Terms and Conditions

J. reine angew. Math. **490** (1997), 81—100

Journal für die reine und angewandte Mathematik

Winding quotients and some variants of Fermat's Last Theorem

By *Henri Darmon* at Montréal and *Loïc Merel* at Paris

Introduction

The motivation (or perhaps the excuse?) for this paper is the study of the following variants of Fermat's equation $x^n + y^n = z^n$:

$$x^n + y^n = 2z^n, \tag{1}$$

$$x^n + y^n = z^2, \tag{2}$$

$$x^n + y^n = z^3. \tag{3}$$

An integer solution (x, y, z) to one of the above equations is called *primitive* if $\gcd(x, y, z) = 1$. The equations (2) and (3) typically have infinitely many non-primitive solutions. For example, if n is odd, and a and b are any two integers with $a^n + b^n = c$, then

$$(ac)^n + (bc)^n = (c^{\frac{n+1}{2}})^2,$$

giving an abundant but rather uninteresting supply of solutions to equation (2). It is natural to restrict ones attention to the primitive solutions, which is what we will do from now on.

Equations (1), (2) and (3) also have certain obvious "trivial" solutions: a solution is called *trivial* if $xyz = 0$ or ± 1, and is called *non-trivial* otherwise.

The work of Hellegouarch, Frey [12], Serre [27], and Ribet [24] relating Fermat's Last Theorem to the Shimura-Taniyama conjecture (and the precise form of this conjecture given by Weil) can also be applied to other ternary equations analogous to Fermat's equation. (See, for example, [12], [19], or sec. 4.3 of [27].) The idea of applying modular form techniques to equations (1), (2) and (3) appeared in [3] for equations (2) and (3), and then in [25] for equation (1). As a consequence of the results in [25] and [3], one has:

Let $n = p > 12$ be prime. Then:

1. *Equation* (1) *has no non-trivial solution if* $n \equiv 1 \pmod 4$.

2. *Equation* (2) *has no non-trivial primitive solution if* $n \equiv 1 \pmod 4$.

3. *Assume that every elliptic curve over* $\mathbb{Q}$ *is modular. Then equation* (3) *has no non-trivial primitive solution if* $n \equiv 1 \pmod 3$.

The goal of this article is to dispose of the case of general n, and, in particular, of primes which are congruent to $-1 \bmod 4$ for equations (1) and (2), and primes which are congruent to $-1 \bmod 3$ for equation (3). Our main result is:

Main Theorem. *Let the exponent n be an arbitrary positive integer.*

1. *The equation* $x^n + y^n = 2z^n$ *has no non-trivial primitive solution when* $n \geqq 3$.

2. *The equation* $x^n + y^n = z^2$ *has no non-trivial primitive solution when* $n \geqq 4$.

3. *Assume that every elliptic curve over* $\mathbb{Q}$ *is modular. Then* $x^n + y^n = z^3$ *has no non-trivial primitive solution when* $n \geqq 3$.

The work of Dénes [7] and Poonen [22] establishes the Main Theorem when the exponent n is a small integer. (See the discussion below.) Hence it is enough to deal with the case where the exponent n is a prime $p \geqq 7$. In essence, the proof of the Main Theorem then follows the same general strategy as in the proof of Fermat's Last Theorem. It relies on a variant of the key idea of Frey and Serre, together with the "lowering the level" result of Ribet which was used to show that the Shimura-Taniyama conjecture implies Fermat's Last Theorem. Enough of the Shimura-Taniyama conjecture is now known, thanks to Wiles' breakthrough in this direction, to make our results on equation (1) and (2) (but not equation (3)) independent of any conjecture.

Central to the study of Fermat's equation is Mazur's theorem that an elliptic curve over $\mathbb{Q}$ cannot have a rational point of order p if $p > 7$. Our work requires an additional result (Theorem 8.1) of a similar nature concerning elliptic curves whose associated mod p Galois representation maps to the normalizer of a nonsplit Cartan subgroup of $\mathbf{GL}_2(\mathbb{F}_p)$. This theorem, whose statement was suggested by Serre's problem (see [26], 4.3), may be of some independent interest. Its proof borrows heavily from the techniques of Mazur [17], [18]. It also exploits (in a manner similar to [20]) a finiteness criterion (predicted by the conjecture of Birch and Swinnerton-Dyer) for the Mordell-Weil groups of modular Jacobians due to Gross-Zagier and Kolyvagin-Logachev.

We now collect a few miscellaneous comments on equations (1), (2), and (3).

The equation $x^n + y^n = 2z^n$. Equation (1) was studied in 1952 by Dénes [7], who conjectured part 1 of our Main Theorem, and proved it for $2 < n < 31$. He was motivated by the problem of finding perfect powers in arithmetic progressions: for if (x, y, z) is a solution to equation (1), then x^n, z^n, y^n forms an arithmetic progression. Part 1 of the Main Theorem implies that there can be no such three-term arithmetic progression when $n \geqq 3$.

The study of equation (1) was later taken up by Ribet in [25], exploiting the link between Fermat's Last Theorem and the Shimura-Taniyama conjecture. Ribet was able to show that equation (1) has no solutions when n is divisible by a prime which is congruent to $1 \bmod 4$, or when xyz is even. In the course of proving the Main Theorem, we will retrace the steps in Ribet's argument.

The reader is referred to [25] for a more thorough historical discussion of equation (1). (See also [4] for a study of the related equation $x^4 - y^4 = z^n$.)

The equation $x^n + y^n = z^2$. When $n \leqq 3$, equation (2) has infinitely many primitive solutions. For $n = 2$, these solutions correspond to primitive pythagorean triples. For $n = 3$, an infinite family of primitive solutions is given by the equations:

$$x = u(u^3 - 8v^3), \quad y = 4v(u^3 + v^3), \quad z = (u^6 + 20u^3v^3 - 8v^6),$$

$$3 \nmid u + v, \quad 2 \nmid u, \quad \gcd(u, v) = 1 .$$

(For the general solution, see for example [5], sec. 7.2.)

The smallest case covered by part 2 of the Main Theorem is the case $n = 4$. This was established by Fermat. Perhaps the best-known application of the method of descent, Fermat's proof involved the elliptic curve $y^2 = x^3 - x$ with complex multiplication by $\mathbb{Q}(i)$.

In [22], Poonen shows that equation (2) has no non-trivial primitive solution when $n = 5$, 6, and 9, using classical descent arguments. Thanks to this work, it is enough to prove part 2 of the Main Theorem when the exponent n is a prime $p \geqq 7$.

The equation $x^n + y^n = z^3$. The equation $x^n + y^n = z^3$ has *infinitely many* primitive solutions if $n = 2$, for example:

$$x = u^3 + 3uv^2, \quad y = v^3 + 3vu^2, \quad z = u^2 + v^2, \quad \gcd(u, v) = 1 .$$

The smallest case covered by part 3 of the Main Theorem is the case $n = 3$. This is just Fermat's Last Theorem with exponent 3, which was proved by Euler by elementary methods. Euler's descent involves the elliptic curve $x^3 + y^3 = 1$ with complex multiplication by $\mathbb{Q}(e^{2\pi i/3})$.

The case $n = 4$ was handled by Lucas (cf. [10], p. 630) and the case $n = 5$ was disposed of by Poonen [22], through a descent on the Jacobian of a curve of genus 3. Hence we are also reduced to proving part 3 of the Main Theorem in the case where the exponent n is a prime $p \geqq 7$.

In this case, our proof of part 3 of the Main Theorem still requires the hypothesis that the elliptic curves involved in the study of the equation $x^n + y^n = z^3$ are modular. This requirement is not a consequence of the results of Wiles, not even of the strengthenings due to Conrad, Diamond, and Taylor, since the conductor of these elliptic curves is divisible by 27. The problem of showing that a cube cannot be expressed as a sum of two

relatively prime nth powers ($n \geqq 3$) gives a Diophantine incentive for proving the entire Shimura-Taniyama conjecture...

