A fourteenth lecture on Fermat's Last Theorem*

Henri Darmon

September 9, 2007

> I informed you earlier of the death of Fermat. He is still alive, and we no longer fear for his health, even though we had counted him among the dead a short time ago... *Letter of Bernard Medon to Nicholas Heinsius, 1652.*

The title of this lecture alludes to Ribenboim's delightful treatise on Fermat's Last Theorem [Rib1]. Fifteen years after the publication of [Rib1], Andrew Wiles finally succeeded in solving Fermat's 350-year-old conundrum. That same year, perhaps to console himself of Fermat's demise, Ribenboim published a second book, this time on *Catalan's conjecture* that there are no consecutive perfect powers other than 8 and 9. As we have learned at this Congress, Preda Mihailescu has just disposed of this conjecture as well. His breakthrough comes only 8 years after the publication of Ribenboim's book on Catalan's equation.

Such is the magic of Ribenboim's books: the age-old problems which they treat have invariably been solved, in comparatively short order! So it is with some eagerness that we await the publication of Ribenboim's next tome (hoping it will be devoted to the Riemann Hypothesis, or the Birch and Swinnerton-Dyer conjecture...)

This "fourteenth lecture" is meant as a tribute both to Ribenboim and to the spirit of Fermat: the fascination with concrete Diophantine problems, especially those that draw us, seemingly inexorably, to central topics in the subject (cyclotomic fields, elliptic curves, reciprocity laws, modular forms...)

*This is a transcription of the author's Ribenboim Prize lecture given at the CNTA meeting in Montreal in May 2002.

Since Fermat's and Catalan's equations have surrendered their mysteries, let me begin with a question which combines features of both: the so-called *generalised Fermat equation*

$$x^p + y^q = z^r. \tag{1}$$

This equation, in which the exponents in Fermat's equation are allowed to be different, is discussed in [DG], [Kr2], and [Da1]; most of this lecture is merely an *update* of the expository article [Da1]. Its main novelty resides in the connection which is drawn, following [Da3], between hypergeometric abelian varieties and (1), as well as in the point of view which differs, albeit superficially, from that of [Da1].

It is useful to allow the exponents p, q or r to be equal to ∞. In that case the somewhat ad-hoc convention

$$z^\infty = \begin{cases} 1 & \text{if } z = 1; \\ \infty & \text{otherwise} \end{cases}$$

is adopted, so that the Catalan equation becomes a special case of (1) with $r = \infty$. The usual convention that $1/\infty = 0$ is adopted as well.

In studying (1), it is important to consider only the so-called *primitive* integer solutions, satisfying

$$\gcd(x, y, z) = 1.$$

This is because non-primitive solutions are often not hard to generate and are therefore less interesting. For example, if a and b are arbitrary integers, the exponent n is odd, and $c = a^n + b^n$, then $(ac, bc, c^{\frac{n+1}{2}})$ is a parametrised family of non-primitive solution to the generalised Fermat equation with exponents $(n, n, 2)$. When considering only primitive solutions, the story is quite different, as the following conjecture suggests.

Conjecture 1 *When $1/p + 1/q + 1/r < 1$, the generalized Fermat equation has no non-trivial primitive solutions except for*

$$1^p + 2^3 = 3^2, \quad 2^5 + 7^2 = 3^4, \quad 7^3 + 13^2 = 2^9,$$
$$2^7 + 17^3 = 71^2, \quad 3^5 + 11^4 = 122^2,$$
$$17^7 + 76271^3 = 21063928^2, \quad 1414^3 + 2213459^2 = 65^7,$$
$$9262^3 + 15312283^2 = 113^7, \quad 43^8 + 96222^3 = 30042907^2$$
$$33^8 + 1549034^2 = 15613^3.$$

Andrew Granville (a student of Ribenboim, who also holds the first Ribenboim prize) and the author proved the following theorem in the Spring of 1993, shortly before Wiles' momentous announcement in June of that year.

Theorem 2 *Suppose (p, q, r) are fixed and satisfy $1/p+1/q+1/r < 1$. Then the equation $x^p + y^q = z^r$ has finitely many primitive integer solutions.*

Proof: The proof (which is also presented, in slightly different forms, in [DG], [Kr2], and [Da1]) relies on the properties of the *Hecke Triangle Group* defined by the abstract presentation:

$$\Gamma_{p,q,r} := \langle \gamma_0, \gamma_1, \gamma_\infty \quad | \quad \gamma_0^p = \gamma_1^q = \gamma_\infty^r = \gamma_0\gamma_1\gamma_\infty = 1 \rangle. \tag{2}$$

(If one of the exponents p, q or r is equal to ∞, the corresponding γ_j is taken to be of infinite order.)

