, W. McCallum has recently introduced a technique, based on the method of Chabauty and Coleman, which suggests new directions for approaching Fermat's Last Theorem via the cyclotomic theory. An application of McCallum's method to showing the *second* case of Fermat's Last Theorem for regular primes is explained in [Mc].

so-called "trivial solutions" of Fermat's equation. Replacing Fermat curves by modular curves, one could ask for a complete list of all the rational points on the curves $X_0(\ell)$ and $X_1(\ell)$. This problem is perhaps even more compelling than Fermat's Last Theorem: rational points on modular curves correspond to objects with natural geometric and arithmetic interest, namely, elliptic curves with cyclic subgroups or points of order ℓ. In [Maz1] and [Maz2], B. Mazur gave essentially a complete answer to the analogue of Fermat's Last Theorem for modular curves. More precisely, he showed that if $\ell \neq 2, 3, 5$ and 7, (i.e., $X_1(\ell)$ has genus > 0) then the curve $X_1(\ell)$ has no rational points other than the "trivial" ones, namely cusps. He proved analogous results for the curves $X_0(\ell)$ in [Maz2], which implied, in particular, that an elliptic curve over $\mathbb{Q}$ with square-free conductor has no rational cyclic subgroup of order ℓ over $\mathbb{Q}$ if ℓ is a prime which is strictly greater than 7. This result appeared a full ten years before Faltings' proof of the Mordell conjecture.

Frey's strategy: In 1986, Gerhard Frey had the insight that these constructions might provide a precise link between Fermat's Last Theorem and deep questions in the theory of elliptic curves, most notably the Shimura Taniyama conjecture. Given a solution $a^\ell + b^\ell = c^\ell$ to the Fermat equation of prime degree ℓ, we may assume without loss of generality that $a^\ell \equiv -1 \pmod 4$ and that $b^\ell \equiv 0 \pmod{32}$. Frey considered (following Hellegouarch, [He], p. 262; cf. also Kubert-Lang [KL], ch. 8, §2) the elliptic curve

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

This curve is *semistable*, i.e., it has square-free conductor. Let $E[\ell]$ denote the group of points of order ℓ on E defined over some (fixed) algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$, and let L denote the smallest number field over which these points are defined. This extension appears as a natural generalization of the cyclotomic fields $\mathbb{Q}(\zeta_\ell)$ studied by Kummer. What singles out the field L for special attention is that it has *very little ramification*: using Tate's analytic description of E at the primes dividing abc, it could be shown that L was ramified only at 2 and ℓ, and that the ramification of L at these two primes was rather restricted. (See theorem 2.15 of section 2.2 for a precise statement.) Moreover, the results of Mazur on the curve $X_0(\ell)$ could be used to show that L is *large*, in the following precise sense. The space $E[\ell]$ is a vector space of dimension 2 over the finite field $\mathbb{F}_\ell$ with ℓ elements, and the absolute Galois group $G_{\mathbb{Q}} = \mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q})$ acts $\mathbb{F}_\ell$-linearly on $E[\ell]$. Choosing an $\mathbb{F}_\ell$-basis for $E[\ell]$, the action is described by a representation

$$\bar{\rho}_{E,\ell} : \mathrm{Gal}\,(L/\mathbb{Q}) \hookrightarrow GL_2(\mathbb{F}_\ell).$$

Mazur's results in [Maz1] and [Maz2] imply that $\bar{\rho}_{E,\ell}$ is *irreducible* if $\ell > 7$ (using the fact that E is *semi-stable*). In fact, combined with earlier results of Serre [Se6], Mazur's results imply that for $\ell > 7$, the representation $\bar{\rho}_{E,\ell}$ is surjective, so that $\mathrm{Gal}\,(L/\mathbb{Q})$ is actually isomorphic to $GL_2(\mathbb{F}_\ell)$ in this case.