Acknowledgements. The authors would like to thank D. Abramovich, I. Chen, B. Edixhoven, and K. Ribet for many useful discussions related to this work.

1. Frey curves

We will assume in this section that the exponent n arising in equations (1), (2), and (3) is a prime $p \geqq 7$.

Let (a, b, c) be a non-trivial primitive solution to equation (1), (2), or (3). Following [25] and [3], we associate to the solution (a, b, c) a *Frey curve* E as follows:

Equation (1). If (a, b, c) is a primitive solution to equation (1), then a and b are odd. By multiplying (a, b, c) by -1 if necessary, assume that $a \equiv -1 \pmod 4$. Let E be the elliptic curve given by the Weierstrass equation:

$$Y^2 = X(X - a^p)(X - 2c^p). \tag{4}$$

Equation (2). If ab is even, we assume without loss of generality that a is even and that $c \equiv 1 \pmod 4$. If ab is odd, we assume without loss of generality that $a \equiv -1 \pmod 4$. This can always be done, by interchanging a and b and replacing c by $-c$ if necessary.

Let E be the elliptic curve given by the Weierstrass equation:

$$Y^2 + XY = X^3 + \frac{(c-1)}{4}X^2 + \frac{a^p}{2^6}X \quad \text{if } ab \text{ is even}, \tag{5}$$

$$Y^2 = X^3 + 2cX^2 + a^pX \quad \text{if } ab \text{ is odd}. \tag{6}$$

Equation (3). Define the elliptic curve E by the following Weierstrass equation, when $c = 2c_0$ is even:

$$Y^2 + b^pY = X^3 - 3(c_0^3 + b^p)c_0X - c_0^3(2c_0^3 - 5b^p). \tag{7}$$

When ab is even, assume without loss of generality that a is odd and b is even, and define E by the equation:

$$Y^2 + cXY = X^3 - c^2X^2 - \frac{3}{2}cb^pX + b^p\left(a^p + \frac{5}{4}b^p\right). \tag{8}$$

Arithmetic invariants of E. Let Δ be the discriminant of the Weierstrass equation defining E. One finds, for equation (1),

$$\Delta = 2^6(abc)^{2p};$$

for equation (2),

$$\Delta = \frac{1}{2^{12}}(a^2b)^p \quad \text{if } 2|ab, \quad \Delta = 2^6(a^2b)^p \quad \text{if } 2\nmid ab;$$

and for equation (3),

$$\Delta = 3^3(a^3b)^p.$$

If M is a positive integer, denote by $\mathrm{rad}(M)$ the product of the primes which divide M.

Proposition 1.1. *Let N be the conductor of the curve E constructed above.*

1. *For equation* (1), $N = \mathrm{rad}(abc)$ *if abc is even, and* $N = 2^5\,\mathrm{rad}(abc)$ *if abc is odd.*

2. *For equation* (2), $N = \mathrm{rad}(ab)$ *if ab is even, and* $N = 2^5\,\mathrm{rad}(ab)$ *if ab is odd.*

3. *For equation* (3), $N = \mathrm{rad}(ab)$ *if* 3 *divides ab, and* $N = 3^3\,\mathrm{rad}(ab)$ *if* 3 *does not divide ab.*

Sketch of Proof. The bad reduction types of the curve E, and the conductor N of E, can be computed using Tate's algorithm [28]. (We also used the results of [9] in the case of equations (1) to analyze the bad fiber at 2 of E; for the case of equations (2) and (3), see also [3], lemmas 2.1 and 3.2.) Note that the condition $\gcd(a, b, c) = 1$ directly implies that the curve E which arises in equations (1) and (2) (resp. equation (3)) has multiplicative reduction at all primes except possibly 2 (resp. 3).

Lemma 1.2. *The curve E constructed above (from a non-trivial primitive solution) has at least one odd prime of multiplicative reduction.*

Proof. For otherwise, Proposition 1.1 implies that abc (resp. ab) is a power of 2 in the case of equation (1) (resp. (2) and (3)). For equation (1), this implies directly that (a, b, c) is a trivial solution, contrary to our assumption. In the case of equation (2) or (3) one can only conclude that (a, b, c) gives rise to a solution to the special case of Catalan's equation: $x^p \pm 1 = z^2$ or $x^p \pm 1 = z^3$ (of a very restricted sort, since x must be a power of 2!). This equation is proved to have no non-trivial solutions aside from $8 + 1 = 9$; cf. [23], (A 6.1) and (A 7.3).

Extra level structures. Crucial to our proof is the fact that the curve E is equipped with some auxiliary level structure, namely, a rational point of order 2 or 3.

Lemma 1.3. *Let E be the Frey curve constructed above.*

1. *In the case of equation* (1), *the curve E has all its points of order* 2 *defined over* $\mathbb{Q}$.

2. *In the case of equation* (2), *the curve E has a point of order* 2 *defined over* $\mathbb{Q}$.

3. *In the case of equation* (3), *the curve E has a rational point (and therefore a rational subgroup) of order* 3.

Proof. Parts 1 and 2 can be seen directly. For part 3, one checks that the point on E given by the (X, Y) coordinates:

$$\left(\frac{3c^2}{4}, \frac{a^p - b^p}{2}\right), \quad (3c^2, 4a^p)$$

in the equations (7) and (8) respectively is of order 3. (We mention in passing a mistake in the proof of lemma 3.1 of [3]. The point of order 3 on the curve denoted $E_{a,b,c}$ has coordinates $(x, y) = (3c^2, 4b)$, and in particular it is a rational point.)

Modularity of E. As in the proof of Fermat's Last Theorem, the modularity of E plays a key role in our analysis.

Theorem 1.4 (Wiles, Taylor, Diamond). *The curve E associated to a solution of equation* (1) *or* (2) *is modular.*

Proof. The curve associated to a solution of equation (1) or (2) is semistable at 3 and 5, and its modularity follows from Diamond's extension [8] of the work of Wiles [30] and Taylor-Wiles [29]. (In the case of equation (1), one does not need the full power of Diamond's extension. Since the curve there is of the form $Y^2 = X(X - A)(X + B)$, its modularity follows more directly from the results of [30] and [29], as is explained in [9]. For more details, see the discussion in [25].)

Remark. For equation (3), even Diamond's results are not enough, since the curve E has additive reduction at 3: this is why the results on equation (3) are still conditional.

2. Galois representations

In this section, as well as in sections 3 and 4, we continue to suppose that $n = p$ is a prime which is $\geqq 7$.

We choose, once for all, an isomorphism of $\mathbb{F}_p$-vector spaces between points of order p of $E(\bar{\mathbb{Q}})$ and $\mathbb{F}_p^2$.

Let

$$\varrho : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathbf{GL}_2(\mathbb{F}_p)$$

be the mod p Galois representation attached to the Frey curve E constructed in the previous section. Let $N(\varrho)$ be the conductor of ϱ in the sense of [27], i.e., with the factor at p removed.

Lemma 2.1. *The conductor $N(\varrho)$ takes on the following values for each of equations* (1), (2), *and* (3).

1. *Equation* (1): $N(\varrho) = 2$ *if abc is even, and* $N(\varrho) \mid 32$ *if abc is odd.*

2. *Equation* (2): $N(\varrho) = 2$ *if ab is even, and* $N(\varrho) \mid 32$ *if ab is odd.*

3. *Equation* (3): $N(\varrho) = 3$ *if* 3 *divides ab, and* $N(\varrho) \mid 27$ *if* 3 *does not divide ab.*

Proof. By Proposition 1.1, the curve E is semistable at all primes except possibly r, where $r = 2$ in the case of equations (1) and (2), and $r = 3$ in the case of equation (3). Furthermore, for all primes $\ell \neq r$, one has $\mathrm{ord}_\ell(\Delta_{\min}) \equiv 0 \pmod{p}$, where $\Delta_{\min}$ is the discriminant of the minimal Weierstrass model for E. The result follows directly. (See for example [27], 4.1.12.)