The group $\Gamma_{p,q,r}$ is finite if and only if $1/p+1/q+1/r > 1$, and is *infinite and non-abelian* if and only if $1/p + 1/q + 1/r < 1$. The reason for this is suggested by the "triangle group" terminology: if τ_0, τ_1, and τ_∞ denote the reflections about the sides of a triangle $\Delta_{p,q,r}$ with opposite angles π/p, π/q and π/r, then the rotations

$$\sigma_0 = \tau_1\tau_\infty, \quad \sigma_1 = \tau_\infty\tau_0, \quad \sigma_\infty = \tau_0\tau_1$$

satisfy precisely the same relations as the γ_j in (2). Identifying γ_j with σ_j realises $\Gamma_{p,q,r}$ as a group of isometries of the underlying plane geometry, with two copies of $\Delta_{p,q,r}$ as fundamental region. The triangle $\Delta_{p,q,r}$ can be drawn on the sphere if $1/p + 1/q + 1/r > 1$, on the usual Euclidean plane if $1/p + 1/q + 1/r = 1$, and on the Poincaré upper half plane

$$\mathcal{H} = \{z \in \mathbb{C} | \mathrm{Im}(z) > 0\}$$

if $1/p + 1/q + 1/r < 1$. In the latter case one obtains a model for $\Gamma_{p,q,r}$ as a discrete group acting by isometries on $\mathcal{H}$ with respect to the hyperbolic metric, and hence an embedding $\Gamma_{p,q,r} \subset \mathbf{PSL}_2(\mathbb{R})$. It is an instructive exercise to show that this embedding (by a judicious choice of $\Delta_{p,q,r}$) can be conjugated into an embedding

$$\Gamma_{p,q,r} \longrightarrow \mathbf{PSL}_2(\mathcal{O}), \tag{3}$$

where $\mathcal{O}$ is the ring of integers of a number field. (More precisely: the real subfield of the field of pqr-th roots of unity.) In particular it follows that

$\Gamma_{p,q,r}$ has plenty of finite non-abelian quotients (obtained by reducing the embedding of (3) modulo an ideal of $\mathcal{O}$).

By definition, the group $\Gamma_{p,q,r}$ is a quotient of the group $\Gamma_{\infty,\infty,\infty}$, which can itself be identified with the fundamental group of $\mathbb{P}_1 - \{0, 1, \infty\}$, by assigning to γ_0, γ_1 and γ_∞ certain well-chosen homotopy classes of loops around 0, 1 and ∞. By composing with the natural surjective homomorphism

$$\Gamma_{\infty,\infty,\infty} \longrightarrow \Gamma_{p,q,r},$$

any surjective homomorphism $\varphi : \Gamma_{p,q,r} \longrightarrow G$ gives a surjection

$$\pi_1(\mathbb{P}_1 - \{0, 1, \infty\}) \longrightarrow G,$$

which corresponds in the usual way to a ramified Galois covering of topological spaces with Galois group G:

$$\pi : X \longrightarrow \mathbb{P}_1. \tag{4}$$

This covering is of "signature (p, q, r)" in the sense of [Se], sec. 6.4: it is unramified over $\mathbb{P}_1 - \{0, 1, \infty\}$, and has ramification indices dividing p, q, and r over 0, 1, and ∞ respectively.

The finite covering X inherits through π a natural structure of a *compact Riemann surface*. By Riemann's existence theorem, X arises from an algebraic curve defined over $\mathbb{C}$. Such a curve can of course be defined over a finitely generated extension of $\mathbb{Q}$, and hence, by specialisation, over some number field K. We will now view X in this way, so that π becomes a rational function defined over K.

Let $\Pi_{p,q,r}$ be the set of primitive solutions to $x^p + y^q = z^r$, and let

$$\Sigma_{p,q,r} = \{a^p/c^r \text{ such that } (a, b, c) \in \Pi_{p,q,r}\} \subset \mathbb{P}_1(\mathbb{Q}).$$

To prove Theorem 2 it is enough to show that $\Sigma_{p,q,r}$ is finite. For $t \in \mathbb{P}_1(\mathbb{Q})$, let $L_t = K(\pi^{-1}(t))$ be the field generated over K by the coordinates of $\pi^{-1}(t)$.

The following proposition, an extension of the Chevalley-Weil theorem for ramified coverings, gives some control over the primes of ramification of L_t:

Proposition 3 *There exists a finite set S of primes of K (depending on π, but not on $t \in \Sigma_{p,q,r}$) such that L_t is unramified outside of S, for all $t \in \Sigma_{p,q,r}$.*

Sketch of Proof. Let S be the finite set of primes at which (4) has bad reduction and which divide pqr. For each place v of K which is not in S, write K_v and $\mathcal{O}_v$ for the completion of K at v and its ring of integers respectively. The main point is that, since (a, b, c) is a *primitive* solution,

1. The numerator of $t = a^p/c^r$ is a pth power in $K_v^\times/\mathcal{O}_v^\times$;
2. The numerator of $t - 1 = -b^q/c^r$ is a qth power in $K_v^\times/\mathcal{O}_v^\times$;
3. The denominator of $t = a^p/c^r$ is an rth power in $K_v^\times/\mathcal{O}_v^\times$.

