POINTS ON CURVES

LEVENT HASAN ALİ ALPÖGE

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
MATHEMATICS
ADVISOR: MANJUL BHARGAVA

SEPTEMBER 2020

Abstract.

In this thesis we discuss finiteness theorems for integral and rational points on curves over number fields. We do this in two parts.

Part I of the thesis is concerned with upper bounds. We give pointwise upper bounds for the number of: $\mathfrak{o}_{K,S}$-points on an elliptic curve (this is joint work with Wei Ho), $\mathfrak{o}_{K,S}$-points on a hyperelliptic curve, and large K-points on a hyperbolic curve. We also give upper bounds on average for: ranks of elliptic curves in certain thin families, and the number of $\mathbb{Q}$-points on odd genus two curves. For the former we introduce a technique that should allow one to treat counts on invariant quadrics in arithmetic statistics in considerable generality.

Part II of the thesis is concerned with effectivity. We give an algorithm which, on input (g, K, S), outputs the g-dimensional abelian varieties over K with good reduction outside S. We prove this algorithm always terminates under standard motivic conjectures. (This is joint work with Brian Lawrence.) Using a theorem of Bogomolov-Tschinkel, we give an algorithm that, assuming strong modularity conjectures for GL_2 over all number fields, computes $C(K)$ for C/K a hyperbolic hyperelliptic curve. We give an *unconditional* algorithm that determines the $\mathfrak{o}_{K,S}$-points on a Hilbert modular variety in finite time when K is an odd-degree totally real field. Because complete curves in such varieties abound this effectivizes Faltings' Theorem in some cases. Using a construction of Cohen-Wolfart, we give an algorithm that, assuming the existence of motives associated to weight zero cuspidal automorphic representations of GL_2 over CM fields, computes $C(K)$ when K is CM and C/K admits a Belyi map over K with sufficiently divisible ramification degrees. Finally we work out an example in detail: given $K/\mathbb{Q}$ totally real of odd degree and $a \in K^\times$, we explain how to unconditionally compute $C_a(K)$, where $C_a : x^6 + 4y^3 = a^2$.

Acknowledgements.

I would like to thank Professor Manjul Bhargava for being a fantastic advisor and for always managing to encourage me, and I would like to thank Professor Peter Sarnak for his advice and teachings throughout the years, as well as for always optimally motivating me to push for more.

Looking back, it is staggering how superb my teachers and mentors have been. I would like to thank all of them, including Alan Blayne, Noam Elkies, Dennis Gaitsgory, Benedict Gross, Joe Gallian, Anita Kapner, Matthew Maroney, Steven J. Miller, Jacob Tsimerman, and many others. I would also like to thank Peter Diao, Daniel Shapiro, and others involved in the Ross Program for introducing me to number theory proper. Having successfully shoehorned Gauss and Jacobi sums into this thesis, I am reminded of Peter asking me what $\frac{1}{2}$ *is* (in order to motivate inversion in modular arithmetic) and his endless patience in teaching a difficult student who knew nothing.

This thesis features joint work with two collaborators, namely Wei Ho and Brian Lawrence, and I thank them with admiration. I hope it is clear just how much of this thesis flows from those works.

I have further benefited from discussions with Mladen Dimitrov, Tony Feng, Curtis McMullen, Will Sawin, Christopher Skinner, Richard Taylor, Paula Tretkoff, Akshay Venkatesh, Rafael von Känel, Jürgen Wolfart, and Shou-Wu Zhang, as well as a large proportion of those already listed above. I thank Will in particular for immediately catching an error in a construction I was describing to him in conversation. At the time I thought I must have simply misunderstood the paper I was referencing, but eventually it became clear that he had saved me from building on an error in the literature.

I would also like to thank Michael Stoll, who organized "Rational Points 2019", and David Harari, Emmanuel Peyre, and Alexei Skorobogatov, who organized the

"Reinventing Rational Points" program, for providing an outstanding environment in which to think about these problems, and also for their invitations at a time when I felt maximally useless.

Finally I thank my family and friends for being wonderful. My grandmother, Suna Güven, has by now given multiple proofs that mathematicians are, in fact, machines for turning ayva tatlısı into theorems. She, my mother Simay, my father Oğuz, my sister Elif, the rest of my family, and my incredible friends have given me everything. Olağanüstüsünüz. I would also be remiss if I did not thank Anirudha Balasubramanian for absolutely critical discussions about the Irwin-Hall distribution.

И Нина, ты просто совершенство. Мне так повезло.

Cin gibi Anneciğim, Babacığım, Anneanneciğim, ailem için.

Ne şanslıymışım.

Contents.

Chapter 1

Introduction.

Fortasse difficilior videtur res quae nondum familiaris est,
***nam male sperant incipientium animi**. . .*

Diophantus, *Arithmetica*[1] (p. 1, Book I).

In this thesis we study Diophantine equations. Matiyasevich, building on work of Davis, Putnam, and Robinson, has shown that there cannot be a finite-time algorithm that determines if an input $f \in \mathbb{Z}[x_1, \ldots, x_n]$ has a zero in $\mathbb{Z}^n$. It is expected that the same conclusion holds when the inputs are restricted to $\mathbb{Z}[x_1, x_2, x_3]$. In this thesis we are interested in the same problem when the inputs are restricted to $\mathbb{Z}[x_1, x_2]$, or, what one quickly sees is the same, in the problem of finding points on curves.

It is evident that there is a finite-time algorithm in the case of one variable — i.e. when the inputs are restricted to $\mathbb{Z}[x_1]$, so that one is dealing with equations of the form $f(x) = 0$ with $f \in \mathbb{Z}[x]$. In the case of two variables one similarly expects a positive solution of this restricted form of Hilbert's tenth problem, in contrast to Matiyasevich: there should be a finite-time algorithm that determines if an input $f \in \mathbb{Z}[x, y]$ has a zero in $\mathbb{Z}^2$, and indeed one expects the same to hold over $\mathbb{Q}$ as

[1] — of course, and for obvious reasons, in Latin translation [99]. Note, too, the relevance of Chortasmenos' famous scholium.

well. By desingularizing and treating the genus zero cases by hand, we arrive at the core of the problem: is there a finite-time algorithm that, on input a smooth projective curve $C/\mathbb{Q}$ of positive genus, outputs[2] $C(\mathbb{Q})$?

Note that, for this question to even be sensible, we must know that we can always output $C(\mathbb{Q})$ in finite time in the first place.

1.1 Part I.

Thus in Part I of this thesis we study finiteness theorems.

In Chapter 2, which is based on joint work with Wei Ho, we optimize Mordell's original invariant-theoretic proof from 1914 that there are only finitely many solutions to $y^2 = x^3+Ax+B$ in $(x, y) \in \mathbb{Z}^2$ when $A, B \in \mathbb{Z}$ and $\Delta_{A,B} := -4A^3-27B^2 \neq 0$. We prove specifically that the number of such solutions is

$$\ll 2^{\operatorname{rank} E_{A,B}(\mathbb{Q})} \cdot O(1)^{\#|\{p : p^2 | \Delta_{A,B}\}|},$$

where $E_{A,B}/\mathbb{Q}$ is the elliptic curve with affine Weierstrass model $\mathcal{E}^{\text{aff.}}_{A,B} : y^2 = x^3 + Ax + B$.[3] This leads to an immediate corollary controlling the average behaviour of $\#|\mathcal{E}^{\text{aff.}}_{A,B}(\mathbb{Z})|$ when varying A and B, via Bhargava-Shankar's results controlling sizes of n-Selmer groups, and thus expressions like $n^{\operatorname{rank} E_{A,B}(\mathbb{Q})}$, on average, at least when $n \leq 5$.

Mordell, of course, came to his finiteness theorem by first studying his namesake curves $y^2 = x^3 + k$. In Chapter 3 we discuss a method to restrict the analysis of Bhargava-Shankar to this subfamily $E_{0,k}$, at least for $n = 2$. This amounts to considering only those 2-Selmer elements, represented by binary quartics $F \in \mathbb{Z}[X, Y]$, that lie on the locus $I(F) = 0$ — here the crucial point is that the classical invariant

[2] Naturally in the case of elliptic curves $E/\mathbb{Q}$ we ask for $E(\mathbb{Q})_{\text{tors.}}$ and a basis of $E(\mathbb{Q})/\text{tors.}$ in the output instead.

[3] We note that this result generalizes to S-integral points over a number field K.

I is a quadric in the five coefficients of F. The fundamental idea, which is to count solutions to the quadric $I(F) = 0$ in five variables using the circle method, is due to Ruth in his Princeton PhD thesis. We simply provide a trivial observation which replaces difficult ad hoc analysis in his treatment. This allows for generalization, and we treat the case of the yet thinner family E_{0,k^2}, where 2-Selmer elements are represented by pairs of binary cubics with vanishing invariant pairing, as another example of the technique.

We turn next to higher-genus curves. In Chapter 4 we optimize the proof by the famous mathematician[4] X in 1926 that there are only finitely many solutions to $y^2 = f(x)$ in $(x, y) \in \mathbb{Z}^2$, where $f \in \mathbb{Z}[x]$ is monic of degree $\deg f \geq 3$ with nonzero discriminant $\Delta_f \neq 0$. Specifically, we prove that the number of such solutions is

$$\ll 2^{\operatorname{rank} \operatorname{Jac} C_f(\mathbb{Q})} \cdot |\Delta_f|^{o(1)},$$

where $C_f/\mathbb{Q}$ is the hyperelliptic curve with affine Weierstrass model $\mathcal{C}_f^{\text{aff.}} : y^2 = f(x)$.[5] The method is to simply do a 2-descent (i.e. bound the sizes of the fibres of the map $\mathcal{C}_f^{\text{aff.}}(\mathbb{Z}) \to \operatorname{Jac} C_f(\mathbb{Q})/2$), though carefully — the best known bound, due to Evertse-Silverman in 1986, was $\ll \#|\mathrm{Cl}(L_f)[2]|^2 \cdot O(1)^{\#|\{p : p | \Delta_f\}|}$, where $L_f/\mathbb{Q}$ is a number field containing at least three roots of f.

However there is a more interesting finiteness theorem available in this context, namely the incredible theorem of Faltings, formerly conjecture of Mordell, that a smooth projective hyperbolic[6] curve over a number field has finitely many rational points. In Chapter 5 we work as explicitly as possible, showing in the particular

[4] Here $X \simeq$ Siegel, in his only published work between his arrival in Frankfurt in 1922 and his 1929 proof, combining his improvement on Thue's theorem as an undergraduate with Weil's recent thesis, of his famous theorem on integral points on affine curves. Amusingly $X^3 \simeq$ Weil, a corollary of one of many publications from his Nancago years: a 1957 note in Italian giving a counterexample to a 1909 conjecture of Severi.

[5] Again this result generalizes to S-integral points over a number field K.

[6] We call curves of genus at least two hyperbolic, for obvious reasons (consider the universal cover of the complex points).

test case of genus two curves that

$$\operatorname*{Avg}_{H(f)\leq X} \#|C_f(\mathbb{Q})| \ll 1$$

for all $X \in \mathbb{R}^+$, the average taken over $f(x) =: \sum_{i=0}^{5} a_i \cdot X^{5-i}$, with $a_0 := 1, a_1 := 0$, and $|a_i| \leq X^i$ for all i. The technique is a combination of an explicit form of Mumford's gap principle, which was arguably the first step towards proving the Mordell conjecture, with an explicit form, due to Bombieri, of Vojta's gap principle, which Vojta used to give a second proof of Faltings' Theorem. Ultimately we control $\#|C_f(\mathbb{Q})|$ in terms of $2^{\mathrm{rank}\,\mathrm{Jac}\,C_f(\mathbb{Q})}$, and then cite a theorem of Bhargava-Gross controlling the average of $\#|\mathrm{Sel}_2(\mathrm{Jac}\,C_f/\mathbb{Q})|$ over this family, generalizing the theorem of Bhargava-Shankar we use in Chapter 2.

However to do this we must optimize the bound on the number of (large) rational points given by Vojta's gap principle, because the best-known bounds, which are of shape $\ll 7^{\mathrm{rank}\,\mathrm{Jac}\,C_f(\mathbb{Q})}$, are not strong enough to be controlled by 2-descent. We do so using the Kabatiansky-Levenshtein bound on sizes of spherical codes. Because the argument works in full generality, rather than only in the particular case of genus two curves with a marked rational Weierstrass point, we present it instead in Chapter 6. The bound we get is

$$\#|C(K)^{\mathrm{large}}| \ll 1.872^{\mathrm{rank}\,\mathrm{Jac}\,C(K)},$$

where "large" indicates that we are restricting to points of large (i.e. $\gg h(C)$ using e.g. the tricanonical embedding $C \hookrightarrow \mathbb{P}^{5g-6}$) height, and the base can be improved from 1.872 to 1.311 once the genus g of C/K is sufficiently large.

Thus ends our discussion of finiteness theorems.