Theorem 2.2. *The representation ϱ is absolutely irreducible.*

Proof. When $\varrho \geqq 17$, this is a direct consequence of Corollary 4.4 of Mazur [18], since the curve E has at least one odd prime of multiplicative reduction (Lemma 1.2). If $7 \leqq p \leqq 13$ and ϱ is reducible, then the curve E gives rise (by Lemma 1.3) to a rational point on one of the modular curves $X_0(N)$ with $N = 14, 22, 26, 21, 33$, or 39. These rational points are known to be either cuspidal, or CM, and cannot correspond to an elliptic curve with non-integral j-invariant. (This follows from work of Mazur [18], Kubert [16], and Kenku [14] for the case $N = 39$.)

3. Modular forms

From Theorems 2.2 and 1.4, it follows that ϱ is an absolutely irreducible modular representation in the case of equations (1) and (2), and that a similar statement is true for equation (3) if one assumes the Shimura-Taniyama conjecture. We now invoke the "lowering the level" theorem of [24] which implies an important part of Serre's ε-conjecture [27]:

Theorem 3.1 (Ribet). *There is a cusp form $f \bmod p$ of weight 2 and level $N(\varrho)$ which is associated to ϱ in the sense that*

$$a_\ell(f) = \mathrm{trace}\left(\varrho(\mathrm{Frob}_\ell)\right)$$

for all $\ell \nmid pN(\varrho)$.

Theorem 3.1 immediately yields the following corollary which was already established in [25] and [3].

Corollary 3.2. *Suppose that $n \geqq 7$ is prime.*

1. *There are no non-trivial primitive solutions to equation* (1) *with* $2 \mid abc$.

2. *There are no non-trivial primitive solutions to equation* (2) *with* $2 \mid ab$.

3. *Assume the Shimura-Taniyama conjecture. There are no non-trivial primitive solutions to equation* (3) *with* $3 \mid ab$.

Proof. A non-trivial primitive solution of equations (1), (2), or (3) satisfying the hypothesis of Corollary 3.2 would give rise, by Lemma 2.1, to a modular irreducible mod p Galois representation of conductor 2 or 3. By Theorem 3.1, this representation corresponds to a mod p eigenform of weight 2 and level 2 or 3. This is impossible since there are no weight 2 cusp forms of these levels in characteristic p.

4. Complex multiplication

Pushing our investigation further, we now know that ϱ corresponds to a mod p eigenform of weight 2 and level 32 in the case of equations (1) and (2). In the case of equation (3), it corresponds to an eigenform of level 27 provided the curve E is modular.

The curves $X_0(32)$ and $X_0(27)$ are both curves of genus 1 with a rational point, i.e., elliptic curves, and they are given by the equations

$$X_0(32): Y^2 = X^3 - X, \quad X_0(27): Y^2 = X^3 + 16.$$

(Note that these curves are also the Frey curves that are associated to the trivial solution $(1,1,1)$ of equation (1), to the trivial solution $(-1,1,0)$ of equation (2), and to the trivial solution $(-1,1,0)$ of equation (3).) It follows that ϱ is isomorphic to the Galois representation given by the action on the p-division points of the elliptic curve $X_0(32)$, in the case of equation (1) and (2), and of the elliptic curve $X_0(27)$, in the case of equation (3). Both of these curves have complex multiplication: $X_0(32)$ by the ring of Gaussian integers $\mathbb{Z}[i]$, and $X_0(27)$ by the ring $\mathbb{Z}[\zeta_3]$ where ζ_3 is a primitive cube root of unity.

Remark. The reader will note that the curve $Y^2 = X^3 - X$ already appears (in a much more elementary guise!) in Fermat's study of the minimal case $n = 4$ of equation (2). The curve $Y^2 = X^3 + 16$ plays a similar role in Euler's study of the minimal case $n = 3$ of equation (3).

We will now apply the theory of complex multiplication to get a precise understanding of the Galois representation ϱ, and in particular of its image. Let G be the image of ϱ in $\mathbf{GL}_2(\mathbb{F}_p)$.

Proposition 4.1. 1. *In the case of equations* (1) *and* (2), *the group G is the normalizer of a Cartan subgroup of* $\mathbf{GL}_2(\mathbb{F}_p)$. *This Cartan subgroup is split if* $p \equiv 1 \pmod 4$, *and is non-split if* $p \equiv -1 \pmod 4$. *Moreover, the field cut out by ϱ is an abelian extension of* $\mathbb{Q}(i)$.

2. *Assume the Shimura-Taniyama conjecture. In the case of equation* (3), *the group G is the normalizer of a Cartan subgroup of* $\mathbf{GL}_2(\mathbb{F}_p)$. *This Cartan subgroup is split if* $p \equiv 1 \pmod 3$, *and is non-split if* $p \equiv -1 \pmod 3$. *Moreover, the field cut out by ϱ is an abelian extension of* $\mathbb{Q}(\zeta_3)$.

The reader will note that a proof of the following proposition already appears, for the most part, in [3] and [25], where it is assumed that the exponent p is $\geqq 17$. We repeat the entire proof here for the sake of completeness, and also to indicate how to handle the small primes $p = 7$ and 13.

Proposition 4.2. *Let* $n = p \geqq 7$ *be prime.*

1. *Equation* (1) *has no non-trivial proper solution if* $p \equiv 1 \pmod 4$.

2. *Equation* (2) *has no non-trivial proper solution if* $p \equiv 1 \pmod 4$.

3. *Equation* (3) *has no non-trivial proper solution if* $p \equiv 1 \pmod 3$, *assuming the Shimura-Taniyama conjecture.*

Proof. If p satisfies the congruences in Proposition 4.2, then Proposition 4.1 shows that the curve E has a pair of subgroups of order p which are defined over $\mathbb{Q}(i)$ or $\mathbb{Q}(\zeta_3)$, and are interchanged by $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, so that the set of these two subgroups is defined over $\mathbb{Q}$. Hence E gives rise to a rational point on the curve denoted by $X_{\mathrm{split}}(p)$ in [21] (and $X^s(p)$ below). Suppose first that $p \geqq 17$. Then by Prop. 3.1. of [21], E has potentially good reduction at all primes $\ell \neq 2$, contradicting Lemma 1.2.

To complete the proof of parts 1 and 2, it remains to dispose of the case $n = 13$ of equations (1) and (2). One can proceed in various ad hoc ways. For example, one can note that the curve E gives rise to a point on $X_0(26)$ which is defined over $\mathbb{Q}(i)$. The Jacobian $J_0(26)$ is a two-dimensional abelian variety which is isogenous to the product of two elliptic curves of rank 0, denoted $26A$ and $26B$ in Cremona's tables [2]. Their twists over $\mathbb{Q}(i)$ are curves of conductor $208 = 2^4 13$, denoted $208A$ and $208D$. One checks again from Cremona's tables that the curve $208A$ has rank 0. Hence $J_0(26)$ has a non-trivial quotient, isogenous to $26A$, which is of rank 0 over $\mathbb{Q}(i)$. Now, Corollary 4.3. of [18] implies that $abc = \pm 1$ in the case of equation (1), and that $ab = \pm 1$, in the case of equation (2) (where we have also used Corollary 3.2).

To finish the proof of part 3, one makes a similar argument to handle the exponents $n = 7$ and $n = 13$: the key points that need to be checked are that the modular Jacobians $J_0(21)$ and $J_0(39)$ have quotients of rank 0 over the field $\mathbb{Q}(\sqrt{-3})$. For example, the Jacobian $J_0(21)$ is isogenous to the elliptic curve denoted $21A$ in Cremona's tables, which has rank 0 and whose twist over $\mathbb{Q}(\sqrt{-3})$ (the curve $63A$) also has rank 0. In the case of the variety $J_0(39)$, the quotient which is of rank 0 over $\mathbb{Q}(\sqrt{-3})$ is of dimension two and hence is not listed in the tables of Cremona. But its existence follows, for example, from Proposition 2.1 of [13].