The result follows from the fact that X has ramification indices p, q, r over $0, 1, \infty$ by a variant of the Chevalley-Weil theorem [Be] for ramified coverings.

To control the possible L_t's that could arise, we invoke the following fundamental result of Hermite-Minkowski.

Proposition 4 *There exists only finitely many extensions of K of degree $\leq d$ which are unramified outside S.*

Hence the fields L_t, as t ranges over $\Sigma_{p,q,r}$, are finite in number. In particular, the compositum:

$$L = \cup_{t \in \Sigma_{p,q,r}} L_t$$

is a *finite* extension of K. Note that

$$\Sigma_{p,q,r} \text{ is contained in } \pi(X(L)), \tag{5}$$

so that Theorem 2 is reduced to a question about the finiteness of algebraic points on the curve X over a single number field L.

Proposition 5 *The curve X has genus strictly greater than* 1.

Proof: It is a consequence of the Riemann-Hurwitz formula that the Euler characteristic $2 - 2g(X)$ of X is equal to

$$\chi(X) = \deg(\pi)\left(\frac{1}{p} + \frac{1}{q} + \frac{1}{r} - 1\right).$$

Note that it is here that the assumption on (p, q, r) made in the statement of Theorem 2 enters into the proof.

To complete the proof of Theorem 2, it remains to roll out the big ammunition: Falting's theorem, which asserts that $X(L)$ is finite since X is a curve of genus > 1 and L is a number field. Theorem 2 now follows from (5).

Remarks.

1. The reduction of Theorem 2 to Faltings' theorem is similar to an approach of Elkies used to show that the *abc* conjecture implies the Mordell conjecture [Elk]. The proof of Theorem 2 simply reverses the direction of Elkies' argument, using the Mordell conjecture to deduce a very special case of the *abc* conjecture.

2. The proof sketched above is also explained in [DG] and in [Da1]. Its main virtue lies in the fact that it suggests a general approach to equation (1) (and, in particular, to Fermat's equation). It is summmarised in the following loosely stated principle, which is in some sense the main thesis of this lecture:

Principle 6 *There is a dictionary between the distinct strategies for studying* $x^p + y^q = z^r$ *and the finite quotients of the Hecke triangle group* $\Gamma_{p,q,r}$.

To be more precise, let G be a finite quotient of $\Gamma_{p,q,r}$ and let φ be the natural surjective homomorphism. The attendant strategy for gaining insights into (1) may be described as follows.

1. The geometric part: Study the corresponding G-covering

$$\pi : X \longrightarrow \mathbb{P}_1.$$

   (Eg: Find an equation – or, beter yet, a moduli interpretation – for X, a "minimal" field of definition K, the set S of primes of K where π has bad reduction, etc.)

2. The arithmetic part: determine a finite extension L of K satisfying (5).

3. The Diophantine part: understand the collection $X(L)$ as precisely as possible.

As the reader might expect, carrying out this strategy represents a fairly tall order, but this breakdown provides a conceptual template with which to organise the various approaches that have been followed over the years to

study equation (1). To illustrate this point, we begin with a discussion of the "classical" Fermat equation

$$x^p + y^p = z^p, \quad p \text{ an odd prime.} \tag{6}$$

In increasing order of sophistication, one might approach (6) by exploiting:

I. Abelian quotients of $\Gamma_{p,p,p}$. Note that $\Gamma_{p,p,p}^{\mathrm{ab}} = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. If we take for φ the natural map from $\Gamma_{p,p,p}$ to its abelianisation:

$$\varphi : \Gamma_{p,p,p} \longrightarrow \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z},$$

the corresponding curve X is the Fermat curve itself, with equation

$$F_p : u^p + v^p = w^p, \quad \text{where } \pi(u,v,w) = (u/w)^p.$$

Any primitive solution (a,b,c) to (6) gives a rational point on F_p, for which $\pi(a,b,c) = a^p/c^p$ belongs to $\Sigma_{p,q,r}$. Therefore the extension L can be chosen to be $\mathbb{Q}$. The study of (6) has thus been "reduced", tautologically, to the study of rational points on the Fermat curve. Note here that the arithmetic step is trivial; the diffficulty is concentrated in the study of the Diophantine equation (6) which has not yet succumbed to a "direct" attack.

II. Solvable quotients of $\Gamma_{p,p,p}$. Just beyond the abelian quotients lie the solvable ones. Suppose for example that $\varphi : \Gamma_{p,p,p} \longrightarrow G$ has solvable image, and that $G/[G,G] = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. In that case X is an unramified covering of the Fermat curve. This probably accounts for the key role played by the study of the Jacobians of Fermat curves in certain approaches to (6). (Cf. for example [Mc].)