1.2 Part II.

Secure in our knowledge that $C(K)$ is indeed finite when C/K is a smooth projective hyperbolic curve, we may again ask: is there a finite-time algorithm that, on input $(K, C/K)$ with $K/\mathbb{Q}$ a number field and C/K a smooth projective hyperbolic curve, outputs $C(K)$?

In Chapter 7, which is based on joint work with Brian Lawrence, we answer: yes, assuming standard conjectures about motives. Specifically, we produce an algorithm that takes input $(K, C/K)$ and, if it terminates, indeed outputs $C(K)$. However, to prove it terminates, we must assume standard conjectures: the Fontaine-Mazur conjecture, the Grothendieck-Serre conjecture, the Tate conjecture, and the absolute Hodge conjecture. This is, in a very weak sense, reminiscent of the descent algorithm to compute the rank of an elliptic curve $E/\mathbb{Q}$, which, if it terminates, outputs the correct answer, whereas to prove the algorithm does terminate one assumes the finiteness of $Ш_{E/\mathbb{Q}}[2^\infty]$. However the latter algorithm actually terminates in practice, whereas no effort is made — here or in any other part of this thesis — to remotely optimize runtimes.

In Chapter 8 we explore the idea of substituting standard modularity conjectures for the above motivic conjectures. Specifically we use a trick of Bogomolov-Tschinkel, and the observation that the triangle group $\Delta(2, 6, 6)$ is arithmetic, to reduce the question of finding the rational points on a hyperelliptic curve to that of finding those Jacobians of genus two curves over a number field K with good reduction everywhere. Assuming a bijection between isogeny classes of non-potentially-CM fake elliptic curves over K and characters of a Hecke algebra acting on degree-$[K : \mathbb{Q}]$ cohomology of the locally symmetric space associated to GL_2/K, we explain how the latter question can be solved in finite time. What results is an algorithm to find the rational points on a hyperelliptic curve assuming standard, but completely out of reach (because one is forced to work over arbitrary $K/\mathbb{Q}$), modularity conjectures.

Even so, in Chapter 9 we succeed in producing from the ideas of Chapter 8 an algorithm for a restricted class of curves which is completely unconditional. We prove that there is a finite-time algorithm that, on input $(\mathfrak{o}, K, S)$, with $\mathfrak{o}$ an order in a totally real field $\operatorname{Frac}\mathfrak{o} =: F/\mathbb{Q}$, $K/\mathbb{Q}$ an odd-degree[7] totally real field, and S a finite set of places of K, outputs $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$, the finite set[8] of $[F : \mathbb{Q}]$-dimensional abelian varieties A/K with good reduction outside S and admitting a map $\mathfrak{o} \hookrightarrow \operatorname{End}_K(A)$. To do this we prove a result, which in the particular case of Galois representations associated to Hilbert modular forms is an explicit form of a theorem of Dimitrov, controlling explicitly in terms of F, K, and S the primes for which the corresponding two-dimensional Galois representation associated to such an A/K has small residual image. Using this, we then compute the finitely many extensions (which depend on A/K through its residual representation at the given prime, a representation with a priori bounded image and ramification) produced by Taylor's potential modularity theorem. Ultimately we reduce to the following argument, already observed by von Känel in the case $K = \mathbb{Q}$ when Serre's conjecture is known: if A/K is modular and has good reduction outside S, with corresponding parallel weight two Hilbert modular eigencuspform f over K, then A is a K-quotient of $\operatorname{Jac} C_{v_K}(\mathfrak{m})^{\times \dim A}/K$, where $v_K|\infty$ is a chosen place at infinity of K, $\mathfrak{m}$ is bounded explicitly in terms of S, and $C_{v_K}(\mathfrak{m})/K$ is the Shimura curve with full level-$\mathfrak{m}$ structure corresponding to the quaternion algebra ramified at exactly the infinite primes distinct from v_K. Thus $\operatorname{Jac} C_{v_K}(\mathfrak{m})^{\times \dim A} \sim_K A \times B$ by Poincaré complete reducibility, so that one produces a bound on $h(A)$ by using Bost's lower bound on $h(B)$ and the Masser-Wüstholz isogeny theorem. Ar-

[7] We discuss the even-degree case as well, but we leave the discussion of this case to Chapter 9 because the absolute Hodge conjecture intervenes in a mild way.

[8] Here, and for the rest of the thesis, we ignore stack-theoretic issues, since they are irrelevant for these Diophantine questions.

guably the key point in the chapter is the following trivial observation: the extensions guaranteed by Moret-Bailly's theorem are computable, since said theorem may be rephrased as the statement that a certain recursively enumerable set is nonempty. We note that, because Hilbert modular varieties are quasiprojective with zero-dimensional boundary, complete curves on them, and, a fortiori, mapping to them, abound, so that this gives an algorithm determining the rational points, over all odd-degree extensions, of a class of curves over, say, $\mathbb{Q}$. Unfortunately we lack a nontrivial criterion to characterize these curves!

In other words, it is not clear precisely which smooth projective hyperbolic curves admit families of GL_2-type abelian varieties defined, along with the relevant endomorphisms, over a totally real field. However there is a very large class of curves admitting such a family over a CM field, thanks to a construction of Cohen-Wolfart. In Chapter 10 we generalize the methods of Chapter 9 to treat exactly this class: the curves C/K over a CM field $K/\mathbb{Q}$ admitting a Belyi map $C \to \mathbb{P}^1$ defined over K with all ramification indices over $0, 1, \infty$ respectively divisible by positive integers $a, b, c \in \mathbb{Z}^+$ such that $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} < 1$. Using such a Belyi map one produces a non-isotrivial family of (hypergeometric) abelian varieties $A \to C$ of dimension $\varphi(N)$ and admitting $\mathbb{Z}[\zeta_N] \hookrightarrow \mathrm{End}_{C/K(\zeta_N)}(A)$. Thus by using this family we see that, to determine $C(K)$, it suffices to determine the finitely many $\varphi(N)$-dimensional abelian varieties A/K with good reduction outside S and admitting $\mathbb{Z}[\zeta_N] \hookrightarrow \mathrm{End}_{K(\zeta_N)}(A)$. Because of recent potential modularity results over CM fields, we may simply imitate Chapter 9 to compute a finite set Π of pairs (L, π), with $L/\mathbb{Q}(\zeta_N)$ CM and π a weight zero cuspidal automorphic representation of GL_2/L, such that all $P \in C(K)$ have fibre A_P/K corresponding to π after base change to L. However, because we are working over a CM field and not an odd-degree totally real field, where we could have applied the same Jacquet-Langlands transfer argument as above, currently there is no way of determining in finite time whether or not there is an abelian variety A_π/L associated to π, so that we are not able to compute $C(K)$ given Π unconditionally, though we certainly may by simply assuming e.g.[9] the existence of motives associated to such π.

Finally, in the spirit of the *Arithmetica*, we work out an example. For $K/\mathbb{Q}$ a totally real field of odd degree and $a \in K^\times$, we explain in detail in Chapter 11 how the above techniques allow one to unconditionally compute $C_a(K)$, with $C_a : x^6 + 4y^3 = a^2$. Mimicking the technique we use in Chapter 10, we form the hypergeometric family $A \to C_a$ associated to the hypergeometric function ${}_2F_1\left(\begin{array}{cc|c} \frac{1}{6} & \frac{1}{3} & \lambda \\ & \frac{5}{6} & \end{array}\right)$ and arising from the arithmetic triangle group $\Delta(3,6,6)$ — explicitly, for $P =: (x,y)$ with $x^6 \neq 0, a^2, \infty$, A_P is the evident two-dimensional quotient of the Jacobian of the (desingularization of the) genus 3 curve $t^6 = s^4(1-s)^3\left(1 - \frac{x^6}{a^2} \cdot s\right)$. Thanks to an identity of finite-field analogues of hypergeometric functions, we find that the corresponding abelian surface A_P/K has quaternionic multiplication over $K(\zeta_3)$ and indeed is of GL_2-type over K. Moreover, because the family does not degenerate as $x^6 \to 0$, a^2, or ∞, it follows that the conductors of the various A_P/K for $P \in C(K)$ are explicitly bounded. Since we have already seen how to determine the $\mathrm{GL}_2(F)$-type abelian varieties over K with good reduction outside S given (F, K, S), it is an easy matter to conclude.

Let us get to the details.

1.3 Notation and logical (in)dependence of the chapters.

We set notation. By $f \ll_\theta g$, or equivalently $f \leq O_\theta(g)$, we mean that $|f| \leq C_\theta \cdot g$, where $C_\theta \in \mathbb{R}^+$ is a constant depending only on θ. Of course $f \gg_\theta g$, or equiv-

[9] We discuss other, likely weaker, hypotheses in Chapter 10.

alently $f \geq \Omega_\theta(g)$, simply means $g \ll_\theta f$. By $f \asymp_\theta g$, or equivalently $f = \Theta_\theta(g)$, we mean $f \ll_\theta g$ and $g \ll_\theta f$. By $f \leq o_{\theta\to\infty}(g)$ we mean that, for all $\varepsilon \in \mathbb{R}^+$, once $\theta \gg_\varepsilon 1$, we have that $|f| \leq \varepsilon \cdot g$. Similarly $f \geq \omega_{\theta\to\infty}(g)$ means $g \leq o_{\theta\to\infty}(f)$. In this thesis all constants left implicit will be effective.

To be precise, to say that C_θ depends effectively on θ means for us that there is a finite-time algorithm, i.e. a Turing machine that terminates on all inputs, that computes the function $\theta \mapsto C_\theta$. Because it seems this standard definition is not completely universal[10], we have largely avoided the use of the word "effective" below and used instead the terms "effectively computable", or, equivalently, "computable", to mean the same.

We also use the word "explicit", which does not have a precise mathematical meaning, besides implying effective computability. We use the standard meaning: $\theta \mapsto C_\theta$ is explicit if it is not only effectively computable, but, informally, "one could also actually write a formula down in terms of θ if one wanted to".

We will use arithmetic Frobenii and give cyclotomic characters Hodge-Tate weight -1, though we assure the reader that there will be no delicate calculations actually relying on these conventions.

Finally we comment that the chapters that follow are logically independent from one another, except that if we state a standard result in one chapter we will not restate it in another, that Chapter 5 will refer to Chapter 6, and that Chapters 10 and 11 will refer to Chapter 9.

[10] For example, the phrase "effective Mordell conjecture" does *not* usually refer to an effective form of Faltings' Theorem, but rather something closer to the abc conjecture and thus immeasurably stronger.

Part I

Upper bounds.

Chapter 2

$\mathcal{E}^{\text{aff.}}_{A,B}(\mathfrak{o}_{K,S})$: S-integral points on elliptic curves.

This chapter is based on joint work with Wei Ho [7].

Abstract.

Let K be a number field. Let $A, B \in \mathfrak{o}_K$ be such that $\Delta_{A,B} := -16(4A^3 + 27B^2) \neq 0$. Let S be a finite set of places of K containing all infinite places and all primes $\mathfrak{p}$ for which $\mathfrak{p}^2|\Delta_{A,B}$. Let $\mathcal{E}^{\text{aff.}}_{A,B} : y^2 = x^3 + Ax + B$ be the affine Weierstrass model of the elliptic curve $E_{A,B}/K$. In this chapter we prove:

$$\#|\mathcal{E}^{\text{aff.}}_{A,B}(\mathfrak{o}_{K,S})| \ll 2^{\operatorname{rank} E_{A,B}(K)} \cdot O(1)^{\#|S|} \cdot \#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|.$$

In the discriminant aspect this improves the best-known bound, due to Helfgott-Venkatesh [55], from 2006. The proof is an exercise in optimizing a technique introduced by Mordell in 1914 [70][1].

[1] It is amusing to note that this paper was, famously, rejected. It was eventually published in a journal featuring, just two papers before Mordell's, a paper titled "Tables of $1 \pm 2^{-n} + 3^{-n} \pm 4^{-n} +$&c. and $1 + 3^{-n} + 5^{-n} + 7^{-n} +$ &c. to 32 places of decimals" [52].

2.1 Introduction.

2.1.1 Main theorem.

In this chapter we prove the following theorem. The proof is quite straightforward: the technique can be summarized as making explicit Mordell's classic proof of finiteness of $\mathcal{E}^{\mathrm{aff.}}_{A,B}(\mathfrak{o}_{K,S})$ and applying a bound due to Evertse [45] in the endgame.[2]

Theorem 2.1.1. *Let $C := 7^{2^8}$. Let K be a number field. Let $A, B \in \mathfrak{o}_K$ be such that $\Delta_{A,B} := -16(4A^3 + 27B^2) \neq 0$. Let S be a finite set of places of K containing all infinite places and all primes $\mathfrak{p}$ for which $\mathfrak{p}^2 | \Delta_{A,B}$. Let $\mathcal{E}^{\mathrm{aff.}}_{A,B} : y^2 = x^3 + Ax + B$ be the affine Weierstrass model of the elliptic curve $E_{A,B}/K$. Then:*

$$\#|\mathcal{E}^{\mathrm{aff.}}_{A,B}(\mathfrak{o}_{K,S})| \leq 2^{\operatorname{rank} E_{A,B}(K)} \cdot C^{\#|S|+1} \cdot \#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|.$$

2.1.2 A corollary.

It is routine to deduce from this (and the bounds of Bhargava-Shankar [19–22], though see Bhargava-Shankar-Wang [23] for the same theorems over number fields) the following corollary.