Serre's conjectures: In [Se7], Jean-Pierre Serre made a careful study of mod ℓ Galois representations $\bar{\rho} : G_{\mathbb{Q}} \longrightarrow GL_2(\mathbb{F}_\ell)$ (and, more generally, of representations into $GL_2(k)$, where k is any finite field). He was able to make very precise conjectures (see section 3.2) relating these representations to modular forms mod ℓ. In the context of the representations $\bar{\rho}_{E,\ell}$ that occur in Frey's construction, Serre's conjecture predicted that they arose from modular forms (mod ℓ) of weight two and level two. Such modular forms, which correspond to differentials on the modular curve $X_0(2)$, do not exist because $X_0(2)$ has genus 0. Thus Serre's conjecture implied Fermat's Last Theorem. The link between fields with Galois groups contained in $GL_2(\mathbb{F}_\ell)$ and modular forms mod ℓ still appears to be very deep, and Serre's conjecture remains a tantalizing open problem.

Ribet's work: lowering the level: The conjecture of Shimura and Taniyama (cf. section 1.8) provides a direct link between elliptic curves and modular forms. It predicts that the representation $\bar{\rho}_{E,\ell}$ obtained from the ℓ-division points of the Frey curve arises from a modular form of weight 2, albeit a form whose level is quite large. (It is the product of all the primes dividing abc, where $a^\ell + b^\ell = c^\ell$ is the putative solution to Fermat's equation.) Ribet [R5] proved that, if this were the case, then $\bar{\rho}_{E,\ell}$ would *also* be associated with a modular form mod ℓ of weight 2 and level 2, in the way predicted by Serre's conjecture. This deep result allowed him to reduce Fermat's Last Theorem to the Shimura-Taniyama conjecture.

Wiles' work: proof of the Shimura-Taniyama conjecture: In [W3] Wiles proves the Shimura-Taniyama conjecture for semi-stable elliptic curves, providing the final missing step and proving Fermat's Last Theorem. After more than 350 years, the saga of Fermat's Last theorem has come to a spectacular end.

The relation between Wiles' work and Fermat's Last Theorem has been very well documented (see, for example, [R8], and the references contained therein). Hence this article will focus primarily on the breakthrough of Wiles [W3] and Taylor-Wiles [TW] which leads to the proof of the Shimura-Taniyama conjecture for semi-stable elliptic curves.

From elliptic curves to ℓ-adic representations: Wiles' opening gambit for proving the Shimura-Taniyama conjecture is to view it as part of the more general problem of relating two-dimensional Galois representations and modular forms. The Shimura-Taniyama conjecture states that if E is an elliptic curve over $\mathbb{Q}$, then E is modular. One of several equivalent definitions of modularity is that for some integer N there is an eigenform $f = \sum a_n q^n$ of weight two on $\Gamma_0(N)$ such that

$$\#E(\mathbb{F}_p) = p + 1 - a_p$$

for all but finitely primes p. (By an eigenform, here we mean a cusp form which is a normalized eigenform for the Hecke operators; see section 1 for definitions.)

This conjecture acquires a more Galois theoretic flavour when one considers the two dimensional ℓ-adic representation

$$\rho_{E,\ell} : G_\mathbb{Q} \longrightarrow GL_2(\mathbb{Z}_\ell)$$

obtained from the action of $G_\mathbb{Q}$ on the ℓ-adic Tate module of E: $T_\ell E = \varprojlim E[l^n](\bar{\mathbb{Q}})$. An ℓ-adic representation ρ of $G_\mathbb{Q}$ is said to arise from an eigenform $f = \sum a_n q^n$ with integer coefficients a_n if

$$\operatorname{tr}(\rho(\operatorname{Frob}_p)) = a_p,$$

for all but finitely many primes p at which ρ is unramified. Here Frob_p is a Frobenius element at p (see section 2), and its image under ρ is a well-defined conjugacy class.