If one assumes further that G is a two-step nilpotent p-group, then equations for π can be chosen in such a way that $\mathbb{Q}(\pi^{-1}(a^p/c^p))$ is an extension of $\mathbb{Q}(\zeta_p)$ which is unramified outside p. This approach leads naturally to a careful study of unramified abelian extensions of $\mathbb{Q}(\zeta_p)$. A highlight of this circle of ideas is the celebrated theorem of Kummer on Fermat's equation.

Theorem 7 (Kummer) *If there is a non-trivial solution to* $x^p + y^p = z^p$, *then there is an unramified cyclic extension of* $\mathbb{Q}(\zeta_p)$ *of degree* p.

It can be seen that some of the difficulty has been *transferred* from the Diophantine to the arithmetic part of the argument. Kummer's theorem opens

a new universe of deep questions about cyclotomic fields, class numbers, special values of zeta functions, Bernouilli numbers, congruences... all of which are very nicely documented in Ribenboim's "13 lectures".

While extremely fruitful both in proving various cases of (6) and in spurring the further study of cyclotomic fields, Kummer's approach ultimately runs into the following difficulties:

1. Sometimes, an unramified cyclic extension of $\mathbb{Q}(\zeta_p)$ of degree p does exist. (p is then called an *irregular prime*.) It is known that there are infinitely many irregular primes, and, although Fermat's Last Theorem is now proved, we have not yet exhausted the wealth of deep questions raised by the possible occurence of such primes.

2. Even more germane to this lecture, the approach based on solvable quotients of $\Gamma_{p,q,r}$ does not extend to equation (1) in general. For example if the exponents p, q, and r in this equation are pairwise coprime, then $\Gamma_{p,q,r}$ has no abelian, and hence no solvable, quotients! This reflects the elementary observation that equation (1) in this case cannot be attacked by a factorisation like the one

$$(x+y)(x+e^{2\pi i/p}y)\cdots(x+\epsilon^{2\pi i(p-1)/p}y)=z^p$$

which is the starting point for Kummer's work on Fermat's Last Theorem.

For these two reasons, it becomes appealing to explore strategies based on

III. Non-solvable quotients of $\Gamma_{p,q,r}$. Perhaps the most obvious non-solvable finite quotients are obtained by considering the linear representations of $\Gamma_{p,q,r}$ with coefficients, say, in a finite field F. The first non-abelian examples of such representations are, of course, the two-dimensional ones. One gains some flexibility by considering *projective* representations

$$\Gamma_{p,q,r}\longrightarrow\mathbf{PGL}_2(F),$$

which amounts to studying the linear representations of the central extension

$$\begin{aligned}\tilde{\Gamma}_{p,q,r} \;:=\; \langle\gamma_0,\gamma_1,\gamma_\infty \;\;|\\ \gamma_0^p=\gamma_1^q=\gamma_\infty^r=1,\quad \gamma_0\gamma_1\gamma_\infty=-1\rangle,\end{aligned}$$

(where -1 denotes a central involution).

Definition 8 *A* Frey representation *attached to $x^p+y^q=z^r$ with coefficients in F is a representation*

$$\varphi : \tilde{\Gamma}_{p,q,r} \longrightarrow \mathbf{GL}_2(F).$$

Two Frey representations are identified if one can be obtained from the other by extending the field F of scalars and conjugating. A Frey representation φ is said to be *even* if $\varphi(-1)=1$, and *odd* if $\varphi(-1)=-1$.

With these concepts in hand, let us now return to discussing the classical Fermat equation of prime exponent p. In this context it is most natural perhaps to study Frey representations with coefficients in a field of characteristic p. The following theorem (attributed, somewhat anachronistically, to Hecke) indicates that in pursuing this line of enquiry one's options become very restricted. This is actually an encouraging sign, since experience indicates that an approach to a problem which is canonical and allows for little "wriggle room" is often more fruitful and deserving of study.

Proposition 9 *There is a unique irreducible Frey representation attached to $x^p+y^p=z^p$ with coefficients in characteristic p. This representation is odd.*

We refer the reader to Theorem 1.5 of [Da3] for a proof of this purely group-theoretic statement.

To construct the Frey representation of Proposition 9, the generators γ_0 and γ_1 can be sent to the matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}.$$

It is hardly surprising that the corresponding covering

$$\pi : X \longrightarrow \mathbb{P}_1 - \{0, 1, \infty\},$$

being essentially unique, should correspond to a well-studied mathematical object. This is indeed the case: the curve X can be realised as the covering of modular curves

$$\pi : X(2p) \longrightarrow X(2),$$

where $X(n)$ is the modular curve attached to the full congruence subroup of level n. An equation for a universal elliptic curve over $X(2)$ is provided by the Legendre family

$$y^2 = x(x-1)(x-t).$$

With this choice of π, it can be seen that the field $\mathbb{Q}(\pi^{-1}(a^p/c^p))$ is closely related to the field of definition of the points of order p of the Frey curve

$$E_{a,b,c} : y^2 = x(x - a^p)(x + b^p).$$

This chain of reasoning has led us to recover Frey's proposed strategy for studying Fermat's Last Theorem.