Corollary 2.1.2. *Let K be a number field. Let S be a finite set of places of K containing all infinite places of K. Then:*

$$\operatorname*{Avg}_{E/K, H(E) \leq X} \#|E(\mathfrak{o}_{K,S})|^{\frac{\log 5}{\log 2}} \ll_{K,\#|S|} 1.$$

Here $H(E)$ denotes the absolute Faltings height of E, though the same statement in terms of the usual naïve height follows as well (note that for convenience one may

[2] One can optimize the implicit constants by applying various different bounds in the literature, but we have not bothered because it seems that in the best case one would simply improve a constant that is number theoretical to one that is merely astronomical.

take S sufficiently large so that $\mathrm{Cl}(\mathfrak{o}_{K,S}) = 0$ in order to easily define quasi-minimal affine Weierstrass models etc.).

The only insight required in proving Corollary 2.1.2 from Theorem 2.1.1 is that

$$\#|\{\mathfrak{p} \mid \Delta_E : \mathrm{Nm}\,\mathfrak{p} \geq H(E)^\varepsilon\}| \ll \frac{1}{\varepsilon}.$$

In other words, one can ignore large primes dividing the discriminant, after which the analysis is straightforward.

2.2 Proof of Theorem 2.1.1.

Proof. It is classical (see e.g. Lemma 2 of Birch and Swinnerton-Dyer's [24]) that:

$$\mathrm{Sel}_2(E_{A,B}/K) \simeq \{F \in \mathfrak{o}_K[X,Y] : F \text{ homogeneous}, I(F) = -48A, J(F) = -1728B,$$
$$Z^2 = F(X,Y) \text{ everywhere locally soluble}\}/\mathrm{PGL}_2(K)$$

as sets, where I and J are the classical invariants of a binary quartic — specifically, writing $F(X,Y) =: a \cdot X^4 + b \cdot X^3Y + c \cdot X^2Y^2 + d \cdot XY^3 + e \cdot Y^4$,

$$I(F) := 12ae - 3bd + c^2$$

and

$$J(F) := 72ace - 27ad^2 - 27b^2e + 9bcd - 2c^3.$$

Thus e.g. the discriminant of F satisfies

$$\Delta(F) = \frac{1}{27}(4I^3 - J^2).$$

We note that PGL_2 acts on the space of binary quartics by

$$(\gamma \cdot F)(X, Y) := (\det \gamma)^{-2} \cdot F((X, Y) \cdot \gamma).$$

Of course the action of $\mathrm{PGL}_2(K)$ does not preserve integrality of the coefficients, but its action still defines an equivalence relation on binary quartics in $\mathfrak{o}_K[X, Y]$, and this is what we implicitly use to define the quotient.

It is also classical (see e.g. Theorem 3.2 of Bhargava-Shankar's [21], though the result is older) that the composition

$$E_{A,B}(K) \twoheadrightarrow E_{A,B}(K)/2 \hookrightarrow \mathrm{Sel}_2(E_{A,B}/K)$$

is simply

$$(x, y) \mapsto X^4 - 6x \cdot X^2Y^2 + 8y \cdot XY^3 + (-4A - 3x^2) \cdot Y^4 \pmod{\mathrm{PGL}_2(K)}.$$

It is evident that the image of $\mathcal{E}^{\mathrm{aff.}}_{A,B}(\mathfrak{o}_{K,S}) \subseteq E_{A,B}(K)$ lies inside $E_{A,B}(K)/2 \hookrightarrow \mathrm{Sel}_2(E_{A,B}/K)$. Because $\#|E_{A,B}(K)/2| = \#|E_{A,B}(K)[2]| \cdot 2^{\mathrm{rank}\, E_{A,B}(K)} \leq 4 \cdot 2^{\mathrm{rank}\, E_{A,B}(K)}$, it follows that, to prove the theorem, it suffices to show that the fibres of the map $\mathcal{E}^{\mathrm{aff.}}_{A,B}(\mathfrak{o}_{K,S}) \to \mathrm{Sel}_2(E_{A,B}/K)$ are of size $\leq 7^{2^7} \cdot C^{\#|S|} \cdot \#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|$.

We introduce the following nomenclature.

Definition 2.2.1. *Let R be a domain with fraction field L. Let $F \in L[X, Y]$ be a binary quartic. Write*

$$F(X, Y) =: \sum_{i=0}^{4} a_i \cdot X^{4-i}Y^i.$$

- *F is* R-integral *if and only if $a_i \in R$ for all i.*

- *F is* R-integer-matrix *if and only if $a_i \in \binom{4}{i} \cdot R$ for all i.*

- *F is* demonic *if and only if $a_0 = 1$ and $a_1 = 0$.*

*We abbreviate $\mathfrak{o}_{K,S}$-integral and $\mathfrak{o}_{K,S}$-integer-matrix as S-*integral *and S-*integer-matrix*, respectively.*

By inspection, evidently the image of $\mathcal{E}^{\text{aff.}}_{A,B}(\mathfrak{o}_{K,S}) \to \mathrm{Sel}_2(E_{A,B}/K)$ lies inside the ($\mathrm{PGL}_2(K)$-equivalence classes of) S-integer-matrix demonic binary quartic forms.

Therefore it suffices to prove the following bound.

Proposition 2.2.2. *Let K be a number field. Let S be a finite set of places of K containing all infinite places of K. Let $F(X,Y) = X^4 + a_2 \cdot X^2Y^2 + a_3 \cdot XY^3 + a_4 \cdot Y^4 \in \mathfrak{o}_{K,S}[X,Y]$ be a demonic S-integer-matrix binary quartic form such that $\Delta(F)$ is squarefree in $\mathfrak{o}_{K,S}$.*[3] *Then:*

$$\#|\{\gamma \in \mathrm{PGL}_2(K) : \gamma \cdot F \textit{ demonic } S\textit{-integer-matrix}\}| \leq 7^{2^7} \cdot C^{\#|S|} \cdot \#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|.$$

To do this we simply follow our nose. Let $F \in \mathfrak{o}_{K,S}[X,Y]$ be a demonic S-integer-matrix binary quartic. Write $F(X,Y) =: X^4 + 6a_2 \cdot X^2Y^2 + 4a_3 \cdot XY^3 + a_4 \cdot Y^4$ with $a_i \in \mathfrak{o}_{K,S}$. Let $\gamma \in \mathrm{PGL}_2(K)$. Write $\gamma =: \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in \mathfrak{o}_K$. We simply expand $(\gamma \cdot F)(X,Y)$ to find the following.

Lemma 2.2.3. *Let K be a number field. Let S be a finite set of places of K containing all infinite places of K. Let $F(X,Y) = X^4 + a_2 \cdot X^2Y^2 + a_3 \cdot XY^3 + a_4 \cdot Y^4 \in \mathfrak{o}_{K,S}[X,Y]$ be a demonic S-integer-matrix binary quartic. Let $\gamma \in \mathrm{PGL}_2(K)$ be such that $\gamma \cdot F$ is demonic and S-integer-matrix. Write $\gamma =: \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in \mathfrak{o}_{K,S}$. Then:*

1. *$(a,b) = (a,b,c,d)$ as ideals of $\mathfrak{o}_{K,S}$, and*

2. *$F(a,b) = (\det \gamma)^2$ divides $\Delta(F) \cdot (a,b)^4$ in $\mathfrak{o}_{K,S}$.*

Remark 2.2.4. *Note that in the above lemma the hypothesis that $\Delta(F)$ is squarefree in $\mathfrak{o}_{K,S}$ does* not *appear. However, in our situation $\Delta(F)$ is squarefree in $\mathfrak{o}_{K,S}$, and this*

[3] This can always be arranged by enlarging S, of course. We state the proposition in this manner because it corresponds to the situation of the theorem: S contains all $\mathfrak{p}$ for which $\mathfrak{p}^2|\Delta(F)$.

has the following consequence in the setup of the above lemma. Evidently (since F is a homogeneous quartic) $F(a,b) \in (a,b)^4$, i.e. $(a,b)^4 \mid (F(a,b))$. Moreover, since $F(a,b) = (\det\gamma)^2$ and since $(\Delta(F))$ is squarefree in $\mathfrak{o}_{K,S}$, it follows that

$$(\det\gamma)^2 = (F(a,b)) \mid (\Delta(F)) \cdot (a,b)^4 \implies (F(a,b)) \mid (a,b)^4.$$

Therefore we have that $(\det\gamma)^2 = (F(a,b)) = (a,b)^4$. Hence by unique factorization we also have that

$$(\det\gamma) = (a,b)^2,$$

in other words that (a,b) represents a 2-torsion class in $\mathrm{Cl}(\mathfrak{o}_{K,S})$. This is the source of the factor of $\#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|$ in the bound of Proposition 2.2.2.

In any case, let us turn to the proof of Lemma 2.2.3. Because the lemma is invariant under scaling the tuple (a,b,c,d) and its claims are local, it suffices to prove the following.

Lemma 2.2.5. *Let R be a principal ideal domain with fraction field L. Let $F(X,Y) = X^4 + a_2X^2Y^2 + a_3XY^3 + a_4Y^4 \in R[X,Y]$ be a demonic R-integer-matrix binary quartic. Let $\gamma \in \mathrm{PGL}_2(L)$ be such that $\gamma \cdot F$ is demonic and R-integer-matrix. Write $\gamma =: \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in R$ such that $(a,b,c,d) = (1)$ as ideals of R. Then:*

1. *$(a,b) = (1)$ as ideals of R, and:*

2. *$F(a,b) = (\det\gamma)^2$ divides $\Delta(F)$ in R.*

Proof of Lemma 2.2.5. Write $(a,b) =: (g)$, so that there exist $\alpha,\beta \in R$ with $a = g\alpha$ and $b = g\beta$. Since $(\alpha,\beta) = (1)$, there are $\widetilde{\alpha},\widetilde{\beta} \in R$ such that $\alpha\widetilde{\alpha} - \beta\widetilde{\beta} = 1$. Let $\widetilde{\gamma} := \left(\begin{smallmatrix} \alpha & \beta \\ \widetilde{\beta} & \widetilde{\alpha} \end{smallmatrix}\right) \in \mathrm{SL}_2(R)$, and let $\eta := c\widetilde{\alpha} - d\widetilde{\beta} \in R$. Let

$$U := \gamma\widetilde{\gamma}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} \widetilde{\alpha} & -\beta \\ -\widetilde{\beta} & \alpha \end{pmatrix} = \begin{pmatrix} g & 0 \\ \eta & \frac{\det\gamma}{g} \end{pmatrix},$$

so that $\gamma = U\widetilde{\gamma}$.

We now show that g divides all the entries of U, namely $g^2 \mid \det\gamma$ and $g \mid \eta$. Let $\widetilde{F} := \widetilde{\gamma}\cdot F \in R[X,Y]$, with no twisting necessary since $\widetilde{\gamma} \in \mathrm{SL}_2(R)$. Note that $\widetilde{F}$ is R-integer-matrix, since the property of being R-integer-matrix is preserved by the action of $\mathrm{SL}_2(R)$. Write

$$\widetilde{F}(X,Y) =: \widetilde{a}_0 \cdot X^4 + \widetilde{a}_1 \cdot X^3Y + \widetilde{a}_2 \cdot X^2Y^2 + \widetilde{a}_3 \cdot XY^3 + \widetilde{a}_4 \cdot Y^4 \in R[X,Y].$$

Then $(\gamma\cdot F)(X,Y) = (\det\gamma)^{-2}\cdot(U\cdot\widetilde{F})(X,Y) = (\det\gamma)^{-2}\cdot\widetilde{F}\left(gX+\eta Y, \frac{\det\gamma}{g}Y\right)$. Expanding, we compute that the X^4-coefficient in $\gamma\cdot F$ is

$$(\gamma\cdot F)(1,0) = F(a,b) = g^4\cdot F(\alpha,\beta) = \frac{g^4\cdot\widetilde{a}_0}{(\det\gamma)^2}.$$

Since it is also 1 by hypothesis, we find that $\frac{(\det\gamma)^2}{g^4} = \widetilde{a}_0 \in R$. Thus g^4 divides $(\det\gamma)^2$, so g^2 divides $\det\gamma$. Now the X^3Y-coefficient of $\gamma\cdot F$ is

$$\frac{4g^3\eta\cdot\widetilde{a}_0 + g^2(\det\gamma)\cdot\widetilde{a}_1}{(\det\gamma)^2} = 0.$$

Substituting for $\widetilde{a}_0$, we find that $\widetilde{a}_1 = -4\cdot\frac{(\det\gamma)\cdot\eta}{g^3} \in 4R$ (since $\widetilde{f}$ is R-integer-matrix). Finally, the X^2Y^2-coefficient of $\gamma\cdot f$ is

$$\frac{6g^2\eta^2\cdot\widetilde{a}_0 + 3g\eta(\det\gamma)\cdot\widetilde{a}_1 + (\det\gamma)^2\cdot\widetilde{a}_2}{(\det\gamma)^2} = -\frac{6\eta^2}{g^2} + \widetilde{a}_2$$

after substituting for $\widetilde{a}_0$ and $\widetilde{a}_1$. Since $\widetilde{a}_2 \in 6R$ and this coefficient lies in $6R$ as well (since both are R-integer-matrix), we deduce that g^2 divides η^2, so g divides η.