Thanks to Proposition 4.2, it now remains to prove the Main Theorem in the case where the exponent $p \geqq 7$ is in addition $\equiv -1 \pmod 4$ in the case of equations (1) and (2), and is $\equiv -1 \pmod 3$ in the case of equation (3). We shall assume that p satisfies these congruences for the rest of section 4.

Recall that G denotes the image of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $\mathbf{GL}_2(\mathbb{F}_p)$ under ϱ. Let D_p be a decomposition group at p, and let G_p be its image under ϱ.

Proposition 4.3. $G_p = G$, *and in particular* G_p *is the normalizer of a non-split Cartan subgroup of* $\mathbf{GL}_2(\mathbb{F}_p)$.

Proof. This follows directly from the theory of complex multiplication: the field cut out by ϱ is an abelian extension of the quadratic imaginary field $K = \mathbb{Q}(i)$ or $\mathbb{Q}(\zeta_3)$ of degree $(p-1)(p+1)$ which is totally ramified at the unique prime of K above p.

Corollary 4.4. *The prime* p *does not divide* abc *in the case of equation* (1), *and does not divide* ab *in the case of equations* (2) *and* (3).

Proof. Otherwise, E would have multiplicative reduction at p. The group G_p would be contained in a Borel subgroup of $\mathbf{GL}_2(\mathbb{F}_p)$ by Tate's analytic theory, contradicting Proposition 4.3.

5. The modular curve $X_0^{ns}(r,p)$

Let K^{ns} be the normalizer of a nonsplit Cartan subgroup of $\mathbf{GL}_2(\mathbb{F}_p)$. Let $X(p)$ be the modular curve classifying elliptic curves with full level p structure equipped with its action of the group $\mathbf{GL}_2(\mathbb{F}_p)$. Let $X^{ns}(p) = X(p)/K^{ns}$, and define (for $r = 2$ or 3)

$$X^{ns}(r,p) := X^{ns}(p) \times_{X(1)} X(r), \quad X_0^{ns}(r,p) := X^{ns}(p) \times_{X(1)} X_0(r).$$

By combining the "mod p" information given by Proposition 4.1 with the "mod r" information ($r = 2$ or 3) of Lemma 1.3, we find that the Frey curve E constructed in section 1 corresponds to a rational point on the following modular curve:

Equation (1): $X = X^{ns}(2,p)$.

Equation (2): $X = X_0^{ns}(2,p)$.

Equation (3): $X = X_0^{ns}(3,p)$.

Let $x \in X(\mathbb{Q})$ be the rational point associated to the curve E, and let $\underline{x}$ be its natural image in the curve $\underline{X} = X(2)$, $X_0(2)$, or $X_0(3)$ obtained by "forgetting the level p structure".

The curve $\underline{X}$. The curve $X(2)$ has over $\mathrm{Spec}(\mathbb{Z}[1/2])$, a canonical smooth proper model and is identified with the classical λ-line of Legendre. It has three cusps which correspond to the values $\lambda = 0$, 1, and ∞. These cusps are permuted transitively by the Galois group $\mathrm{Gal}(X(2)/X(1)) \simeq S_3$, and are denoted $\underline{0}$, $\underline{1}$ and $\underline{\infty}$.

The curve $X_0(r)$ has a smooth proper model over $\mathrm{Spec}(\mathbb{Z}[1/r])$ which is isomorphic to the projective line when $r = 2$ or 3. The curve $X_0(r)$ (with r any prime) has two cusps, which are denoted $\underline{0}$ and $\underline{\infty}$ because they correspond to the images of $\tau = 0$ and $\tau = \infty$ on the upper half plane.

The curve $X(2)$ has three distinct projections π_0, π_1, and π_∞ to $X_0(2)$ given by "forgetting" various parts of the level 2 structure. The map π_j is characterized by the fact that $\pi_j(j) = \infty$.

Given a rational prime ℓ which is not equal to r, and two sections P and Q of $X(2)$ or of $X_0(r)$ over $\mathrm{Spec}(\mathbb{Z}[1/r])$, we let $(P \cdot Q)_\ell$ denote the usual arithmetic intersection number. Note that the intersection number $(\underline{x} \cdot \xi)_\ell$ is strictly positive, for some cusp ξ of $\underline{X}$, if and only if the curve E has multiplicative reduction at ℓ, i.e., if and only if ℓ divides abc in the case of equation (1) or ab in the case of equations (2) and (3).

The curve X. The genus of the curve X ($= X^{\mathrm{ns}}(2,p)$, $X_0^{\mathrm{ns}}(2,p)$ or $X_0^{\mathrm{ns}}(3,p)$) is indicated below. (Here, χ_2 and χ_3 are the primitive Dirichlet characters associated to $\mathbb{Q}(i)$ and $\mathbb{Q}(\zeta_3)$ respectively.)

$$X^{\mathrm{ns}}(2,p)\colon \quad g = \frac{p^2-4p+7}{4};$$

$$X_0^{\mathrm{ns}}(2,p)\colon \quad g = \frac{p^2-6p+11+2\chi_2(p)}{8};$$

$$X_0^{\mathrm{ns}}(3,p)\colon \quad g = \frac{p^2-4p+8+\chi_3(p)}{6}.$$

Note that each of these curves has non-zero genus when $p > 3$.

We now give a description of the cusps of X. The map of modular curves $X \to \underline{X}$ is a (non-Galois) covering of degree $d = p(p-1)/2$. There are exactly $m = (p-1)/2 = d/p$ cusps above each cusp of $\underline{X}$. Each of these maps to $\underline{X}$ with ramification index equal to p. These m cusps are all defined over the totally real subfield $\mathbb{Q}(\mu_p)^+$ of $\mathbb{Q}(\mu_p)$ and are permuted transitively by the absolute Galois group of $\mathbb{Q}$. In particular, X has no rational cusps. We denote the cusps above $\underline{\xi}$ by $\xi_1, \ldots, \xi_m$.

Specifically, the curve $X^{\mathrm{ns}}(2,p)$ has $3m$ cusps, denoted

$$0_1, \ldots, 0_m, \quad 1_1, \ldots, 1_m, \quad \infty_1, \ldots, \infty_m.$$

The curve $X_0^{\mathrm{ns}}(r,p)$ has $2m$ cusps, denoted

$$0_1, \ldots, 0_m, \quad \infty_1, \ldots, \infty_m.$$

6. Connection with the curve $X_0(rp^2)/w_p$

Assume for now that p is any odd prime, and that r is an integer which is not divisible by p. (In all our applications, we will have either $r = 2$ or $r = 3$.) Let J be the Jacobian of X. We write also $J = J^{\mathrm{ns}}(2,p)$, $J_0^{\mathrm{ns}}(2,p)$, and $J_0^{\mathrm{ns}}(3,p)$ to denote the Jacobians of $X^{\mathrm{ns}}(2,p)$, $X_0^{\mathrm{ns}}(2,p)$, and $X_0^{\mathrm{ns}}(3,p)$ in the case of equations (1), (2), and (3) respectively.

The following basic result of Imin Chen [1] relates J to the Jacobians of the form $J_0(N)$. Chen's original proof, based on the trace formula, was indirect and did not exhibit an explicit isogeny. Theorem 6.1 follows from work of Edixhoven [11]. (Note that Edixhoven does not explicitly state Theorem 6.1, but the "functorial" nature of his proof provides it as an application of Theorem 1.3 of [11], using the fact that $J_0(r)$ is trivial for $r = 2$ or 3.) Recall that the curve $X_0(rp^2)$ is equipped with two involutions denoted w_p and w_r. Let $J_0'(rp^2)$ be the p-new quotient of the Jacobian $J_0(rp^2)$, and let $J_0'(rp^2)/w_p$ be its quotient by the Atkin-Lehner involution w_p. Note that $J_0(rp^2)/w_p$ is isogenous to the Jacobian of the curve $X_0(rp^2)/w_p$.