Let

$$\rho_{a,b,c} : G_{\mathbb{Q}} \longrightarrow \mathbf{GL}_2(\mathbb{F}_p)$$

be the Galois representation attached to the p-division points of the elliptic curve $E_{a,b,c}$.

Carrying out the arithmetic step requires a precise understanding of the behaviour of $\rho_{a,b,c}$ at the "bad primes" – more precisely, of its restriction to the inertia groups at these primes. Since so much is known about the local properties of Galois representations attached to elliptic curves, this can be done, and the outcome is summarised in the following proposition, which can be viewed as a non-abelian analogue of Theorem 7 of Kummer.

Theorem 10 (Frey) *If there is a non-trivial solution to $x^p + y^p = z^p$, then there is a Galois representation*

$$\rho_{a,b,c} : G_{\mathbb{Q}} \longrightarrow \mathbf{GL}_2(\mathbb{F}_p)$$

(with cyclotomic determinant) which is
1. Unramified outside $2p$.
2. "Finite" at p.
3. "Semistable" at 2.

We will not go into a definition of the technical terms marked in quotes, referring to Section 2.2. of [DDT] for a more detailed discussion.

We are now faced with the problem of understanding or classifying the representations satisfying the conclusion of Theorem 10. This leads to considerations involving *non-abelian* class field theory, which, in its modern formulation, attempts to understand non-abelian extensions of $\mathbb{Q}$ (or of number fields), and more specifically, finite-dimensional representations of $G_{\mathbb{Q}}$. The Langlands philosophy predicts that these representations should be related to *modular forms*. And this is precisely what has been proved by Wiles [Wi] for the representation $\rho_{a,b,c}$. More precisely,

Theorem 11 (Wiles) *Suppose that $\rho_{a,b,c}$ is irreducible. Then it is attached to a cusp form of level $N = \prod_{\ell|abc} \ell$.*

The following deep result of Ribet, which makes crucial use of the fact that ρ is unramified outside of 2 and p and is finite at p, then leads to the following conclusion:

Theorem 12 (Ribet) *The representation $\rho_{a,b,c}$ is attached to a cusp form (mod p) of level 2.*

But a direct calculation now reveals that there are no such modular forms! (The modular curve $X_0(2)$ has genus 0 and hence there are no cusp forms of weight 2 on $\Gamma_0(2)$.) Hence:

$$\text{The representation } \rho_{a,b,c} \text{ must be reducible.} \tag{7}$$

The Diophantine step in the argument now consists in showing that a mod p Galois representation arising from the division points of an elliptic curve cannot be reducible, at least if p is large enough. This amounts to studying the rational points on modular curves of the form $X_0(p)$ (or, eventually, in this situation, $X_0(2p)$). In some sense, the difficult argument used to prove (7) has enabled us to *transfer* a Diophantine question about Fermat curves, to a similar one, of ostensibly similar difficulty, about modular curves. Fortunately, something has been gained in this transfer. Indeed we know, thanks to the fundamental work of Mazur on the Eisenstein ideal [Ma1], [Ma2], that the modular curve $X_0(p)$ is *more tractable* than the Fermat curve of exponent p. More precisely, one has

Theorem 13 (Mazur) *If $p > 163$, then the curve $X_0(p)$ has no rational points other than the cusps, and hence $\rho_{a,b,c}$ is irreducible.*

This contradiction to (7) proves Fermat's Last Theorem, at least for $p > 163$. (Of course, the small exponents had already been handled a long time ago, and in any case many could be disposed of by running through the above argument more carefully.)

Let us now return to the generalised Fermat equation $x^p + y^q = z^r$. As in the case of the Frey representations attached to $x^p + y^p = z^p$, one has the following

Rigidity principle: The Frey representations attached to a given triple (p, q, r) of exponents, with coefficients in a field $\mathbb{F}$ of a given characteristic, are *essentially* unique.

For precise statements in this direction see Theorems 1.7 and 1.8 of [Da3].

Once again, the Frey representations attached to (1), being essentially unique, should correspond to familiar, previously studied mathematical objects. This is indeed the case. The two-dimensional representations of $\tilde{\Gamma}_{p,q,r}$, correspond to rank two local systems on $\mathbb{P}_1 - \{0, 1, \infty\}$ with prescribed monodromies (of finite orders, p, q, and r) over the three deleted points. Such systems have been well-studied classically (by mathematicians like Gauss, Riemann, and Poincaré), and are intimately connected to hypergeometric differential equations.