Since g divides all the entries of U, we see that g divides all the entries of $U\cdot\widetilde{\gamma} = \gamma$, implying that g divides $(a,b,c,d) = (1)$, whence $(g) = (1)$, proving the first claim. Thus without loss of generality $g = 1$.

Now since $g = 1$ it follows that $\widetilde{a}_1 = -4\eta \det \gamma$, so $\det \gamma$ divides $\widetilde{a}_1$, and of course $(\det \gamma)^2$ divides $\widetilde{a}_0$. We thus find that $(\det \gamma)^2$ divides $\Delta(\widetilde{F}) = \Delta(F)$ (since every term of $\Delta(\widetilde{F})$ is a multiple of either $\widetilde{a}_0$ or $\widetilde{a}_1^2$). □

Thus we have proven Lemma 2.2.5, and hence (by localizing at all primes of $\mathfrak{o}_{K,S}$ and applying Lemma 2.2.5 to the resulting discrete valuation ring) Lemma 2.2.3.

We may now prove Proposition 2.2.2.

Proof of Proposition 2.2.2. The key point is a trick that we will use again in Chapter 4: we represent the relevant ideal classes of $\mathfrak{o}_{K,S}$ by primes to make the application of Evertse's bound more efficient.

Recall that $\mathrm{Cl}(\mathfrak{o}_{K,S})$ is a quotient of $\mathrm{Cl}(\mathfrak{o}_K)$. Let P be a set of prime ideals of $\mathfrak{o}_K$ for which the canonical map $P \to \mathrm{Cl}(\mathfrak{o}_{K,S})$, taking an element of P to its ideal class, is a bijection onto $\mathrm{Cl}(\mathfrak{o}_{K,S})[2]$, the 2-torsion subgroup of $\mathrm{Cl}(\mathfrak{o}_{K,S})$. Note that such a set of representatives exists by Chebotarev's density theorem applied to the Hilbert class field of K.

Lemma 2.2.3 shows that for any $\gamma \in \mathrm{PGL}_2(K)$ (represented by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathfrak{o}_{K,S}$) such that $\gamma \cdot F$ is demonic and S-integer-matrix, we have that $(a, b) = (a, b, c, d)$ as ideals of $\mathfrak{o}_{K,S}$ and $F(a, b) = (\det \gamma)^2$ is a square dividing $\Delta(F) \cdot (a, b, c, d)^4$ in $\mathfrak{o}_{K,S}$. As noted in Remark 2.2.4, it follows from the fact that $\Delta(F)$ is squarefree in $\mathfrak{o}_{K,S}$ that $(a, b)^2 = (\det \gamma)$.

Therefore, by our choice of P, there is a $\mathfrak{p} \in P$ and an $\alpha \in K^\times$ for which $\mathfrak{p} = \alpha \cdot (a, b, c, d)$. Since $\mathfrak{p} \subseteq \mathfrak{o}_K \subseteq \mathfrak{o}_{K,S}$ it follows that $\alpha a, \ldots, \alpha d \in \mathfrak{o}_{K,S}$. Scaling each of a, b, c, d through by α (note that this does not change $\gamma \in \mathrm{PGL}_2(K)$) we may without loss of generality assume that $(a, b, c, d) = (a, b) = \mathfrak{p}$ in $\mathfrak{o}_{K,S}$.

Thus, given $\gamma \in \mathrm{PGL}_2(K)$ for which $\gamma \cdot f$ is both demonic and S-integer-matrix, we get a pair $(a, b) \in \mathfrak{o}_{K,S}^2$, well defined up to the action of $\mathfrak{o}_{K,S}^\times$ (since γ is an equivalence class of matrices in $\mathrm{GL}_2(K)$ modulo scaling by $K^\times$, and we have pinned down the ideal (a,b) via $(a,b) = \mathfrak{p}$ with $\mathfrak{p} \in P$).

We now claim that the map

$$\begin{aligned}\Phi\colon &\{\gamma \in \mathrm{PGL}_2(K) : \gamma \cdot F \text{ is demonic and } S\text{-integer-matrix}\}\\ &\to \bigcup_{\mathfrak{p}\in P} \{(a,b) \in \mathfrak{o}_{K,S}^2 \mid (a,b) = \mathfrak{p}, (F(a,b)) = \mathfrak{p}^4\}/\mathfrak{o}_{K,S}^\times,\end{aligned}$$

taking γ as above to the equivalence class of (a,b), defined as above, is injective.

Indeed, if $\gamma, \gamma' \in \mathrm{PGL}_2(K)$ map to the same $(a,b) \in \mathfrak{o}_{K,S}^2/\mathfrak{o}_{K,S}^\times$, write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ and $\gamma' = \left(\begin{smallmatrix} a & b \\ c' & d'\end{smallmatrix}\right)$, and note that

$$\gamma'\gamma^{-1} = \begin{pmatrix} 1 & 0 \\ \frac{c'd - cd'}{\det\gamma} & 1 \end{pmatrix}.$$

Let $\lambda := \frac{c'd-cd'}{\det\gamma} \in K$. Since

$$(\gamma' \cdot f)(X,Y) = ((\gamma'\gamma^{-1}) \cdot (\gamma \cdot f))(X,Y) = (\gamma \cdot f)(X + \lambda Y, Y)$$

and both $\gamma \cdot f$ and $\gamma' \cdot f$ are demonic by hypothesis, it follows that $\lambda = 0$ and so $\gamma = \gamma'$, as desired.

Thus, the size of the domain of Φ is bounded by the size of the codomain of Φ. To bound the size of the codomain we do the following. Write

$$M_{\mathfrak{p}} := \{(a,b) \in \mathfrak{o}_{K,S}^2 \mid (a,b) = \mathfrak{p}, (F(a,b)) = \mathfrak{p}^4\}.$$

Thus the codomain is $\bigcup_{\mathfrak{p}\in T} M_{\mathfrak{p}}/\mathfrak{o}_{K,S}^\times$.

Of course $(F(a,b)) = \mathfrak{p}^4$ implies that $F(a,b) \in \mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^\times$. Thus we may upper bound the size of each term $M_{\mathfrak{p}}/\mathfrak{o}_{K,S}^\times$ as follows.

Note first that the canonical map $M_{\mathfrak{p}}/\mathfrak{o}_{K,S}^{\times} \to M_{\mathfrak{p}}/\mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times}$, taking equivalence classes of elements of $M_{\mathfrak{p}}$ modulo the diagonal action of $\mathfrak{o}_{K,S}^{\times}$ to equivalence classes modulo the action of the larger $\mathfrak{o}_{K,S\cup\{p\}}^{\times}$, is in fact a bijection. This is simply the fact that, given $\alpha \in K^{\times}$ and $(a,b) \in \mathfrak{o}_{K,S}$ for which $(a,b) = \alpha \cdot (a,b) = \mathfrak{p}$ as ideals of $\mathfrak{o}_{K,S}$, it follows that $\alpha \in \mathfrak{o}_{K,S}^{\times}$.

Next we enlarge $M_{\mathfrak{p}}$ as follows. Observe that

$$M_{\mathfrak{p}} \subseteq \{(a,b) \in \mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{2} : F(a,b) \in \mathfrak{o}_{K,S\cup\{p\}}^{\times}\},$$

and hence

$$M_{\mathfrak{p}}/\mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times} \subseteq \{(a,b) \in \mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times} : F(a,b) \in \mathfrak{o}_{K,S\cup\{p\}}^{\times}\}/\mathfrak{o}_{K,S\cup\{p\}}^{\times}.$$

Now we invoke Theorem 3 of Evertse's [45]:

Theorem 2.2.6 (Theorem 3 of Evertse's [45].)**.** *Let K be a number field. Let S be a finite set of places of K containing all infinite places of K. Let $F \in \mathfrak{o}_{K,S}[X,Y]$ be a homogeneous polynomial with at least three distinct roots in $\mathbb{P}^1(\overline{\mathbb{Q}})$. Then:*

$$\#|\{(x,y) \in \mathfrak{o}_{K,S}^{2} : F(x,y) \in \mathfrak{o}_{K,S}^{\times}\}/\mathfrak{o}_{K,S}^{\times}| \leq 7^{(\deg F)^3 \cdot ([K:\mathbb{Q}] + 2\cdot\#|S|)}.$$

Applying Theorem 2.2.6 with binary form F and set of places $S \cup \{\mathfrak{p}\}$, we find

$$\begin{aligned}
&\#|M_{\mathfrak{p}}/\mathfrak{o}_{K,S}^{\times}| \\
&= \#|M_{\mathfrak{p}}/\mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times}| \\
&\leq \#|\{(a,b) \in \mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{2} : F(a,b) \in \mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times}\}/\mathfrak{o}_{K,S\cup\{\mathfrak{p}\}}^{\times}| \\
&\leq 7^{2^7} \cdot C^{\#|S|}.
\end{aligned}$$

Summing this over $\mathfrak{p} \in P$ (and recalling that $\#|P| = \#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|$) to bound the size of the codomain of Φ, we find that:

$$\#|\{\gamma \in \mathrm{PGL}_2(K) : \gamma{\cdot}f \text{ is demonic and } S\text{-integer-matrix}\}| \leq 7^{2^7}{\cdot}C^{\#|S|}{\cdot}\#|\mathrm{Cl}(\mathfrak{o}_{K,S})[2]|,$$

as desired. □

Thus we have proven Proposition 2.2.2. As already noted, Theorem 2.1.1 follows. □

Chapter 3

$E_{0,B}(\mathbb{Q}), E_{0,B^2}(\mathbb{Q})$: quadrics in arithmetic statistics.

Abstract.

In this chapter we introduce a trick that allows one to apply Bhargava's counting technique to count points on invariant quadrics. As example applications we give easy arguments bounding the average sizes of 2-Selmer groups in the families $E_{0,B} : y^2 = x^3 + B$ and $E_{0,B^2} : y^2 = x^3 + B^2$. The fundamental idea, which is to use the circle method to easily obtain precise counts in the "bulk", is due to Ruth in his Princeton PhD thesis [87]. Our contribution is a trick that trivializes the counting in the "tail". We also introduce a smoothed form of Bhargava's counting method, as well as a trick that allows us to deduce what the Selmer average along the invariant quadric must be from knowledge of the corresponding unconstrained average.

3.1 Introduction.

In this chapter we introduce a trick that allows one to use Bhargava's counting technique to count points on invariant quadrics. We note that, up until now, with the sole exception of Ruth's Princeton PhD thesis [87], all applications of Bhargava's counting technique have been to counting points in various affine spaces — in other words, they have ultimately reduced the question to counting lattice points in sufficiently round subsets of Euclidean space, traditionally handled using Davenport's standard lemma.[1]

We detail this trick by way of treating two example applications. Write $E_{A,B} : y^2 = x^3 + Ax + B$. We will only work over $\mathbb{Q}$ for simplicity.

3.1.1 Main theorems.

The first example application is a generalization of the main theorem of Ruth's Princeton PhD thesis [87] (see Theorem $1.1.2$ of [87]) to sets defined by congruence conditions.

Theorem 3.1.1. *Let $\mathcal{B} \subseteq \mathbb{Z} - \{0\}$ be a set of positive upper density defined by congruence conditions. Then:*

$$\underset{B \in \mathcal{B} : |B| \leq X}{\mathrm{Avg}} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q})| \leq 3 + O_{\mathcal{B}}(o_{X \to \infty}(1)).$$

Here we do not allow a "congruence" at ∞ (i.e. a specification of the sign of B) to spare ourselves the notation, but such a result also follows from our technique. In fact we also prove equality in the above when e.g. $\mathcal{B}$ is defined by finitely many congruence conditions — or more generally is "large" in a sense analogous to that of Bhargava-Shankar's [21] — but rather than discuss the precise hypotheses on $\mathcal{B}$

[1] We point out that, using the smoothed form of Bhargava's counting method described in Section 3.2, the relevant lattice point count follows from Poisson summation rather than Davenport.

needed for equality in the above statement we have left the theorem as a general upper bound. Nonetheless we will explain how to prove the relevant uniformity estimate, and thus the matching lower bound for such $\mathcal{B}$, in the course of the proof.