A direct computation shows that $\#E(\mathbb{F}_p) = p + 1 - \operatorname{tr}(\rho_{E,\ell}(\operatorname{Frob}_p))$ for all primes p at which $\rho_{E,\ell}$ is unramified, so that E is modular (in the sense defined above) if and only if for some ℓ, $\rho_{E,\ell}$ arises from an eigenform. In fact the Shimura-Taniyama conjecture can be generalized to a conjecture that every ℓ-adic representation, satisfying suitable local conditions, arises from a modular form. Such a conjecture was proposed by Fontaine and Mazur [FM].

Galois groups and modular forms

Viewed in this way, the Shimura-Taniyama conjecture becomes part of a much larger picture: the emerging, partly conjectural and partly proven correspondence between certain modular forms and two dimensional representations of $G_{\mathbb{Q}}$. This correspondence, which encompasses the Serre conjectures, the Fontaine-Mazur conjecture, and the Langlands program for GL_2, represents a first step toward a higher dimensional, non-abelian generalization of class field theory.

Two-dimensional representations of $G_{\mathbb{Q}}$: In the first part of this century, class field theory gave a complete description of $G_{\mathbb{Q}}^{\mathrm{ab}}$, the maximal (continuous) abelian quotient of $G_{\mathbb{Q}}$. In fact the Kronecker-Weber theorem asserts that $G_{\mathbb{Q}}^{\mathrm{ab}} \cong \prod_p \mathbb{Z}_p^\times$, and one obtains a complete description of all one-dimensional representations of $G_{\mathbb{Q}}$. In the second half of this century much attention has focused on attempts to understand the whole group $G_{\mathbb{Q}}$, or more precisely to describe all its representations. Although there has been a fair degree of success in using modular forms to construct representations of $G_{\mathbb{Q}}$, less is known about how exhaustive these constructions are. The major results in the latter direction along these lines are the work of Langlands [Ll2] and the recent work of Wiles ([W3] completed by [TW]). Both concern two-dimensional representations of $G_{\mathbb{Q}}$ and give significant evidence that these representations are parametrised (in a very precise sense) by certain modular forms. The purpose of this article is to describe both the proven and conjectural parts of this theory, give a fairly detailed exposition of Wiles' recent contribution and explain the application to Fermat's Last theorem. To make this description somewhat more precise let us distinguish three types of representation.

Artin representations and the Langlands-Tunnell theorem: Continuous representations $\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{C})$ are called (two-dimensional) Artin representations. Such representations necessarily have finite image, and are therefore semi-simple. We restrict our attention to those which are irreducible. They are conjectured to be in bijection (in a precise way) with certain newforms (a special class of eigenforms). Those ρ which are odd (i.e. the determinant of complex conjugation is -1), should correspond to weight 1 holomorphic newforms. Those which are even should correspond to certain non-holomorphic (Maass) newforms. Two partial but deep results are known.

(a) (Deligne-Serre) If f is a holomorphic weight one newform then the corresponding Artin representation can be constructed ([DS]).

(b) (Langlands-Tunnell) If ρ is a two dimensional Artin representation with soluble image then the corresponding modular form exists ([Ll2] and [Tu]).

The proof of the latter result is analytic in nature, invoking the trace formula and the theory of L-functions.

ℓ-adic representations and the Fontaine-Mazur conjecture: By an ℓ-adic representation we shall mean any continuous representation $\rho : G_{\mathbb{Q}} \to GL_2(K)$ which is unramified outside a finite set of primes and where K is a finite extension of $\mathbb{Q}_\ell$ (generalizing slightly the notion of ℓ-adic representation that was introduced before). Given a holomorphic newform f one can attach to f a system of ℓ-adic representations, following Eichler, Shimura, Deligne and Serre. These ℓ-adic representations are called modular. The Fontaine-Mazur conjecture (see [FM]) predicts if ρ is an odd, irreducible, ℓ-adic representation whose restriction to the decomposition group at ℓ is well enough behaved, then ρ is modular. (The restriction on the behaviour of the representation on the decomposition group at ℓ is essential in this conjecture; it is not true that all odd, irreducible two dimensional ℓ-adic representation are modular.) Before Wiles' work almost nothing was known about this conjecture, except that certain very special cases could be deduced from the work of Hecke, Langlands and Tunnell.