Theorem 6.1 (Chen-Edixhoven). *There is an isogeny between*

$$J_0^{ns}(r,p) \quad \text{and} \quad J_0'(rp^2)/w_p$$

when $r = 1, 2$ or 3 which is compatible with the action of the Hecke operators T_n (n a positive integer prime to p).

Remark. When $r = 1$, the L-functions attached to the weight 2 cusp forms on $J_0'(p^2)/w_p$ have sign -1 in their functional equation, and hence vanish at $s = 1$. As a result, Theorem 6.1 combined with the Birch and Swinnerton-Dyer conjecture leads one to expect that *every non-zero quotient of $J^{ns}(p)$ has infinite Mordell-Weil group.* This is why the curve $X^{ns}(p)$ has proved stubbornly resistant to Mazur's methods. The auxiliary level r structure ($r = 2$ or 3) plays a providential role in our argument, by allowing us to work with the curve $X_0^{ns}(r,p)$, whose Jacobian does have a non-trivial quotient of rank 0, as will be shown in section 7.

We define below an explicit correspondence α_r between modular curves. From this correspondence one deduces an homomorphism of abelian varieties between the Jacobian of the curves. We expect (but do not need to prove) this homomorphism to be an explicit description of Chen's isogeny of Theorem 6.1.

We begin with the case $r = 1$. Let K^s be the group consisting of diagonal or antidiagonal matrices of $\mathbf{GL}_2(\mathbb{F}_p)$; it is the normalizer of a split Cartan subgroup of $\mathbf{GL}_2(\mathbb{F}_p)$. Define the curve $X^s(p)$ to be $X(p)/K^s$. Both the curves $X^{ns}(p)$ and $X^s(p)$ are quotients of the curve $X'(p) = X(p)/(K^s \cap K^{ns})$. The morphism

$$X'(p) \to X^s(p) \times X^{ns}(p)$$

(given by the obvious pair of degeneracy maps) defines a correspondence π from $X^s(p)$ to $X^{ns}(p)$. This correspondence is of degree $(p-1)/2$ (the index of $K^s \cap K^{ns}$ in K^s) and is defined over $\mathbb{Q}$.

The modular curve $X^s(p)$ is isomorphic to $X_0(p^2)/w_p$. The isomorphism u from $X^s(p)$ to $X_0(p^2)/w_p$ is deduced from the multiplication by $1/p$ in the upper half-plane. In particular it sends the cusp ∞ of $X^s(p)$ to the cusp ∞ of $X_0(p^2)/w_p$.

Let α be the correspondence from $X_0(p^2)/w_p$ to $X^{ns}(p)$ obtained by composing π with u. We say that a divisor of $X_0(p^2)/w_p$ (resp. $X^{ns}(p)$) is *p-old* if it is in the image of the correspondence $X_0(p) \to X_0(p^2)/w_p$ (resp. $X_0(1) \to X^{ns}(p)$) deduced from any of the two degeneracy maps $X_0(p^2) \to X_0(p)$ (resp. from the degeneracy map $X^{ns}(p) \to X_0(1)$). We will need two properties of the correspondence α.

Lemma 6.2. (a) *The image by α of a p-old divisor of $X_0(p^2)/w_p$ is a p-old divisor of $X^{ns}(p)$.*

(b) *The image by α of the cusp ∞ of $X_0(p^2)/w_p$ is equal to the sum of the cusps of $X^{ns}(p)$.*

Proof. The property (a) can be seen more clearly by considering the curve $X^s(p)$ instead of the curve $X_0(p^2)/w_p$. In this context a p-old divisor of $X^s(p)$ has an inverse image in $X(p)$ which is the sum of a B^+ and of a B^--invariant divisor, where B^+ and B^-

are the upper and lower triangular Borel subgroups of $\mathbf{GL}_2(\mathbb{F}_p)$ respectively. The image in $X^{\mathrm{ns}}(p)$ of such a divisor is $\mathbf{GL}_2(\mathbb{F}_p)$-invariant since $K^{\mathrm{ns}}B = \mathbf{GL}_2(\mathbb{F}_p)$ ($B = B^+$ or B^-).

We see the property (b) from the fact that $\alpha(\infty)$ is necessarily a divisor of degree $(p-1)/2$ defined over $\mathbb{Q}$. The sum of the cusps of $X^{\mathrm{ns}}(p)$ is the only cuspidal divisor with this property.

By working over the base $X_0(r)$ we deduce a correspondence α_r from $X_0(rp^2)/w_p$ to $X_0^{\mathrm{ns}}(r,p)$. This correspondence defines functorially a $\mathbb{Q}$-linear homomorphism of groups

$$\alpha_{r*}: \mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q}) \to \mathrm{H}_1(X_0^{\mathrm{ns}}(r,p), \mathbb{Q}),$$

and a homomorphism of abelian varieties from the jacobian of $X_0(rp^2)/w_p$ to $J_0^{\mathrm{ns}}(r,p)$.

Corollary 6.3. *The p-old part of* $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})$ *is in the kernel of* α_{r*}.

Proof. Lemma 6.2 part (a) adapted to our situation with an extra $\Gamma_0(r)$-structure implies that the p-old part of $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})$ is sent by α_{r*} into the image of the group homomorphism $\mathrm{H}_1(X_0(r), \mathbb{Q}) \to \mathrm{H}_1(X_0^{\mathrm{ns}}(r,p), \mathbb{Q})$ deduced functorially from the natural map $X_0^{\mathrm{ns}}(r,p) \to X_0(r)$. Since $X_0(r)$ has genus 0 this image is 0.

Recall the notation introduced in section 5 for the cusps of $X_0^{\mathrm{ns}}(r,p)$. Let D be the cuspidal divisor of degree 0 defined by

$$D = (0_1) + \cdots + (0_m) - (\infty_1) - \cdots - (\infty_m).$$

Corollary 6.4. *The image by* α_r *of the divisor* $(0) - (\infty)$ *of* $X_0(rp^2)/w_p$ *is* D.

Proof. This is just the version over the base $X_0(r)$ of part (b) of Lemma 6.2.

Proposition 6.5. *The class of* D *in* $J_0^{\mathrm{ns}}(r,p)(\mathbb{Q})$ *has order* p. *In particular, it is non-trivial.*

Proof. Since $X_0(r)$ has genus zero, we may choose an identification of $X_0(r)$ with $\mathbb{P}^1$ which send the cusps 0 and ∞ to 0 and ∞. Let $\phi: X_0^{\mathrm{ns}}(r,p) \to X_0(r) \simeq \mathbb{P}^1$ be the natural projection. The divisor $\mathrm{Div}(\phi)$ is equal to pD, so that the order of D divides p. To see that D has order p, suppose otherwise. Then we would have $\phi = \alpha^p$, for some rational function α on $X_0^{\mathrm{ns}}(r,p)$. Let X' be an irreducible component of the modular curve classifying elliptic curves with a subgroup of order r and *full* level p structure. It is a Galois covering of $X_0(r)$ with Galois group $\mathrm{PSL}_2(\mathbb{F}_p)$. Consider the sequence of coverings of complex curves:

$$X' \xrightarrow{\pi} X_0^{\mathrm{ns}}(r,p) \xrightarrow{\alpha} \mathbb{P}^1 \xrightarrow{\nu} X_0(r) = \mathbb{P}^1,$$

where π is the natural quotient map, and ν is the cyclic covering of degree p which sends a suitable parameter t of $\mathbb{P}^1$ to t^p. Both the coverings $\nu\alpha\pi$ and ν are Galois, with Galois groups isomorphic to $\mathbf{PSL}_2(\mathbb{F}_p)$ and $\mathbb{Z}/p\mathbb{Z}$ respectively. This is a contradiction: there are no surjective homomorphisms from $\mathbf{PSL}_2(\mathbb{F}_p)$ to $\mathbb{Z}/p\mathbb{Z}$ when $p \geqq 5$. (The reader will note

that when $p = 3$, the curve $X_0^{ns}(2,3)$ has genus 0, so that the divisor D *is* principal. Indeed, the group $\mathbf{PSL}_2(\mathbb{F}_3) \simeq A_4$ does have a cyclic quotient of order 3 in this case.)