We comment that we use the circle method to deduce that the right-hand side is a product of local densities, and then we use knowledge of the average of $\#|\mathrm{Sel}_2(E_{A,B}/\mathbb{Q})|$ over *both* A and B (namely, that it is $3 + o_{X\to\infty}(1)$, by Bhargava-Shankar's [21]) to calculate said product without any local calculations — specifically, we deduce that the right-hand side is $3 + o_{X\to\infty}(1)$ as well.

As in Bhargava-Shankar's [21], the problem immediately reduces to a question about the number of binary quartic forms F over $\mathbb{Z}$ with bounded (classical) invariants $I(F)$ and $J(F)$ — however, because we are only dealing with the elliptic curves $E_{0,B}$ (i.e. $A = 0$), we are only interested in those binary forms F with $I(F) = 0$. Of course, $I(F)$ is a quadratic form in five variables, namely the coefficients of F.[2]

The second application is to a question which arises in forthcoming joint work with Manjul Bhargava and Ari Shnidman [6]. Note that the family is now $y^2 = x^3 + B^2$.

Theorem 3.1.2. *Let $\mathcal{B} \subseteq \mathbb{Z} - \{0\}$ be a set of positive upper density defined by congruence conditions. Then:*

$$\underset{B\in\mathcal{B}:|B|\leq X}{\mathrm{Avg}} \#|\mathrm{Sel}_2(E_{0,B^2}/\mathbb{Q})| \leq 3 + O_{\mathcal{B}}(o_{X\to\infty}(1)).$$

We note that we again employ the smoothed form of Bhargava's counting method described in Section 3.2, and we again evaluate (sans calculation) the product of local densities that arises by repeating the trick of passing from our

[2] Writing $F(X,Y) =: a\cdot X^4 + b\cdot X^3Y + c\cdot X^2Y^2 + d\cdot XY^3 + e\cdot Y^4$,

$$I(F) = 12ae - 3bd + c^2.$$

family $y^2+2B\cdot y = x^3$ to a larger one — specifically, the family $y^2+a_2\cdot xy+a_6\cdot y = x^3$ — and using 2-Selmer average results on the larger family which do not employ the circle method (see part (f) of Theorem 1.1 in Bhargava-Ho's [18] — the relevant parametrization is by triply symmetric hypercubes). We also again prove the uniformity estimate required to deduce equality in the theorem for "large" $\mathcal{B}$ in the course of the proof.

Here the relevant thing to count is no longer binary quartic forms F over $\mathbb{Z}$ with bounded invariants $I(F)$ and $J(F)$ with $I(F) = 0$ as in Theorem 3.1.1, but rather pairs of binary cubic forms[3] $\vec{F} := (F_1, F_2)$ over $\mathbb{Z}$ with bounded invariants $I_2(\vec{F}), I_6(\vec{F})$, where I_2 and I_6 are explicit polynomials in the coefficients of the F_i that are invariant under the natural $\mathrm{SL}_2 \times \mathrm{SL}_2$ action, with one SL_2 acting via changes of variable and the other SL_2 acting on the vector (F_1, F_2).[4]

Moreover, the condition that these arise from a curve of the form E_{0,B^2} is precisely the condition that $I_2(\vec{F}) = 0$. This is a quadric in eight variables, namely the coefficients of the F_i.

3.1.2 Main technique.

Having stated the main theorems, let us discuss the method of proof. In order to be specific, let us put ourselves in the situation of the first theorem: we would like to count, up to $\mathrm{PGL}_2(\mathbb{Q})$-equivalence, the binary quartic forms $F \in \mathbb{Z}[X, Y]$ with $I(F) = 0$ and $0 \neq |J(F)| \leq X$. Write then $V := \mathrm{Sym}^4(2)$ for the space of binary

[3] To reconcile this parametrization with the usual parametrization of 2-Selmer elements of elliptic curves by binary quartics, note that the discriminant form $\mathrm{disc}_{X,Y}(x \cdot F_1(X, Y) + y \cdot F_2(X, Y)) \in \mathbb{Z}[x, y]$ is a binary quartic form associated to the pair of binary cubic forms (F_1, F_2) with $F_i \in \mathbb{Z}[X, Y]$.

[4] Of course $I_2(\vec{F})$ is simply the invariant alternating bilinear form on the space of binary cubics — specifically, writing $F_i(X, Y) =: a_i \cdot X^3 + b_i \cdot X^2Y + c_i \cdot XY^2 + d_i \cdot Y^3$,

$$I_2(\vec{F}) = 3a_1d_2 - b_1c_2 + c_1b_2 - 3d_1a_2.$$

quartic forms. It is now standard that, using a method introduced in Bhargava's Princeton PhD thesis [14] and first carried out in this context by Bhargava-Shankar [21] (let us ignore our smoothing for the sake of this discussion), the problem immediately reduces to the problem of obtaining an asymptotic of the following form:

$$\int_{1\ll\lambda\ll X^{\frac{1}{24}}} d^\times\lambda \int_{|u|\ll 1} du \int_{1\ll t\ll\lambda} t^{-2} d^\times t \,\#|\{F\in\lambda\cdot n_u\cdot a_t\cdot G_0\cdot L\cap V(\mathbb{Z})^{\text{irred.}} : I(F)=0\}|$$
$$\sim \text{const.}\cdot X^{\frac{1}{2}}.$$

Here $n_u := \left(\begin{smallmatrix}1 & 0\\ u & 1\end{smallmatrix}\right)$, $a_t := \left(\begin{smallmatrix}t^{-1} & 0\\ 0 & t\end{smallmatrix}\right)$, and $d^\times z := \frac{dz}{z}$. We note that $d^\times\lambda\, du\, t^{-2} d^\times t\, dk$ is the Haar measure on $\mathrm{GL}_2^+(\mathbb{R})$ (the "+" denoting positive determinant) induced by Haar measures on Λ, N, A, K under the Iwasawa decomposition $\mathrm{GL}_2^+(\mathbb{R}) = \Lambda\cdot N\cdot A\cdot K$, where Λ are the scalars, N are the lower unipotents, A is the diagonal torus, and $K := \mathrm{SO}_2(\mathbb{R})$ is the maximal compact.

For the sake of exposition, the reader should imagine $G_0\cdot L\subseteq V(\mathbb{R})$ as a small ball of binary quartic forms of height $\asymp 1$, and indeed should imagine (though also keep in mind that this is a slight oversimplification) that the set $\lambda\cdot n_u\cdot a_t\cdot G_0\cdot L\cap V(\mathbb{Z})^{\text{irred.}}$ is simply the set $S_{\text{approx.}}$, say, of $(a,b,c,d,e)\in\mathbb{Z}$ satisfying $a, e\neq 0$ and:

$$|a| \ll \frac{\lambda^4}{t^4},$$
$$|b| \ll \frac{\lambda^4}{t^2},$$
$$|c| \ll \lambda^4,$$
$$|d| \ll t^2\cdot\lambda^4,$$
$$|e| \ll t^4\cdot\lambda^4.$$

Examining the integral, one evidently reduces to needing to treat

$$\#|\{F\in\lambda\cdot n_u\cdot a_t\cdot B\cdot L\cap V(\mathbb{Z})^{\text{irred.}} : I(F)=0\}|.$$

Were there no condition $I(F)=0$, this would essentially be counting lattice points in a somewhat "round" set scaled by a parameter λ, which is simple — one invokes an easy lemma of Davenport, which is precisely tailored for this sort of counting problem, in the standard treatment. Indeed, examining the defining inequalities of $S_{\text{approx.}}$, it is evidently quite simple to evaluate $\#|S_{\text{approx.}}|$.

Thus the entire difficulty is dealing with the condition $I(F)=0$. However we are simply asking to count zeroes of a quadratic form in five variables that lie in the scalings of a somewhat "round" set, and this is easy if the set is in fact quite "round" — in other words, if we are in the "bulk" and $1 \ll t \ll_\varepsilon \lambda^\varepsilon$, say, then the defining inequalities of $S_{\text{approx.}}$ are all, up to $X^{O(\varepsilon)}$, the same, so that $S_{\text{approx.}}$ is essentially a scaled cube, and it is straightforward to count the number of solutions of a quadratic form in five variables lying in such a set, by the oldest forms of the circle method. For convenience we follow Ruth and use the smoothed delta symbol method, which trivializes the problem. In the end we obtain the asymptotic

$$\#|\{F \in \lambda \cdot n_u \cdot a_t \cdot B \cdot L \cap V(\mathbb{Z})^{\text{irred.}} : I(F)=0\}| \sim \text{const.} \cdot \lambda^{12} + O_\varepsilon(\lambda^{12-\delta+O(\varepsilon)})$$

for a positive absolute constant $\delta > 0$.

Thus the entire problem has reduced to treating the count when we are in the "tail", i.e. $\lambda^\varepsilon \ll_\varepsilon t \ll \lambda$. Here Ruth uses a difficult argument that again involves the smoothed delta symbol method and is quite specific to the situation — so much so that it is unclear if his argument generalizes from $\mathcal{B} = \mathbb{Z} - \{0\}$ to sets defined by finitely many congruence conditions, let alone to other quadratic forms.

Our observation is that the "tail" case is also obviously trivial. Specifically, it is obvious that

$$\#|\{F \in \lambda \cdot n_u \cdot a_t \cdot B \cdot L \cap V(\mathbb{Z})^{\text{irred.}} : I(F)=0\}| \ll \lambda^{12+o(1)}$$

by the divisor bound: $12ae - 3bd + c^2 = 0$ implies that $a, e \,|\, 3bd - c^2$. Because we are dealing with elements in $V(\mathbb{Z})^{\text{irred.}}$ (and thus $a, e \neq 0$), it follows that (b, c, d) determine (a, e) up to $\ll \lambda^{o(1)}$ choices, and the number of (b, c, d) is[5] $\ll \lambda^{12}$.

Because of the $o(1)$ in the exponent, this bound is not quite as sharp as we needed in the "bulk", but since we are in the "tail" we needn't work so hard. The point is that the condition $t \gg_\varepsilon \lambda^\varepsilon$ combines with the $t^{-2} d^\times t$ in the Haar measure (note that the exponent 2 is more than is needed for convergence, so to speak) to imply that this bound is enough to give a bound $\ll_\varepsilon \lambda^{12-\Omega(\varepsilon)}$ on the "tail" contribution after integrating over u and t.

So in the end we obtain the desired asymptotic, which is equivalent to the theorem via Bhargava's counting method. The argument in the case of pairs of binary cubic forms is the same, and indeed the above argument easily generalizes to other quadrics in at least four variables (so that the circle method analysis in the "bulk" goes through easily).

Having described the method, let us now prove the theorems. We begin with our remark on smoothing in Bhargava's counting method.

3.2 Smoothing in Bhargava's counting method.

Let us now introduce the aforementioned technical convenience that simplifies Bhargava's counting method, as first introduced in Chapter 5 of his Princeton PhD thesis [14], though the trick of averaging over fundamental domains was first introduced in the published version (see Section 2.2 of Bhargava's [15]). The point is that, while an average over G_0 (using the notation of the above outline) improves the situation considerably, one is still integrating a "rough" function, namely $\mathbb{1}_{G_0}$,

[5] While we have seemingly used the special form of $I(F)$ in this argument, this divisor bound argument works in general — the point is that a binary quadratic form represents a nonzero integer few times, because of a divisor bound in at most a quadratic extension.

and it is wiser to instead integrate a compactly supported smooth function. We note that this is completely natural from Bhargava's original formulation — see equation (4) in Section 2.2 of Bhargava's [15], and note that we are taking (in his notation) Φ to be smooth and compactly supported, rather than the indicator function of a box.

In order to be specific, and for the reader's convenience, let us work in the setup of the proof of Theorem 3.1.1 (it will be clear how to modify the construction for Theorem 3.1.2, and indeed in any application of Bhargava's counting method). That is, $V := \mathrm{Sym}^4(2), G := \mathrm{GL}_2$, $L := L^{(0)} \coprod L^{(1)} \coprod L^{(2+)} \coprod L^{(2-)}$ with

$$L^{(0)} := \left\{ X^3Y - \frac{1}{3} \cdot XY^3 + \frac{J}{27} \cdot Y^4 : J \in (-2,2) \right\},$$
$$L^{(1)} := \left\{ X^3Y - \frac{I}{3} \cdot XY^3 \pm \frac{2}{27} \cdot Y^4 : I \in [-1,1) \right\} \cup \left\{ X^3Y + \frac{1}{3} \cdot XY^3 + \frac{J}{27} \cdot Y^4 : J \in (-2,2) \right\},$$
$$L^{(2\pm)} := \pm \left\{ \frac{1}{16} X^4 - \sqrt{\frac{2-J}{27}} \cdot X^3Y + \frac{1}{2} \cdot X^2Y^2 + Y^4 : J \in (-2,2) \right\},$$

and

$$\mathcal{F} := \left\{ \lambda \cdot n_u \cdot a_t \cdot k : \lambda \in \mathbb{R}^+, u \in \nu(t) \subseteq \left[-\frac{1}{2}, \frac{1}{2} \right], t \geq \sqrt{\frac{\sqrt{3}}{2}}, k \in \mathrm{SO}_2(\mathbb{R}) \right\} \subseteq G(\mathbb{R}),$$

Gauss's classical fundamental domain for $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$. Note that L is a fundamental domain for $G(\mathbb{R}) \curvearrowright V(\mathbb{R})^{\Delta \neq 0}$. Observe that the points with $I = 0$ and $\Delta \neq 0$ are all in the $G(\mathbb{R})$-orbit of the two forms $F_\pm(X,Y) := X^3Y \pm \frac{2}{27} \cdot Y^4$, which lie in the interior of $L^{(1)}$ (and thus lie in small compact subsets thereof). Thus the following setup suffices for us.