The Frey representations attached to (1) can be realised on the torsion of certain *hypergeometric abelian varieties*, so-called because their periods are expressed in terms of values of classical hypergeometric functions at rational arguments. A fairly general discussion of this point is given in Section 1.3 of [Da3], and we will content ourselves with focussing on an illustrative special case. Consider from now on the equation

$$x^p + y^p = z^r, \tag{8}$$

where p and r are odd primes, and let $K = \mathbb{Q}(\zeta_r)^+$. We take the point of view that r is fixed and that p is allowed to vary, corresponding to the Diophantine problem of understanding whether an rth power can be expresseed as a sum of two relatively prime perfect powers.

Proposition 14 *There exists two abelian varieties J_r^+ and J_r^- over $\mathbb{Q}(t)$ having good reduction outside of $t = 0$, 1, and ∞ such that*

1. $End_K(J_r^\pm) \otimes \mathbb{Q} \simeq K$;

2. $\dim(J_r^\pm) = [K : \mathbb{Q}]$. (The abelian variety $J_r^\pm$ is said to be of "$\mathbf{GL}_2$-type" over K.) For each prime $\mathfrak{p}$ of K, such an abelian variety gives rise to a two-dimensional mod p representation of $G_{K(t)}$,

$$\rho_{\mathfrak{p}}^\pm : G_{K(t)} \longrightarrow \mathbf{GL}_2(\mathbb{F}),$$

where $\mathbb{F}$ is the residue field of K above $\mathfrak{p}$.

*3. For each rational prime p, and $\mathfrak{p}|p$, the representation $\rho_{\mathfrak{p}}^+$ (resp. $\rho_{\mathfrak{p}}^-$) is the (*unique*, up to twist, and automorphisms of $\mathbb{F}$) even (resp. odd) Frey representation attached to $x^p + y^p = z^r$.*

The hypergeometric abelian varieties of proposition 14 play the same role in the study of equation (8) as the elliptic curves in the Legendre family in the study of equation (6).

Here is what happens in carrying out the main steps in the study of (1), following our earlier sketch of the proof of Fermat's Last Theorem. (For more details, see [Da3].)

To carry out the *arithmetic step*, one is led to consider the problem of establishing the modularity (in a suitable sense, that is made precise in [Da3]) of hypergeometric abelian varieties: this amounts to showing that $J_r^{\pm}(t)$, for $t \in \mathbb{Q}$, say, is modular, in the sense that its ℓ-adic representations correspond to a *Hilbert modular form* over the totally real field K.

Let $\mathfrak{r}$ be the unique prime of K above r. Then the $\mathfrak{r}$-torsion subgroup $J_r^{\pm}(t)[\mathfrak{r}]$ is a two-dimensional $\mathbb{F}_r$-vector space on which G_K acts linearly. It can be shown that this action *extends* to an action of $G_{\mathbb{Q}}$, and that

Theorem 15 *1. The $G_{\mathbb{Q}}$-representation $J_r^{+}(t)[\mathfrak{r}]$ is reducible.*

2. The representation $J_r^{-}(t)[\mathfrak{r}]$ is isomorphic to (a twist of) the representation $E[r]$, where

$$E : y^2 = x(x-1)(x-t).$$

Sketch of Proof: By Proposition 14, the $\mathbb{F}_r[G_{\mathbb{Q}}]$-modules $J_r^{+}[\mathfrak{r}]$ and $J_r^{-}[\mathfrak{r}]$ are even and odd Frey representations attached to the Fermat equation $x^r + y^r = z^r$. Hence Proposition 9 implies that the representation attached to $J_r^{+}[\mathfrak{r}]$ is reducible (since there are no even irreducible Frey representations attached to the classical Fermat equation) and that the $G_{\mathbb{Q}(t)}$-module $J_r^{-}[\mathfrak{r}]$ coincides (up to a twist) with the r-torsion points of the Frey curve.

Applying the modularity result of [BCDT], we then have:

Corollary 16 *The mod r representation $J_r^{\pm}(t)[\mathfrak{r}]$ is modular, i.e., corresponds to a classical modular form over $\mathbb{Q}$, mod r.*

Now applying the technique of base change over an abelian extension of $\mathbb{Q}$, one may conclude that $J_r^{\pm}(t)[\mathfrak{r}]$ corresponds to a Hilbert modular form over K, mod r. Now we are in a good situation to apply the lifting theorems of Wiles, which assert, roughly, that an r-adic representation is modular if the corresponding mod r representation is. (So Wiles' modularity theorem would play the same role, in establishing the modularity of hypergeometric abelian varieties, as the Langlands-Tunnell theorem in Wiles' original argument!) Unfortunately the lifting statement of Wiles and some of its

subsequent generalisations (cf. for example [SW1] and [SW2]) come with a number of technical restrictions which preclude us from deducing the modularity of all the hypergeometric abelian varieties involved in the argument. (See [Da2] for a more detailed discussion.) But note how this strategy for proving the modularity of all hypergeometric abelian varieties emerges naturally from the study of the generalised Fermat equation (and is arguably the most valuable insight to have emerged so far from the author's study of this equation).