Remark. One could also invoke the classical theorem of Galois ("Lorsque $p > 11$ le degré de l'équation modulaire ne s'abaisse pas à p") stating that the group $\mathbf{PSL}_2(\mathbb{F}_p)$ has no subgroup of index p when $p > 11$.

Examples. 1. The curve $X_0^{ns}(2,5)$ is a curve of genus one. By Theorem 6.1, it is isogenous to the elliptic curve $J_0(50)/w_5$. (The modular forms of level 50 are automatically new at 5, since $X_0(2)$ and $X_0(10)$ are of genus 0.) This is the curve denoted $50B$ in the tables of Cremona [2], with equation

$$y^2 + xy + y = x^3 + x^2 - 3x + 1 .$$

The Mordell-Weil group of this curve is a finite group of order 5, generated by D by Proposition 6.5.

2. The curve $X_0^{ns}(3,5)$ is a curve of genus 2, and its Jacobian is isogenous to a product of two elliptic curves of conductor 75, which are denoted $75B$ and $75C$ in [2]. Both of these curves have a finite Mordell-Weil group, and the factor $75C$ has a rational point of order 5.

3. The curve $X_0^{ns}(2,7)$ has genus 2, and $J_0^{ns}(2,7)$ is an irreducible two-dimensional quotient of $J_0(98)$. The Hecke ring acting on this quotient is isomorphic to an order in $\mathbb{Z}[\sqrt{2}]$, and the following table lists the values of the good Hecke operators T_ℓ for the first few primes $\ell \nmid 14$ (the authors are grateful to Jordi Quer for assisting them with this calculation):

ℓ	3	5	11	13	17	23	29	31	37	41	43
T_ℓ	$\sqrt{2}$	$-2\sqrt{2}$	-2	0	$\sqrt{2}$	-4	2	$-6\sqrt{2}$	10	$7\sqrt{2}$	2

The reader will note that the values of $\ell + 1 - T_\ell$ computed from this table are always divisible by the prime ideal $(7, 4 - \sqrt{2})$ of $\mathbb{Z}[\sqrt{2}]$ which lies above 7.

7. Winding quotients

We will construct, as in [20], a "winding quotient" of J. Recall that a quotient abelian variety A of an abelian variety B is *optimal* if the kernel of the map $B \to A$ is connected.

Let us point out that in the proof of the following proposition we make a crucial use of a theorem of Kolyvagin and Logachev [15] (supplemented by work of Gross-Zagier, Bump-Friedberg-Hoffstein and Murty-Murty) in the direction of the conjecture of Birch and Swinnerton-Dyer.

Proposition 7.1. *The abelian variety J possesses a nonzero optimal quotient J_e such that:*

(1) *$J_e(\mathbb{Q})$ is finite.*

(2) *The kernel of the homomorphism of abelian varieties $J \to J_e$ is stable under the action of the Hecke operators T_ℓ ($\ell \nmid p$). Therefore J_e inherits a natural Hecke action from J.*

Proof. Chen's isogeny (cf. Theorem 6.1) provides an isogeny $J_0'(rp^2)/w_p \to J$ which is compatible with the action of Hecke operators of index prime to p. By a result due to Ribet (cf. [18], Proposition 2.1), any optimal quotient of the new quotient of $J_0(N)$ (where N is any integer) inherits an action of Hecke operators compatible with the action on $J_0(N)$. Therefore it is enough to show that the new quotient A of the jacobian of $X_0(rp^2)/w_p$ has a nonzero optimal quotient A_e with finitely many rational points.

The Hecke operators T_ℓ (ℓ a prime number not dividing pr) operate on A in a manner compatible with their action on $J_0(rp^2)$. They generate a commutative semi-simple $\mathbb{Q}$-algebra $\mathbb{T}$ contained in the algebra $\mathrm{End}(A) \otimes \mathbb{Q}$. Let S be the space of cusp forms of weight 2 for $\Gamma_0(rp^2)$ which are new, are invariant under w_p and have rational q-expansion at ∞. The space S is endowed with a natural action of $\mathbb{T}$, and is free of rank one over $\mathbb{T}$ by the Atkin-Lehner theory of newforms. Given $f \in S$, let ω_f be the differential form over $\mathbb{C}$ on $X_0(rp^2)/w_p$ deduced from the differential form $2\pi i f(\tau) d\tau$ on the upper half-plane.

In what follows, the homology groups of modular curves will always be understood as the singular homology of the underlying Riemann surfaces. Let $H = \mathrm{H}_1(A(\mathbb{C}), \mathbb{Q})^+$ be the subspace of $\mathrm{H}_1(A(\mathbb{C}), \mathbb{Q})$ which is invariant under complex conjugation. It is naturally a quotient of $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})^+$, and is endowed with a natural action of $\mathbb{T}$. Given $\gamma \in H$, let $\tilde{\gamma}$ be any lift of γ to $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})$.

The formula

$$\langle f, \gamma \rangle := \int_{\tilde{\gamma}} \omega_f$$

defines a nondegenerate $\mathbb{C}$-valued pairing between S and H. The Hecke operators T_ℓ which act on S and on H are self-adjoint with respect to this pairing. As a result, H is also a free $\mathbb{T}$-module of rank one.

Let e be the unique element of $\mathrm{H}_1(X_0(rp^2)/\omega_p, \mathbb{Q})^+$ such that the integral of any differential form ω on $X_0(rp^2)/\omega_p$ of a cycle of class e is equal to the integral of the pullback in the upper half plane of ω over the geodesic path from 0 to $i\infty$ in the upper half plane. (It is rational by the Drinfeld-Manin Theorem and it is fixed by complex conjugation.) Let e_A be the image of e in H. Let $I_e \subset \mathbb{T}$ be the annihilator of e_A, and let I_e' be the intersection of I_e with the subring of $\mathrm{End}(A)$ generated by the Hecke operators T_ℓ ($\ell \nmid pr$). Define $A_e := A/I_e'A$.

We claim that $A_e(\mathbb{Q})$ is finite. In view of a theorem of Kolyvagin and Logachev [15], it is enough to show that the L-function $L(A_e, s)$ is non-zero at $s = 1$.

Let $S_{\mathbb{C}}$ be the complex vector space of cusp forms generated by S, and let S_{I_e} be the subspace of $S_{\mathbb{C}}$ which is annihilated by I_e. A theorem of Eichler, Shimura, Igusa and Carayol asserts that $L(A_e, s)$ is the product of the L-functions associated to the normalized eigenforms in S_{I_e}. (In its classical formulation the theorem is concerned only with the case

where I_e is a prime ideal, but it extends directly to any ideal of $\mathbb{T}$.) If f is any eigenform in $S_{\mathbb{C}}$, then

$$L(f,1) = 2\pi i \int_0^{i\infty} f(\tau)\, d\tau = \langle f, e_A \rangle .$$

But if f belongs to S_{I_e}, it is orthogonal to $I_e H$, and hence cannot be orthogonal to e_A, since $H = I_e H \oplus \mathbb{T} e_A$ and the pairing $\langle , \rangle$ is nondegenerate. Therefore $L(f,1) \neq 0$.

This proves that $L(A_e,1) \neq 0$ and therefore that $A_e(\mathbb{Q})$ is finite. It remains to show that A_e is non-zero, i.e. that I_e is not the full algebra $\mathbb{T}$. This is a consequence of the following key proposition.

Proposition 7.2. *The homology class e of* $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})$ *associated to the geodesic path from* 0 *to* $i\infty$ *in the upper half-plane does not belong to the old part of* $\mathrm{H}_1(X_0(rp^2)/w_p, \mathbb{Q})$.