Mod ℓ representations and Serre's conjecture: A mod ℓ representation is a continuous representation $\bar{\rho} : G_{\mathbb{Q}} \longrightarrow GL_2(\bar{\mathbb{F}}_\ell)$. For example if $E/\mathbb{Q}$ is an elliptic curve then the action of $G_{\mathbb{Q}}$ on the ℓ-division points of E gives rise to a mod ℓ representation $\bar{\rho}_{E,\ell}$ which is just the reduction modulo ℓ of $\rho_{E,\ell}$. One can use the work of Eichler, Shimura, Deligne and Serre to associate to each mod ℓ eigenform a mod ℓ representation of $G_{\mathbb{Q}}$. The mod ℓ representations which arise in this way are called modular. Serre has conjectured [Se7] that every odd (absolutely) irreducible mod ℓ representation is modular and should arise from a mod ℓ eigenform with certain very specific properties. This conjecture can be thought of as having two parts.

The first asserts that every odd irreducible mod ℓ representation is modular. About this very little is known. It is known for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_2)$ by work of Hecke. It is also known for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_3)$. This latter result is an application of the Langlands-Tunnell theorem using the two accidents that there is a section to the homomorphism $GL_2(\mathbb{Z}[\sqrt{-2}]) \twoheadrightarrow GL_2(\mathbb{F}_3)$ and that $GL_2(\mathbb{F}_3)$ is soluble. Partial results for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_5)$ follow from Wiles' work.

Given a mod ℓ representation arising from a mod ℓ eigenform, the second part of Serre's conjecture predicts the minimal weight and level for that mod ℓ eigenform. Here the situation is much better. There has been a lot of work over the last decade (including ideas from Mazur, Ribet, Carayol and Gross) and the problem is nearly completely resolved (see [Di1]). As was pointed out earlier, Ribet's contribution [R5] implies that, if one can show that the Galois representation $\bar{\rho}_{E,\ell}$ arising from the (semi-stable) Frey curve attached to a solution of Fermat's equation with exponent ℓ is modular, then one can show that this representation does not exist—because it would be modular of weight two and level two— and hence one can deduce Fermat's Last Theorem.

However we have seen that to show $\bar{\rho}_{E,\ell}$ is modular it suffices to show that for some ℓ_0, the ℓ_0-adic representation ρ_{E,ℓ_0} is modular. In particular it suffices to verify that either $\rho_{E,3}$ or $\rho_{E,5}$ is modular. Hence the Shimura-Taniyama conjecture can be reduced to (part of) the Fontaine-Mazur conjecture for $\ell = 3$

and 5. We have seen that for these primes part of Serre's conjecture is known, so it turns out it suffices to prove results of the form "Serre's conjecture for ℓ implies the Fontaine-Mazur conjecture for ℓ". This is the direction of Wiles' work, although nothing quite this general has been proven yet.

Deformation theory: Thus the problem Wiles faces is to show that if ρ is an odd ℓ-adic representation which has irreducible modular reduction $\bar{\rho}$ and which is sufficiently well behaved when restricted to the decomposition group at ℓ, then ρ is modular. In fact he only proves a weakened version of such a result, but one which is sufficient to conclude that all semistable elliptic curves are modular.