When $r = 2$ and 3, the hypergeometric abelian varieties $J_r^{\pm}$ correspond to families of elliptic curves (which include the (uni)versal families over $X_0(2)$ and $X_0(3)$ respectively). This fact allowed Merel and the author [DM] to prove the following theorem, generalising classical results of Fermat and Euler in the case of $n = 4$ and 3 respectively.

Theorem 17 *The equations $x^n + y^n = z^2$ $(n \geq 4)$ and $x^n + y^n = z^3$ $(n \geq 3)$ have no non-trivial primitive solutions.*

Remark: The proofs for small n are handled by Poonen [Po] by classical descent techniques.

In trying to exend the analysis to larger values of r, one runs into considerable difficulties of a more than technical nature. Why? The proof of Theorem 17 exploits the isomorphisms

$$\Gamma_{\infty,\infty,2} = \Gamma_0(2); \quad \Gamma_{\infty,\infty,3} = \Gamma_0(3).$$

When $r \geq 5$, by constrast, the group $\Gamma_{\infty,\infty,r}$ is *non-arithmetic*, and it then seems like a very difficult problem to extend Mazur's Diophantine results on modular curves to the ramified coverings of $\mathbb{P}_1$ obtained by considering the torsion points of hypergeometric abelian varieties over $\mathbb{Q}(\zeta_r)^+(t)$. One of the difficulties of the non-arithmetic case resides in the fact that the corresponding coverings have very few correspondences and there is no analogue of the ring of Hecke operators which plays such an important role in the arguments of [Ma1] and [Ma2].

The difficulty of the general question has not prevented various mathematicians from proving a number of partial results which would fill a handsome chapter in any sequel to Ribenboim's "13 Lectures". For example, Alain Kraus also proved the following result about the equation $x^3 + y^3 = z^p$ by using modular form methods [Kr1]:

Theorem 18 (Kraus,1998) *The equation*

$$x^3 + y^3 = z^p$$

has no non-trivial primitive solution for $17 \leq p \leq 10,000$.

He does this by establishing criteria (strikingly reminiscent of the criteria of Sophie Germain and Legendre for disposing the first case of Fermat's Last Theorem) which guarantee that the equation $x^3 + y^3 = z^p$ has no non-trivial primitive solution. These criteria are presumably satisfied for all p but were checked numerically in the range given by the theorem.

There is also the following theorem of Jordan Ellenberg concerning the equation $x^4 + y^2 = z^p$:

Theorem 19 (Ellenberg, 1999) *The equation*

$$x^4 + y^2 = z^p$$

has no non-trivial primitive solutions, if $p > 211$.

The proof uses some new modularity results on $\mathbb{Q}$-curves obtained in collaboration with Chris Skinner [ES].

Although it falls somewhat outside the scope of these lectures since it concerns the equation

$$Ax^p + By^p = Cz^p, \tag{9}$$

I cannot resist mentionning the following gem which would have to figure prominently in any sequel to [Rib1] concerned with (9)

Theorem 20 (Halberstadt, Kraus) *Suppose A, B and C are odd and relatively prime. There exists a set Π of primes of positive density (depending, in an effectively computable way, on A, B, C) such that, for all $p \in \Pi$, the equation*

$$Ax^p + By^p = Cz^p$$

has no non-trivial solution.

The proof of Halberstadt and Kraus, an application of their "symplectic method", is based on a careful study of the module of p-division points of the Frey curve endowed with its natural symplectic structure arising from the Weil pairing. Ultimately, it relies crucially on the fact that the Frey curve

$$y^2 = x(x - Au^p)(x + Cw^p)$$

has a discriminant of the form

$$\Delta = 2^{-8}(ABC)^2(uvw)^{2p}$$

in which the exponent of 2 that appears is negative!

Results have also been obtained for the more general equation

$$Ax^p + By^q = Cz^r \tag{10}$$

which blends the features of (1) and (9). In this context let us mention the work of Bennett and Skinner [BS] and Ivorra [Iv] on the equation $x^n + y^n = Cz^2$, and of Kraus on the equations $x^m - y^m = Rz^n$ [Kr3].

In conclusion, the generalized Fermat equation seems to share with its classical couterpart the power of generating important mathematical questions. Aside from two key special cases – Fermat's Last Theorem, and Catalan's conjecture – and the fragmentary results in Theorem 17 [DM], Theorem 18 [Kr1], and Theorem 19 [Ell] mentioned above, almost everything about it remains unknown. Ribenboim should be delighted by this state of affairs!