Proof. The $\mathbb{Q}$-vector space $\mathrm{H}_1(X_0^{\mathrm{ns}}(r,p), \mathbb{Q})$ is the sum of its new part, its p-old part and its r-old part. It is also the direct sum of its r-old and r-new parts, where the r-new part is the intersection of the kernels of the group homomorphisms on the homologies deduced from the two degeneracy maps $X_0(rp^2)/w_p \to X_0(p^2)/w_p$.

Let us prove first that e is r-new. Note first that $w_r e = -e$, since the image in $X_0(rp^2)(\mathbb{C})$ of the geodesic path from 0 to $i\infty$ in the upper half plane is reversed by the action of the involution $w_r w_p$ on $X_0(rp^2)$. Therefore if e is killed by one of the two degeneracy maps from level rp^2 to level p^2, it is killed by the other, since these two maps are exchanged by w_r. The image of e in the homology of $X_0(p^2)/w_p$ by the group homomorphism deduced from morphism $X_0(rp^2)/w_p \to X_0(p^2)/w_p$ which forgets the $\Gamma_0(r)$-structure is trivial, since the image in $X_0(p^2)(\mathbb{C})$ of the geodesic path from 0 to $i\infty$ in the upper half-plane is reversed by application of the involution w_p.

It remains to prove that e is not p-old. Suppose it is. Then its image under α_{r*} would be 0 by Corollary 6.3. Now note that the boundary of the path from 0 to ∞ in the upper half-plane is equal to the divisor $(\infty)-(0)$. By Abel's description of the set of complex points of the Jacobian of a curve X as $\mathrm{H}_1(X,\mathbb{R})/\mathrm{H}_1(X,\mathbb{Z})$, the denominator of $\alpha_{r*}(e)$ (i.e. the order of the image of $\alpha_{r*}(e)$ in $\mathrm{H}_1(X,\mathbb{R})/\mathrm{H}_1(X,\mathbb{Z})$ for $X = X_0^{\mathrm{ns}}(r,p)$) is equal to the order of the class in $J_0^{\mathrm{ns}}(r,p)$ of the degree 0 divisor $\alpha_r((\infty)-(0))$. By Corollary 6.4, this divisor is the divisor D of Proposition 6.5. By Proposition 6.5, its class in $J_0^{\mathrm{ns}}(r,p)$ is nonzero. Therefore $\alpha_{r*}(e)$ is nonzero.

8. Galois properties of torsion points of elliptic curves

The purpose of this section is to prove the following theorem which might be viewed as a small advance in the direction of a positive answer to Serre's problem ([26], section 4.3). We reproduce here a series of arguments invented by Mazur [18].

We keep the notations of the previous section. In particular, p is an odd prime, E is an elliptic curve over $\mathbb{Q}$ and

$$\varrho : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathbf{GL}_2(\mathbb{F}_p)$$

denotes the mod p Galois representation associated to E.

Theorem 8.1. *Suppose that*

(1) *E has a $\mathbb{Q}$-rational subgroup of order r, with $r = 2$ or 3.*

(2) *$p \geqq 5$, and the image of ϱ in $\mathbf{GL}_2(\mathbb{F}_p)$ is isomorphic to the normalizer of a nonsplit Cartan subgroup.*

Then $j(E)$ belongs to $\mathbb{Z}[\frac{1}{p}]$.

Remark. We would have liked to conclude that $j(E)$ belongs to $\mathbb{Z}$, but we are prevented from doing so by problems arising from the bad reduction of the abelian variety J at the prime p.

Proof of Theorem 8.1. Suppose that $j(E) \notin \mathbb{Z}[\frac{1}{p}]$, and let $\ell \neq p$ be a prime dividing the denominator of $j(E)$. The curve E with its given level structure gives rise to a $\mathbb{Q}$-rational point P on the modular curve $X = X_0^{\mathrm{ns}}(r, p)$ which meets a cusp of X at a prime above ℓ. Since the cusps of X are defined over $\mathbb{Q}(\mu_p)^+$, ℓ must be split completely in $\mathbb{Q}(\mu_p)^+$, i.e., it must be congruent to ± 1 modulo p. Since $p \geqq 5$ we may suppose that ℓ is not 2 or 3. After a suitable choice of our nonsplit Cartan subgroup (there are $(p-1)/2$ of them naturally indexed by the cusps of $X^{\mathrm{ns}}(p)$) and possibly an application of the involution w_r, we may suppose that the cusp is ∞, i.e. the point defined over $R = \mathbb{Z}[\frac{1}{p}, \mu_p]^+$ corresponding to $i\infty$ in the upper half-plane.

Let A be a nonzero optimal quotient of J on which the Hecke operators T_n (n an integer prime to p) operate in a manner compatible with their action on J. Let $\mathscr{A}$ be the Néron model of A. The morphism obtained by composing the map $X \to J$ (normalized by the fact that it sends the cusp ∞ to 0) with the canonical map $J \to A$ is defined over $\mathbb{Q}(\mu_p)^+$ since the cusp ∞ is defined over that field. It extends by the universal property of the Néron model to a morphism ϕ defined over R

$$\phi : \mathscr{X}_{/R}^{\mathrm{smooth}} \to \mathscr{A}_{/R}$$

from the smooth part of the canonical model $\mathscr{X}$ of X over R to $\mathscr{A}$.

Lemma 8.2. *The map ϕ is a formal immersion in characteristic ℓ at the point ∞.*

Proof. We still follow closely the method of Mazur. (What follows is modelled on [18], Proposition 3.1.) We have to show that the map deduced from ϕ on the cotangent spaces at ∞ is surjective. Observe that since ℓ is congruent to ± 1 modulo p, we are actually working over $\mathbb{F}_\ell$. Since the Jacobian J is a semistable abelian variety, by a theorem of Mazur and Raynaud ([18], Corollary 1.1) the map $J \to A$ gives rise functorially to a direct injection from the cotangent space of $\mathscr{A}$ to the cotangent space of the Néron model $\mathscr{J}$ of J away from characteristics 2 and p. The cotangent space of $\mathscr{J}_{/\mathbb{F}_\ell}$ can be identified with the space $\mathrm{H}^0(\mathscr{X}_{/\mathbb{F}_\ell}, \Omega^1)$ of differential forms on $\mathscr{X}_{/\mathbb{F}_\ell}$, which in turn can be identified with the space of cusp forms of weight 2 on $X_0^{\mathrm{ns}}(r, p)$ in characteristic ℓ.

The Tate curve equipped with its p-torsion points defined over $R[\![q^{1/p}]\!]$ provides simultaneously a q-expansion for any such cusp form f of the type $\sum_{n=1}^{\infty} a_n(f)q^{n/p}$ and a basis (given by $dq^{\frac{1}{p}}/q^{\frac{1}{p}}$) over $\mathbb{F}_\ell$ of $\mathrm{Cotg}_\infty(\mathscr{X}_{/\mathbb{F}_\ell}) \simeq \mathbb{F}_\ell$ (see [6], VII. 2). With the identifications considered above, the image by ϕ^* in $\mathrm{Cotg}_\infty(\mathscr{X}_{/\mathbb{F}_\ell}) \simeq \mathbb{F}_\ell$ of the element of $\mathrm{Cotg}_0(\mathscr{J}_{/\mathbb{F}_\ell})$ associated to f is $a_1(f)$.

Let ω be a non-zero element in $\mathrm{Cotg}_0(\mathscr{A}_{/\mathbb{F}_\ell})$ whose associated modular form f (in characteristic ℓ) has q-expansion $\sum_{n>0} a_n(f)q^{n/p}$. Since A is nonzero such an element exists. The Hecke operators T_n (n prime to p) operate on $\mathscr{A}$ and therefore on its cotangent space. The first Fourier coefficient of the modular form associated to $T_n\omega$ is $a_n(f)$. Therefore the image of $T_n\omega$ in $\mathrm{Cotg}_\infty(X_{/\mathbb{F}_\ell}) \simeq \mathbb{F}_\ell$ is $a_n(f)$. Suppose that ϕ is not a formal immersion in characteristic ℓ. Then we would have $a_n(f) = 0$ whenever n is prime to p. That would imply that f is modular for $\Gamma_0(r)$ and therefore zero since there are no cusp forms of weight 2 for $\Gamma_0(2)$ and $\Gamma_0(3)$ in characteristic $\ell > 3$. This contradiction finishes the proof of Lemma 8.2.