Write, for each $v \in V(\mathbb{R})^{\Delta \neq 0}$, $v_L \in L$ for the unique element of L mapping to the image of v under $V(\mathbb{R})^{\Delta \neq 0} \to V(\mathbb{R})^{\Delta \neq 0}/G(\mathbb{R}) \simeq L$.

Let $\alpha \in C_c^\infty(G(\mathbb{R}))$ and $\beta \in C_c^\infty(L)$ be compactly supported smooth functions such that: α is $\mathrm{SO}_2(\mathbb{R})$-invariant, $\int_{G(\mathbb{R})} \alpha = 1$, $\beta(F_\pm) = 1$, and $\mathrm{supp}\,\beta, \beta^{-1}(\{1\}) \subseteq L^{(1)}$

are both unions of two small compact intervals respectively containing $F_\pm$. Let

$$\varphi(v) := \sum_{g \cdot v_L = v} \alpha(g) \cdot \beta(v_L).$$

Note that this is a finite sum because stabilizers of elements of $V(\mathbb{R})^{\Delta \neq 0}$ are finite.

Via α and β we get a slightly more convenient way[6] to smooth out the various integrals in Bhargava's counting technique — instead of integrating the normalized indicator function $\frac{1}{\int_{G_0} dg} \cdot \mathbb{1}_{G_0}$ of G_0 over $g \in G(\mathbb{R})$ (i.e. integrating over $g \in G_0$) and observing that $g \cdot \mathcal{F}$ is also (the closure of) a fundamental domain for $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$ so that all counts are independent of g, we instead integrate $\alpha(g)$ over $g \in G$ and then make the same observation.

Specifically, we observe that, since

$$\#|\{F \in \mathcal{F} \cdot g \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}} : I(F) = 0, 0 \neq |J(F)| \leq X\}|$$

is constant in g (since $\mathcal{F} \cdot g$ and $\mathcal{F}$ are both fundamental domains for $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$ — the first main observation of Section 2.2 of Bhargava's [15]), it follows that:

$$\frac{\int_{g \in G_0} dg \, \#|\{F \in \mathcal{F} \cdot g \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}} : I(F) = 0, 0 \neq |J(F)| \leq X\}|}{\int_{g \in G_0} dg}$$
$$= \int_{g \in G} dg \, \alpha(g) \cdot \#|\{F \in \mathcal{F} \cdot g \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}} : I(F) = 0, 0 \neq |J(F)| \leq X\}|.$$

The left-hand side is precisely Ruth's $N(Y(\mathbb{Z})^{\text{nontriv.}}, X)$.

We then manipulate this expression just as in Section 2.3 of Bhargava-Shankar's [21] (and implicitly in Bhargava's [15]).

[6] Specifically, this insertion of a smooth weight in Bhargava's main trick in his counting technique saves us the effort required to remove smooth weights when applying the smoothed delta method. We note here that we use β to ensure smoothness of φ — note that $L^{(0)} \coprod L^{(1)}$ is a rectangle missing two corners, and at the other two corners a similar definition of φ with $\beta = 1$ identically would fail to be smooth. One can get around this, of course, but this choice simplifies notation.

Evidently:

$$\begin{aligned}&\int_{g\in G} dg\,\alpha(g)\cdot\#|\{F\in\mathcal{F}\cdot g\cdot L\cap V(\mathbb{Z})^{\text{nontriv.}}:I(F)=0,0\neq|J(F)|\leq X\}|\\&=\sum_{v\in V(\mathbb{Z})^{\text{nontriv.}}:I(v)=0,0\neq|J(v)|\leq X}\int_{G(\mathbb{R})}dh\,\alpha(h)\cdot\#|\{g\in\mathcal{F}:gh\cdot v_L=v\}|.\end{aligned}$$

But, using that $\beta(v_L)=1$ if $I(v)=0$ and $\Delta(v)\neq 0$ (since this implies that $v_L=F_+$ or F_-), each integral can be evaluated as follows:

$$\begin{aligned}&\int_{G(\mathbb{R})}dh\,\alpha(h)\cdot\#|\{g\in\mathcal{F}:gh\cdot v_L=v\}|\\&=\sum_{\gamma\in G(\mathbb{R}):\gamma\cdot v_L=v}\int_{G(\mathbb{R})}dh\,\alpha(h)\cdot\beta(v_L)\cdot\#|\{g\in\mathcal{F}:gh=\gamma\}|\\&=\sum_{\gamma\in G(\mathbb{R}):\gamma\cdot v_L=v}\int_{\mathcal{F}^{-1}\cdot\gamma}dh\,\alpha(h)\cdot\beta(v_L)\\&=\sum_{\gamma\in G(\mathbb{R}):\gamma\cdot v_L=v}\int_{\mathcal{F}^{-1}}dh\,\alpha(h\cdot\gamma)\cdot\beta(v_L)\\&=\int_{\mathcal{F}}dh\sum_{\gamma\cdot v_L=v}\alpha(h^{-1}\cdot\gamma)\cdot\beta(v_L)\\&=\int_{\mathcal{F}}dh\sum_{g\cdot v_L=h^{-1}\cdot v}\alpha(g)\cdot\beta(v_L)\\&=\int_{\mathcal{F}}dh\,\varphi(h^{-1}\cdot v),\end{aligned}$$

by definition.

Therefore we have found that:

$$\begin{aligned}N(Y(\mathbb{Z})^{\text{nontriv.}},X)&=\sum_{v\in V(\mathbb{Z})^{\text{nontriv.}}:I(v)=0,0\neq|J(v)|\leq X}\int_{\mathcal{F}}dh\,\varphi(h^{-1}\cdot v)\\&=\int_{\mathcal{F}}dh\sum_{v\in V(\mathbb{Z})^{\text{nontriv.}}:I(v)=0,0\neq|J(v)|\leq X}\varphi(h^{-1}\cdot v).\end{aligned}$$

The point is that a sum of shape $\sum_{v\in V(\mathbb{Z})}\Phi(v)$, or else e.g. $\sum_{v\in V(\mathbb{Z}):I(v)=0}\Phi(v)$, is much easier to evaluate if Φ is smooth and compactly supported. We note that in the above calculation we did not use anything about β besides that $\beta(v_L)=1$ for the $v\in V(\mathbb{Z})$ we were interested in counting. This suggests that one might, by doing the above, replace the use of Davenport's standard lemma (which is traditionally used to evaluate $\sum_{v\in V(\mathbb{Z})^{\text{nontriv.}}:H(v)\leq X}\varphi(h^{-1}\cdot v)$ when $\alpha=\mathbb{1}_{G_0}$ and $\beta=1$) in usual treatments of the theorems of arithmetic statistics with an application of Poisson summation. However, we repeat that, were $\operatorname{supp}\beta$ not small, φ would a priori not be smooth, so that there would be some subtlety in carrying such an argument out and obtaining strong error terms.

In any case we will use the above calculation to save ourselves the work of unsmoothing the below arguments.

3.3 Proof of Theorem 3.1.1.

Now let us turn to the proof of Theorem 3.1.1.

3.3.1 Reduction to point counting.

For the reader's convenience we use Ruth's notation in what follows. Let $V:=\mathrm{Sym}^4(2)$. Let $G:=\mathrm{GL}_2$. Let, for $F\in V$,

$$I(F):=12ae-3bd+c^2,$$

$$J(F):=72ace+9bcd-27ad^2-27b^2e-2c^3,$$

so that the discriminant

$$\Delta(F)=4I^3-J^2,$$

where we have written $F(X,Y) =: a \cdot X^4 + b \cdot X^3Y + c \cdot X^2Y^2 + d \cdot XY^3 + e \cdot Y^4$. Let $G \curvearrowright V$ via $(g \cdot F)(X,Y) := F((X,Y) \cdot g)$. We note that, for $g \in G$ and $F \in V$, we have that:

$$I(g \cdot F) = (\det g)^4 \cdot I(F),$$
$$J(g \cdot F) = (\det g)^6 \cdot J(F).$$

Let $L := L^{(0)} \coprod L^{(1)} \coprod L^{(2+)} \coprod L^{(2-)}$ with

$$L^{(0)} := \left\{ X^3Y - \frac{1}{3} \cdot XY^3 + \frac{J}{27} \cdot Y^4 : J \in (-2,2) \right\},$$
$$L^{(1)} := \left\{ X^3Y - \frac{I}{3} \cdot XY^3 \pm \frac{2}{27} \cdot Y^4 : I \in [-1,1) \right\} \cup \left\{ X^3Y + \frac{1}{3} \cdot XY^3 + \frac{J}{27} \cdot Y^4 : J \in (-2,2) \right\},$$
$$L^{(2\pm)} := \pm \left\{ \frac{1}{16}X^4 - \sqrt{\frac{2-J}{27}} \cdot X^3Y + \frac{1}{2} \cdot X^2Y^2 + Y^4 : J \in (-2,2) \right\},$$

a fundamental domain for $G(\mathbb{R}) \curvearrowright V(\mathbb{R})^{\Delta \neq 0}$.

Let

$$\mathcal{F} := \left\{ \lambda \cdot n_u \cdot a_t \cdot k : \lambda \in \mathbb{R}^+, u \in \nu(t), t \geq \sqrt{\frac{\sqrt{3}}{2}}, k \in \mathrm{SO}_2(\mathbb{R}) \right\} \subseteq G(\mathbb{R}),$$

where $\nu(t) \subseteq [-\frac{1}{2}, \frac{1}{2}]$ is $[-\frac{1}{2}, \frac{1}{2})$ when $t \gg 1$ or else a union of two subintervals of $[-\frac{1}{2}, \frac{1}{2}]$ when $\sqrt{\frac{\sqrt{3}}{2}} \leq t \ll 1$ (just imagine the usual Gauss fundamental domain for $\mathrm{SL}_2(\mathbb{Z}) \curvearrowright \mathfrak{h}$, and note that t^2 corresponds to $\mathfrak{Im}\,\tau$ and u corresponds to $\mathfrak{Re}\,\tau$). Here we have written

$$n_u := \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix},$$
$$a_t := \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix}.$$

We note that $\mathcal{F}$ is a fundamental domain for $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$.

Let $\alpha \in C_c^\infty(G(\mathbb{R}))$ and $\beta \in C_c^\infty(L)$ be compactly supported smooth functions such that: α is $\mathrm{SO}_2(\mathbb{R})$-invariant, $\int_{G(\mathbb{R})} \alpha = 1$, $\beta(F_\pm) = 1$, and $\operatorname{supp} \beta, \beta^{-1}(\{1\}) \subseteq L^{(1)}$ are both unions of two small compact intervals respectively containing $F_\pm$. Let $G_0 := \operatorname{supp} \beta$. Let

$$\varphi(v) := \sum_{g \cdot v_L = v} \alpha(g) \cdot \beta(v_L).$$

Write $V(\mathbb{Z})^{\Delta \neq 0} := \{F \in V(\mathbb{Z}) : \Delta(F) \neq 0\}$ and

$$V(\mathbb{Z})^{\mathrm{nontriv.}} := \{F \in V(\mathbb{Z})^{\Delta \neq 0} : 0 \notin F(\mathbb{P}^1(\mathbb{Q}))\}.$$

(Here we deviate from Ruth, and indeed Bhargava [14] and Bhargava-Shankar [21] in using the superscript $\mathrm{nontriv.}$ instead of $\mathrm{irred.}$, since in our view it is misleading to call these irreducible.) That is, $V(\mathbb{Z})^{\mathrm{nontriv.}}$ is the subset of $V(\mathbb{Z})$ consisting of binary quartic forms with no root in $\mathbb{P}^1(\mathbb{Q})$ — i.e., those binary quartics that do not have a linear factor defined over $\mathbb{Q}$.