References

[BCDT] Christophe Breuil, Brian Conrad, Fred Diamond and Richard Taylor. *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*. J. Amer. Math. Soc. **14** (2001), no. 4, 843–939.

[Be] Sybilla Beckmann. *Ramified primes in the field of moduli of branched coverings of curves*. J. Algebra **125** (1989), no. 1, 236–255.

[BS] Michael Bennett and Chris Skinner. *Ternary Diophantine equations via Galois representations and modular forms*, Canad. J. Math., to appear.

[Da1] Henri Darmon. *Faltings plus epsilon, Wiles plus epsilon, and the generalized Fermat equation*. C. R. Math. Rep. Acad. Sci. Canada **19** (1997), no. 1, 3–14.

[Da2] Henri Darmon. *Modularity of fibres in rigid local systems*. Ann. of Math. (2) **149** (1999), no. 3, 1079–1086.

[Da3] Henri Darmon. *Rigid local systems, Hilbert modular forms, and Fermat's last theorem*. Duke Math. J. **102** (2000), no. 3, 413–449.

[DDT] Henri Darmon, Fred Diamond, and Richard Taylor. *Fermat's last theorem*. Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), 2–140, Internat. Press, Cambridge, MA, 1997.

[DG] Henri Darmon and Andrew Granville. *On the equations* $z^m = F(x, y)$ *and* $Ax^p + By^q = Cz^r$. Bull. London Math. Soc. **27** (1995), no. 6, 513–543.

[DM] Henri Darmon and Loïc Merel. *Winding quotients and some variants of Fermat's last theorem*. J. Reine Angew. Math. **490** (1997), 81–100.

[Elk] Noam D. Elkies. *ABC implies Mordell*. Internat. Math. Res. Notices 1991, no. 7, 99–109.

[Ell] Jordan S. Ellenberg. *Galois representations attached to* $\mathbb{Q}$*-curves and the generalized Fermat equation* $A^4 + B^2 = C^p$, Amer. J. Math., to appear.

[ES] Jordan S. Ellenberg and Chris Skinner. *On the modularity of* $\mathbb{Q}$*-curves*. Duke Math. J. **109** (2001), no. 1, 97–122.

[HK] Emmanuel Halberstadt et Alain Kraus. *Courbes de Fermat: résultats et problèmes*. J. Reine Angew. Math. **548** (2002), 167–234.

[Iv] Wilfrid Ivorra, Thesis, Université Paris VI, in progress.

[Kr1] Alain Kraus. *Sur l'équation* $a^3 + b^3 = c^p$. Experiment. Math. **7** (1998), no. 1, 1–13.

[Kr2] Alain Kraus. *On the equation* $x^p + y^q = z^r$*: a survey*. Ramanujan J. 3 (1999), no. 3, 315–333.

[Kr3] Alain Kraus. *Une question sur les équations* $x^m - y^m = Rz^n$. Compositio Math. **132** (2002), no. 1, 1–26.

[Ma1] Barry Mazur. *Modular curves and the Eisenstein ideal*. Inst. Hautes Etudes Sci. Publ. Math. No. **47** (1977), 33–186 (1978).

[Ma2] Barry Mazur. *Rational isogenies of prime degree* (with an appendix by D. Goldfeld). Invent. Math. **44** (1978), no. 2, 129–162.

[Mc] William G. McCallum. *The arithmetic of Fermat curves*. Math. Ann. **294** (1992), no. 3, 503–511.

[Mi] Preda Mihailescu. *Primary units and a proof of Catalan's conjecture*, Crelle's Journal, submitted.

[Po] Bjorn Poonen. *Some Diophantine equations of the form* $x^n + y^n = z^m$. Acta Arith. **86** (1998), no. 3, 193–205.

[Rib1] Paulo Ribenboim. Thirteen Lectures on Fermat's Last Theorem. Springer-Verlag, New York-Heidelberg, 1979.

[Rib2] Paulo Ribenboim. Catalan's conjecture. Are 8 and 9 the only consecutive powers? Academic Press, Inc., Boston, MA, 1994.

[Se] Jean-Pierre Serre. Topics in Galois theory. Lecture notes prepared by Henri Damon [Henri Darmon]. With a foreword by Darmon and the author. Research Notes in Mathematics, 1. Jones and Bartlett Publishers, Boston, MA, 1992.

[SW1] Chris Skinner and Andrew Wiles. *Residually reducible representations and modular forms*. Inst. Hautes Etudes Sci. Publ. Math. No. **89** (1999), 5–126.

[SW2] Chris Skinner and Andrew Wiles. *Nearly ordinary deformations of irreducible residual representations*. Ann. Fac. Sci. Toulouse Math. (6) **10** (2001), no. 1, 185–215.

[Wi] Andrew Wiles. *Modular elliptic curves and Fermat's last theorem*. Ann. of Math. (2) **141** (1995), no. 3, 443–551.