Wiles approaches the problem by putting it in a more general setting. On the one hand he considers lifts of $\bar{\rho}$ to representations over complete noetherian local $\mathbb{Z}_\ell$-algebras R. For each finite set of primes Σ, one can consider lifts of type Σ; these are lifts which are well-behaved on a decomposition group at ℓ, and whose ramification at primes not in Σ is rather restricted. In particular, such a lift is unramified outside $\Sigma \cup S$ where S is the set of ramified primes of $\bar{\rho}$. A method of Mazur (see [Maz3]) can then be used to show that if $\bar{\rho}$ is absolutely irreducible, then there is a representation

$$\rho_\Sigma^{\mathrm{univ}} : G_\mathbb{Q} \longrightarrow GL_2(R_\Sigma)$$

which is universal in the following sense. If $\rho : G_\mathbb{Q} \to GL_2(R)$ is a lift of $\bar{\rho}$ of type Σ, then there is a unique local homomorphism $R_\Sigma \longrightarrow R$ such that ρ is equivalent to the pushforward of $\rho_\Sigma^{\mathrm{univ}}$. Thus the equivalence classes of type Σ lifts to $GL_2(R)$ can be *identified* with $\mathrm{Hom}(R_\Sigma, R)$. The local ring R_Σ is called the *universal deformation ring* for representations of type Σ.

On the other hand Wiles constructs a candidate for a universal modular lifting of type Σ

$$\rho_\Sigma^{\mathrm{mod}} : G_\mathbb{Q} \longrightarrow GL_2(\mathbb{T}_\Sigma).$$

The ring $\mathbb{T}_\Sigma$ is constructed from the algebra of Hecke operators acting on a certain space of modular forms. The universal property of R_Σ gives a map $R_\Sigma \to \mathbb{T}_\Sigma$. The problem thus becomes: to show that this map is an isomorphism[2]. In fact, it can be shown to be a surjection without great difficulty, and the real challenge is to prove injectivity, i.e., to show, in essence, that R_Σ is not larger than $\mathbb{T}_\Sigma$.

By an ingenious piece of commutative algebra, Wiles found a numerical criterion for this map to be an isomorphism, and for the ring $\mathbb{T}_\Sigma$ to be a local complete intersection. This numerical criterion seems to be very close to a special case of the Bloch-Kato conjecture [BK]. Wiles further showed (by combining arguments from Galois cohomology and from the theory of congruences between modular forms) that this numerical criterion was satisfied if the minimal version $\mathbb{T}_\emptyset$ of this Hecke algebra (obtained by taking $\Sigma = \emptyset$, i.e., allowing the least possible amount of ramification in the deformations) was a complete intersection. Finally in [TW] it was proved that $\mathbb{T}_\emptyset$ is a complete intersection.

[2] Maps of this kind were already considered in [Maz3] and [BM], and it is conjectured in [MT] that these maps are isomorphisms in certain cases, though not in exactly the situations considered by Wiles.

Outline of the paper

Chapter 1 recalls some basic notions from the classical theory of elliptic curves and modular forms, such as modular forms and modular curves over $\mathbb{C}$ and $\mathbb{Q}$, Hecke operators and q-expansions, and Eichler-Shimura theory. The Shimura-Taniyama conjecture is stated precisely in section 1.8.

Chapter 2 introduces the basic theory of representations of $G_{\mathbb{Q}}$. We describe Mazur's deformation theory and begin our study of the universal deformation rings using techniques from Galois cohomology and from the theory of finite flat group schemes. We also recall some basic properties of elliptic curves, both to explain Frey's argument precisely and illustrate the uses of ℓ-adic representations.

Chapter 3 explains how to associate Galois representations to modular forms. We then describe what was known and conjectured about associating modular forms to Galois representations before Wiles' work. After introducing the universal modular lifts of certain mod ℓ representations, we give the proof of Wiles' main theorems, taking for granted certain results of a more technical nature that are proved in the last two chapters.

Chapter 4 explains how to prove the necessary results concerning the structure of Hecke algebras: the generalization by Taylor and Wiles of a result of de Shalit, and the generalization by Wiles of a result of Ribet.

Chapter 5 establishes the fundamental results from commutative algebra discovered by Wiles, following modifications of the approach of Wiles and Taylor-Wiles proposed by Faltings and Lenstra.