Write

$$Y(\mathbb{Z}) := \{F \in V(\mathbb{Z}) : I(F) = 0, J(F) \neq 0\}$$

and $Y(\mathbb{Z})^{\mathrm{nontriv.}} := Y(\mathbb{Z}) \cap V(\mathbb{Z})^{\mathrm{nontriv.}}$. We note that our $Y(\mathbb{Z})$ and $Y(\mathbb{Z})^{\mathrm{nontriv.}}$ play the role of Ruth's Y and $Y^{\mathrm{irr.}}$, respectively. Similarly, for $M \in \mathbb{Z}^+$ and $F_0 \in V(\mathbb{Z}/M)$, write

$$Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z}) := \{F \in Y(\mathbb{Z}) : F \equiv F_0 \pmod{M}\}$$

and $Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})^{\mathrm{nontriv.}} := Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z}) \cap V(\mathbb{Z})^{\mathrm{nontriv.}}$.

Write, for $S \subseteq V(\mathbb{Z})$,

$$\#_\varphi |B(u, t, \lambda, X) \cap S| := \sum_{F \in S : |J(F)| \leq X} \varphi(a_t^{-1} \cdot n_u^{-1} \cdot (\lambda \cdot \mathrm{id})^{-1} \cdot F),$$

where we have disambiguated the action of λ (which should really be the action of $\left(\begin{smallmatrix}\lambda & 0\\ 0 & \lambda\end{smallmatrix}\right)$) by writing $\lambda \cdot \mathrm{id}$ for clarity. We note that this is different from Ruth's (and indeed Bhargava's) notation precisely because we have smoothed using α and β rather than simply $\mathbb{1}_{G_0}$ — though the reader may well imagine that $\varphi \sim \mathbb{1}_{G_0 \cdot L}$, in which case $\#_\varphi |B(u,t,\lambda,X) \cap S| \sim \#|\lambda \cdot n_u \cdot a_t \cdot G_0 \cdot L \cap S|$.

Just as in Section 2.2 of Ruth's [87], we find that the problem reduces to controlling $\#_\varphi |B(u,t,\lambda,X) \cap Y_{F_0 \pmod M}(\mathbb{Z})^{\text{nontriv.}}|$. We do this with the following two lemmas. Note that Lemma 3.3.2 matches Ruth's Proposition 2.3.1 in form, except that we have included a congruence condition in order to sieve to locally soluble forms — Ruth overlooks doing this in his circle method analysis, but in any case the argument is similarly straightforward. We note also that Ruth overlooks a factor of the form $\sigma_\infty(u,t,\lambda,X)$ (specifically he overlooks the dependence on both λ and X — on page 12 he states that the condition $|J(v)| < X$ is superfluous once $\lambda \ll X^{\frac{1}{24}}$, thus one can drop it — this is false, since his definition of his $\mathcal{F}'$ depends on a parameter C' and were this to be the case the Selmer average would also grow with C', rather than be 3). Again, given his key idea of using the circle method in this context, fixing these small details is simple.

Lemma 3.3.1 (The "tail" estimate.). *Let*

$$\lambda \in \mathbb{R}^+, u \in \left[-\frac{1}{2}, \frac{1}{2}\right], \sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda.$$

Then:

$$\#_\varphi |B(u,t,\lambda,X) \cap Y(\mathbb{Z})^{\text{nontriv.}}| \ll_\varphi \lambda^{12+o(1)}.$$

Lemma 3.3.2 (The "bulk" estimate.). *Let $M \in \mathbb{Z}^+$ and $F_0 \in V(\mathbb{Z}/M)$. Let*

$$\lambda \in \mathbb{R}^+, u \in \left[-\frac{1}{2}, \frac{1}{2}\right], \sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda.$$

Then:

$$\#_{\varphi}|B(u,t,\lambda,X)\cap Y_{F_0\,(\mathrm{mod}\ M)}(\mathbb{Z})| = \sigma_{\infty}(u,t,\lambda,X)\cdot\prod_{p}\sigma_p(Y_{F_0\,(\mathrm{mod}\ M)}(\mathbb{Z})) \\ + O_{\varphi,M}(t^4\cdot\lambda^{8+o(1)}) + O_{\varphi,M,N}(t^N\cdot\lambda^{O(1)-N}),$$

where

$$\sigma_{\infty}(u,t,\lambda,X) := \lim_{\varepsilon\to 0}\frac{\int_{v\in V(\mathbb{R}):|I(v)|\le\varepsilon,|J(v)|\le X} dv\,\varphi(a_t^{-1}\cdot n_u^{-1}\cdot(\lambda\cdot\mathrm{id})^{-1}\cdot v)}{2\varepsilon}$$

and

$$\sigma_p(Y_{F_0\,(\mathrm{mod}\ M)}(\mathbb{Z})) := \lim_{k\to\infty} p^{-4k}\cdot\#|\{F\in V(\mathbb{Z}/p^k) : I(F)\equiv 0\ (\mathrm{mod}\ p^k), F\equiv F_0\ (\mathrm{mod}\ M)\}|.$$

We have written $\sigma_{\infty}(u,t,\lambda,X)$ despite the function being independent of u and t (via $v\mapsto n_u\cdot a_t\cdot v$) for notational convenience.

Just as in e.g. Theorems 2.12 and 2.21 of Bhargava-Shankar's [21], we must introduce a weight function $\phi : V(\mathbb{Z}/M)\to\mathbb{R}$ (which will eventually be taken to be a majorant of, in their notation, $f\mapsto\frac{1}{m(f)}$) with M highly divisible. The following weighted version of Lemma 3.3.2 of course follows immediately from Lemma 3.3.2.

Lemma 3.3.3. *Let $M\in\mathbb{Z}^+$ and $\phi : V(\mathbb{Z}/M)\to\mathbb{R}$. Let*

$$\lambda\in\mathbb{R}^+, u\in\left[-\frac{1}{2},\frac{1}{2}\right], \sqrt{\frac{\sqrt{3}}{2}}\le t\ll\lambda.$$

Then:

$$\sum_{F_0\in V(\mathbb{Z}/M)}\phi(F_0)\cdot\#_{\varphi}|B(u,t,\lambda,X)\cap Y_{F_0\,(\mathrm{mod}\ M)}(\mathbb{Z})|$$

$$= \sigma_\infty(u,t,\lambda,X) \cdot \prod_{p \nmid M} \sigma_p(Y(\mathbb{Z})) \cdot \lim_{k\to\infty} M^{-4k} \cdot \sum_{F \in V(\mathbb{Z}/M^k) : I(F) \equiv 0 \pmod{M^k}} \phi(F \pmod{M})$$
$$+ O_{\varphi,M}(||\phi||_1 \cdot t^4 \cdot \lambda^{8+o(1)}) + O_{\varphi,M,N}(||\phi||_1 \cdot t^N \cdot \lambda^{O(1)-N}).$$

We note that the p-adic local densities are of course exact analogues of the singular integral at infinity — one thickens p-adically by allowing $I(F) \equiv 0 \pmod{p^k}$ only, and then one takes a limit. It is because of this thickening that we easily reduce the calculation of the constants to results of Bhargava-Shankar.

Let us first deduce Theorem 3.1.1 from Lemmas 3.3.1 and 3.3.2.

Proof of Theorem 3.1.1 assuming Lemmas 3.3.1 and 3.3.2. We reduce immediately to the case of $\mathcal{B}$ defined by finitely many congruence conditions. Indeed, assume Theorem 3.1.1 for nonempty subsets of $\mathbb{Z} - \{0\}$ defined by finitely many congruence conditions. Writing $\mathcal{B}_p$ for the closure of $\mathcal{B}$ in $\mathbb{Z}_p$, and $\mathcal{B}_{\leq T} := \{n \in \mathbb{Z} - \{0\} : \forall p \leq T, n \in \mathcal{B}_p\}$ (thus $\mathcal{B} \subseteq \mathcal{B}_{\leq T}$), we find that, assuming Theorem 3.1.1 for sets defined by finitely many congruence conditions (and thus for $\mathcal{B}_{\leq T}$):

$$\sum_{B \in \mathcal{B} : |B| \leq X} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q})|$$
$$\leq \sum_{B \in \mathcal{B}_{\leq T} : |B| \leq X} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q})|$$
$$\leq (3 + O_{\mathcal{B},T}(o_{X\to\infty}(1))) \cdot \#|\{n \in \mathcal{B}_{\leq T} : |n| \leq X\}|$$
$$= (3 + O_{\mathcal{B},T}(o_{X\to\infty}(1))) \cdot \left(\frac{\#|\{n \in \mathcal{B}_{\leq T} : |n| \leq X\}|}{\#|\{n \in \mathcal{B} : |n| \leq X\}|} \right) \cdot \#|\{n \in \mathcal{B} : |n| \leq X\}|$$
$$= (3 + O_{\mathcal{B}}(o_{T\to\infty}(1)) + O_{\mathcal{B},T}(o_{X\to\infty}(1))) \cdot \#|\{n \in \mathcal{B} : |n| \leq X\}|.$$

Taking $X \to \infty$ and then $T \to \infty$ gives that

$$\operatorname*{Avg}_{n \in \mathcal{B} : |n| \leq X} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q})| \leq 3 + O_{\mathcal{B}}(o_{X\to\infty}(1)),$$

as desired.

Thus without loss of generality $\mathcal{B}$ is defined by finitely many congruence conditions, i.e. it is of the form $\mathcal{B} = \{n \in \mathbb{Z} - \{0\} : n \equiv a \pmod{m}\}$ for $m \in \mathbb{Z}^+$ and $a \in \mathbb{Z}/m$.

We will appeal to Ruth's Lemma 2.2.3 (proven in Section 4.3 of his [87] using an analysis of the (monogenized) cubic resolvent ring arising from a binary quartic form) to use the equality $n(F) = m(F)$ outside a negligible set. Here we use the notation of Section 3.2 of Bhargava-Shankar's [21]: $n(F)$ is the number of $\mathrm{PGL}_2(\mathbb{Z})$-orbits in the (intersection of $V(\mathbb{Z})$ and the) $\mathrm{PGL}_2(\mathbb{Q})$-orbit of $F \in V(\mathbb{Z})^{\Delta \neq 0}$, and

$$\begin{aligned} m(F) &:= \sum_{F' \in \mathrm{PGL}_2(\mathbb{Z}) \backslash (V(\mathbb{Z}) \cap \mathrm{PGL}_2(\mathbb{Q}) \cdot F)} \frac{\#|\operatorname{Aut}_{\mathrm{PGL}_2(\mathbb{Q})}(F')|}{\#|\operatorname{Aut}_{\mathrm{PGL}_2(\mathbb{Z})}(F')|} \\ &= \prod_p \sum_{F' \in \mathrm{PGL}_2(\mathbb{Z}_p) \backslash (V(\mathbb{Z}_p) \cap \mathrm{PGL}_2(\mathbb{Q}_p) \cdot F)} \frac{\#|\operatorname{Aut}_{\mathrm{PGL}_2(\mathbb{Q}_p)}(F')|}{\#|\operatorname{Aut}_{\mathrm{PGL}_2(\mathbb{Z}_p)}(F')|} \\ &=: \prod_p m_p(F). \end{aligned}$$

Let $M \in \mathbb{Z}^+$ with $m|M$. Let $\phi : V(\mathbb{Z}/M) \to \mathbb{R}$. We will eventually take a sequence of majorants ϕ_n of $F \mapsto \frac{1}{m(F)}$ — i.e. $\phi_1(F) \geq \cdots \geq \phi_n(F) \geq \cdots \geq \frac{1}{m(F)}$ for all $F \in V(\mathbb{Z})^{\Delta \neq 0}$ — and apply the below to ϕ_n and take $n \to \infty$.