Let us return to the proof of Theorem 8.1. Using Proposition 7.1 we will apply our study to $A = J_e$.

Lemma 8.3. *The point $\phi(P)$ is torsion in $J_e(\mathbb{Q}(\mu_p)^+)$.*

Proof. If $\sigma \in \mathrm{Gal}(\mathbb{Q}(\mu_p)^+/\mathbb{Q})$, then

$$(\sigma - 1)\big((P) - (\infty)\big) = (\infty_i) - (\infty_j),$$

for some $1 \leqq i, j \leqq m$. It is a cuspidal divisor and hence is torsion by the Drinfeld-Manin Theorem. Hence there exists an integer n such that $n\big((P) - (\infty)\big)$ belongs to $J(\mathbb{Q})$. In particular, $n\phi(P)$ belongs to $J_e(\mathbb{Q})$. Since $J_e(\mathbb{Q})$ is torsion by Proposition 7.1, the lemma follows.

We can now finish the proof of Theorem 8.1. Since A has good reduction at ℓ and since the prime ℓ is unramified in R, the torsion in A over $\mathbb{Q}(\mu_p)^+$ injects into $A(R/\ell')$, for all primes ℓ' above ℓ. Since $\phi(P)$ reduces to 0 modulo such a prime, and since $\phi(P)$ is torsion by Lemma 8.3, it follows that $\phi(P) = 0 = \phi(\infty)$. (See [18], 1.c.) The fact that the point P specializes to ∞ modulo a prime above ℓ contradicts the fact that ϕ is a formal immersion in characteristic ℓ ([18], 4.b).

9. Proof of the Main Theorem for large exponents

Corollary 9.1. *Suppose that the exponent n is a prime $p \geqq 7$, and let (a, b, c) be a primitive solution to equation* (1), (2), *or* (3) *with $abc \neq 0$.*

1. *In the case of equation* (1), *we have $abc = \pm 1$.*

2. *In the case of equation* (2), *we have $ab = \pm 1$.*

3. *Assume the Shimura-Taniyama conjecture. In the case of equation* (3) *we have* $ab = \pm 1$

Proof. Assuming that it is modular, the curve E constructed from a non-trivial solution (a, b, c) to equation (1), (2) or (3) satisfies the assumptions of Theorem 8.1, by Propositions 4.2 and 4.1. Hence $j(E)$ belongs to $\mathbb{Z}[\frac{1}{p}]$. It follows that abc (in the case of equation (1)) or ab (in the case of equations (2) and (3)) is equal to a power of p. But by Corollary 4.4, the prime p does not divide abc (in the case of equation (1)) or ab (in the case of equation (2) and (3)). Corollary 9.1 follows. This completes our proof of the Main Theorem.

References

[1] *I. Chen*, The Jacobian of non-split Cartan modular curve, Proc. London Math. Soc., to appear.

[2] *J.E. Cremona*, Algorithms for modular elliptic curves, Cambridge University Press, 1992.

[3] *H. Darmon*, The equations $x^n + y^n = z^2$ and $x^n + y^n = z^3$, Internat. Math. Res. Notices **10** (1993), 263–274.

[4] *H. Darmon*, The equation $x^4 - y^4 = z^p$, C.R. Math. Rep. Acad. Sci. Canada **15** (1993), 286–290.

[5] *H. Darmon, A. Granville*, On the equations $x^p + y^q = z^r$ and $z^m = f(x, y)$, Bull. London Math. Soc. no. 129, **27** part 6, November 1995, pp. 513–544.

[6] *P. Deligne, M. Rappoport*, Les schémas de modules des courbes elliptiques, in: Vol. II of the Proceedings of the International Summer School on modular functions, Antwerp (1972), Lect. Notes Math. **349**, Springer, Berlin–Heidelberg–New York 1973.

[7] *P. Dénes*, Über die Diophantische Gleichung $x^\ell + y^\ell = cz^\ell$, Acta Math. **88** (1952), 241–251.

[8] *F. Diamond*, On deformation rings and Hecke rings, Ann. Math. (2) **144** (1996), no. 1, 137–166.

[9] *F. Diamond, K. Kramer*, Modularity of a family of elliptic curves, Math. Res. Lett. **2** (1995), 299–304.

[10] *L.E. Dickson*, History of the theory of numbers, Chelsea, New York 1971.

[11] *B. Edixhoven*, On a result of Imin Chen, Séminaire de théorie des nombres de Paris, 1995–96, Cambridge University Press, to appear.

[12] *G. Frey*, On elliptic curves with isomorphic torsion structures and corresponding curves of genus 2, in: Elliptic curves, modular forms and Fermat's Last Theorem, J. Coates, S.-T. Yau, eds., International Press, Cambridge, MA (1995), 79–98.

[13] *S. Kamienny*, Points on Shimura curves over fields of even degree, Math. Ann. **286** (1990), 731–734.

[14] *M.A. Kenku*, The modular curve $X_0(39)$ and rational isogenies, Math. Proc. Camb. Phil. **85** (1979), 21–23.

[15] *V.A. Kolyvagin, D.Yu. Logachev*, Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties, Leningrad Math. J. **1** no. 5 (1990), 1229–1253.

[16] *D. Kubert*, Universal bounds on torsion of elliptic curves, Proc. London Math. Soc. (3) **33** (1976), 193–237.

[17] *B. Mazur*, Modular curves and the Eisenstein ideal, Publ. Math. IHES **47** (1977), 33–186.

[18] *B. Mazur*, Rational isogenies of prime degree, Invent. Math. **44** (1978), 129–162.

[19] *B. Mazur*, Questions about number, in: New Directions in Mathematics, to appear.

[20] *L. Merel*, Bornes pour la torsion des courbes elliptiques sur les corps de nombres, Invent. Math. **124** (1966), no. 1–3, 437–449.

[21] *F. Momose*, Rational points on the modular curves $X_{\text{split}}(p)$, Comp. Math. **52** (1984), 115–137.

[22] *B. Poonen*, Some diophantine equations of the form $x^n + y^n = z^m$, to appear.

[23] *P. Ribenboim*, Catalan's conjecture, Academic Press, Boston 1994.

[24] *K. Ribet*, On modular representations of $\text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms, Invent. Math. **100** (1990), 431–476.

[25] *K. Ribet*, On the equation $a^p + 2^\alpha b^p + c^p = 0$, Acta Arith. **79** no. 1 (1997), 7–16.

[26] *J.-P. Serre*, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, Invent. Math. **15** (1972), 259–331.

[27] *J.-P. Serre*, Sur les représentations modulaires de degré 2 de $\text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, Duke Math. J. **54** no. 1 (1987), 179–230.

[28] *J. Tate*, Algorithm for determining the type of a singular fiber in an elliptic pencil, in: Modular Functions of One Variable IV, Springer Lect. Notes Math. **476** (1975), 33–52.

[29] *R. Taylor, A. Wiles*, Ring-theoretic properties of certain Hecke algebras, Ann. Math. **141** (1995), 553–572.
[30] *A. Wiles*, Modular elliptic curves and Fermat's Last Theorem, Ann. Math. **141** (1995), 443–551.

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street-West,
Montreal, PQ H3A 2K6, Canada
e-mail: darmon@math.mcgill.ca

Université Denis Diderot, Mathématiques, Case postale 7012, 2 place Jussieu, F-75251 Paris cedex 05
e-mail: merel@math.jussieu.fr

Eingegangen 17. Juni 1996, in revidierter Fassung 4. März 1997