As we have seen in e.g. Section 3.2,

$$\begin{aligned} &N(Y_{F_0 \pmod{M}}(\mathbb{Z})^{\text{nontriv.}}, X) \\ &= \int_{\mathcal{F}} dh \sum_{v \in V(\mathbb{Z})^{\text{nontriv.}} : I(v) = 0, F \equiv F_0 \pmod{M}} \varphi(h^{-1} \cdot v) \\ &= \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{u \in \nu(t)} du \int_{\sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda} t^{-2} d^\times t \, \#_\varphi |B(u,t,\lambda,X) \cap Y_{F_0 \pmod{M}}(\mathbb{Z})^{\text{nontriv.}}| \\ &= \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{u \in \nu(t)} du \int_{\sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda^\delta} t^{-2} d^\times t \, \#_\varphi |B(u,t,\lambda,X) \cap Y_{F_0 \pmod{M}}(\mathbb{Z})^{\text{nontriv.}}| \\ &\quad + \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{|u| \leq \frac{1}{2}} du \int_{\lambda^\delta \ll t \ll \lambda} t^{-2} d^\times t \, \#_\varphi |B(u,t,\lambda,X) \cap Y_{F_0 \pmod{M}}(\mathbb{Z})^{\text{nontriv.}}| \end{aligned}$$

$$
\begin{aligned}
&= \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{u \in \nu(t)} du \int_{\sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda^\delta} t^{-2} d^\times t \, \begin{pmatrix} \sigma_\infty(u,t,\lambda,X) \cdot \prod_p \sigma_p(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})) \\ +O_{\varphi,M}(t^4 \cdot \lambda^{8+o(1)}) \end{pmatrix} \\
&\quad + \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{|u| \leq \frac{1}{2}} du \int_{\lambda^\delta \ll t \ll \lambda} t^{-2} d^\times t \ O_\varphi(\lambda^{12+o(1)}) \\
&= \prod_p \sigma_p(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})) \cdot \int_{1 \ll \lambda \ll X^{\frac{1}{24}}} d^\times \lambda \int_{u \in \nu(t)} du \int_{\sqrt{\frac{\sqrt{3}}{2}} \leq t \ll \lambda^\delta} t^{-2} d^\times t \, \sigma_\infty(u,t,\lambda,X) \\
&\quad + O_{\varphi,M}(X^{1-\Omega(\delta)+o(1)}) \\
&= \prod_p \sigma_p(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})) \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{\mathcal{F}} d^\times \lambda \, du \, t^{-2} d^\times t \int_{v \in V(\mathbb{R}) : |I(v)| \leq \varepsilon, |J(v)| \leq X} dv \, \varphi((\lambda \cdot a_t \cdot n_u)^{-1} \cdot v) \\
&\quad + O_{\varphi,M}(X^{1-\Omega(\delta)+o(1)}) \\
&= \prod_p \sigma_p(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})) \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{v \in V(\mathbb{R}) : |I(v)| \leq \varepsilon, |J(v)| \leq X} dv \int_{h \in \mathcal{F}} dh \, \varphi(h^{-1} \cdot v) \\
&\quad + O_{\varphi,M}(X^{1-\Omega(\delta)+o(1)}),
\end{aligned}
$$

where each step of the above is either by definition, by Lemmas 3.3.1 and 3.3.2, or simply rearranging integrals or limits.

Recall that we saw in Section 3.2 that, for $v \in V(\mathbb{R})^{\Delta \neq 0}$ with $I(v) = 0$ (recall that then $\beta(v_L) = 1$),

$$
\int_{h \in \mathcal{F}} dh \, \varphi(h^{-1} \cdot v) = \int_{G(\mathbb{R})} dh \, \alpha(h) \cdot \#|\{g \in \mathcal{F} : gh \cdot v_L = v\}|.
$$

Because we arranged that $\beta = 1$ on sufficiently small intervals around $F_\pm$, the same identity holds for $v \in V(\mathbb{R})^{\Delta \neq 0}$ with $|I(v)| \leq \varepsilon$ and $|J(v)| \gg 1$ when ε is sufficiently small. Inserting this into the above, we find that the main term is:

$$
\begin{aligned}
&(1 + O_M(X^{-1})) \cdot N(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})^{\mathrm{nontriv.}}, X) \\
&\ = \prod_p \sigma_p(Y_{F_0 \ (\mathrm{mod}\ M)}(\mathbb{Z})) \\
&\qquad \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{v \in V(\mathbb{R}) : |I(v)| \leq \varepsilon, 1 \ll |J(v)| \leq X} dv \int_{G(\mathbb{R})} dh \, \alpha(h) \cdot \#|\{g \in \mathcal{F} : gh \cdot v_L = v\}|
\end{aligned}
$$

$$
\begin{aligned}
&= \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \\
&\qquad \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{G(\mathbb{R})} dh\, \alpha(h) \int_{v \in V(\mathbb{R}) : |I(v)| \le \varepsilon, 1 \ll |J(v)| \le X} dv\, \#|\{g \in \mathcal{F} : gh \cdot v_L = v\}| \\
&= \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{G(\mathbb{R})} dh\, \alpha(h) \int_{v \in \mathcal{F} \cdot h \cdot L : |I(v)| \le \varepsilon, 1 \ll |J(v)| \le X} dv \\
&= \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \cdot \lim_{\varepsilon \to 0} (2\varepsilon)^{-1} \cdot \int_{v \in \mathcal{F} \cdot L : |I(v)| \le \varepsilon, 1 \ll |J(v)| \le X} dv \\
&= 2X \cdot \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \cdot \lim_{\varepsilon \to 0} (4\varepsilon \cdot X)^{-1} \cdot \int_{v \in \mathcal{F} \cdot L : |I(v)| \le \varepsilon, 1 \ll |J(v)| \le X} dv \\
&= \frac{2X}{m} \cdot \left(m \cdot \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \right) \cdot \lim_{\varepsilon \to 0} (4\varepsilon \cdot X)^{-1} \cdot \int_{v \in \mathcal{F} \cdot L : |I(v)| \le \varepsilon, 1 \ll |J(v)| \le X} dv,
\end{aligned}
$$

where in the third-to-last equality we use that the inner integral is independent of h (since $\mathcal{F} \cdot h$ is also a fundamental domain of $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$) and that $\int_{G(\mathbb{R})} \alpha = 1$. Note that the factor $(1 + O_M(X^{-1}))$ in the first line comes from the fact that we have thrown out the $v \in V(\mathbb{R})^{\Delta \neq 0}$ with $|I(v)| \le \varepsilon$ and $|J(v)| \ll 1$ in order to use the equality $\int_{h \in \mathcal{F}} dh\, \varphi(h^{-1} \cdot v) = \int_{G(\mathbb{R})} dh\, \alpha(h) \cdot \#|\{g \in \mathcal{F} : gh \cdot v_L = v\}|$.

Note that, by Proposition 2.8 of Bhargava-Shankar's [21] (we could avoid using this to get from the leftmost to the rightmost terms in the equality, of course), we have that

$$
\begin{aligned}
(4\varepsilon \cdot X)^{-1} \cdot \int_{v \in \mathcal{F} \cdot L : |I(v)| \le \varepsilon, |J(v)| \le X} dv &= \frac{2\zeta(2)}{27} \cdot (1 + O(\varepsilon \cdot X^{-1})) \\
&= (1 + O(\varepsilon \cdot X^{-1})) \cdot \frac{\int_{\coprod_i \mathcal{R}^{(i)}(X)} dv}{\int_{\coprod_i R^{(i)}(X)} dI dJ},
\end{aligned}
$$

using the notation of Section 2.4 of Bhargava-Shankar's [21]. This is the first step in the trick of reducing the calculation of the product of local densities to the one done in Bhargava-Shankar — at the moment, we have only dealt with the Archimedean restrictions.

So far we have found that:

$$N(Y_{F_0 \pmod M}(\mathbb{Z})^{\text{nontriv.}}, X) = (1+O_M(X^{-1}))\cdot\frac{2X}{m}\cdot\left(m\cdot\prod_p \sigma_p(Y_{F_0 \pmod M}(\mathbb{Z}))\right)\cdot\frac{\int_{\coprod_i \mathcal{R}^{(i)}(X^2/4)} dv}{\int_{\coprod_i R^{(i)}(X^2/4)} dI dJ}.$$

Note that

$$\#|\{J \in \mathcal{B} : 0 \neq |J| \leq X\}| = \frac{2X}{m}\cdot\left(1+O\left(\frac{m}{X}\right)\right).$$

Just as in the passage from Lemma 3.3.2 to Lemma 3.3.3, we find that:

$$\begin{aligned}
&\sum_{F_0 \in V(\mathbb{Z}/M) : J(F_0) \equiv a \pmod m} \phi(F_0)\cdot N(Y_{F_0 \pmod M}(\mathbb{Z})^{\text{nontriv.}}, X)\\
&= (1+O(X^{-1}))\cdot\frac{2X}{m}\cdot\frac{\int_{\coprod_i \mathcal{R}^{(i)}(X^2/4)} dv}{\int_{\coprod_i R^{(i)}(X^2/4)} dI dJ}\\
&\quad\cdot\left(m\cdot\lim_{T\to\infty} n_T^{-4}\cdot\sum_{F\in V(\mathbb{Z}/n_T) : I(F)\equiv 0 \pmod{n_T}, J(F)\equiv a \pmod m}\phi(F \pmod M)\right),
\end{aligned}$$

where $n_T := \prod_{p\leq T} p^T$, say, and without loss of generality $T \geq M$ so that $M|n_T$.

Write $\mathbb{1}_{\text{sol.}}^{(p)}$ for the indicator function of the $\mathbb{Q}_p$-soluble binary quartic forms $F \in V(\mathbb{Z}_p)$ (that is to say, those F for which $Z^2 = F(X,Y)$ admits a nonzero solution in $\mathbb{Q}_p$). Write $\mathbb{1}_{\text{loc. sol.}} := \prod_p \mathbb{1}_{\text{sol.}}^{(p)}$. Write

$$\phi_*(F) := \frac{\mathbb{1}_{\text{loc. sol.}}}{m(F)} = \prod_p \frac{\mathbb{1}_{\text{sol.}}^{(p)}}{m_p(F)} =: \prod_p \phi_*^{(p)}(F)$$

on $V(\mathbb{Z})^{\Delta\neq 0}$. Thus when $1 =: \phi_0^{(p)}(F) \geq \phi_1^{(p)}(F) \geq \cdots \geq \phi_n^{(p)}(F) \geq \cdots \geq \phi_*^{(p)}(F)$ for all $F \in V(\mathbb{Z}_p)^{\Delta\neq 0}$ with $\phi_n^{(p)} : V(\mathbb{Z}_p)^{\Delta\neq 0} \to [0,1]$ factoring through $V(\mathbb{Z}/p^n)$ (and not

$V(\mathbb{Z}/p^{n-1})$) and such that $\phi_n^{(p)}(F) \to \phi_*^{(p)}(F)$ as $n \to \infty$, we find that[7], writing

$$\phi_n := \prod_{p \leq n} \phi_n^{(p)},$$

(and here is where we use Lemma 2.2.3 of Ruth's [87] to replace $m(F)$ by $n(F)$ for all but $\ll X^{\frac{5}{6}}$ forms):

$$\begin{aligned}
&\operatorname*{Avg}_{B \in \mathcal{B}: |B| \leq X} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q}) - \{0\}| \\
&\leq (1 + O_T(o_{X\to\infty}(1))) \cdot (1 + o_{T\to\infty}(1)) \\
&\quad \cdot \frac{\int_{\coprod_i \mathcal{R}^{(i)}(X^2/4)} dv}{\int_{\coprod_i R^{(i)}(X^2/4)} dI dJ} \cdot \left(m \cdot n_T^{-4} \cdot \sum_{F \in V(\mathbb{Z}/n_T) : I(F) \equiv 0 \ (\mathrm{mod}\ n_T), J(F) \equiv a \ (\mathrm{mod}\ m)} \phi_T(F) \right).
\end{aligned}$$

Note that we have "thickened" by changing the constraint $I(F) = 0$ to $I(F) \equiv 0 \ (\mathrm{mod}\ n_T)$. Accordingly, we write

$$\mathrm{Inv}^{(T,\mathcal{B})} := \{(I, J) \in \mathbb{Z}^2 : I \equiv 0 \ (\mathrm{mod}\ n_T), J \equiv a \ (\mathrm{mod}\ m), 4I^3 - J^2 \neq 0\},$$

and $\mathrm{Inv}_p^{(T,\mathcal{B})} \subseteq \mathbb{Z}_p^2$ for its closure in $\mathbb{Z}_p^2$.

The trick is now to notice that, for all $k \in \mathbb{Z}^+$,

$$\begin{aligned}
&\sum_{F \in V(\mathbb{Z}/n_T) : I(F) \equiv 0 \ (\mathrm{mod}\ n_T), J(F) \equiv a \ (\mathrm{mod}\ m)} \phi_T(F) \\
&= n_T^{-5\cdot(k-1)} \cdot \sum_{F \in V(\mathbb{Z}/n_T^k) : I(F) \equiv 0 \ (\mathrm{mod}\ n_T), J(F) \equiv a \ (\mathrm{mod}\ m)} \phi_T(F),
\end{aligned}$$

[7] Of course one always has that, for a convergent sequence $x_k \in \mathbb{R}$, $\lim_{k\to\infty} x_k = x_k + o_{k\to\infty}(1)$, but here we have written $\lim_{k\to\infty} x_k = (1 + o_{k\to\infty}(1)) \cdot x_k$, which is only justified (for k sufficiently large) when $\lim_{k\to\infty} x_k \neq 0$. While we will see that the relevant limit is 2, technically we are not yet justified in doing this and should carry the various additive error terms $o_{k\to\infty}(1)$ through the argument. However we hope the reader will grant us this notational simplification, since it makes no difference to the argument.

so that:

$$\begin{aligned}
&m \cdot n_T^{-4} \cdot \sum_{F \in V(\mathbb{Z}/n_T) : I(F) \equiv 0 \pmod{n_T}, J(F) \equiv a \pmod{m}} \phi_T(F) \\
&= \frac{n_T^{-5k} \cdot \sum_{F \in V(\mathbb{Z}/n_T^k) : I(F) \equiv 0 \pmod{n_T}, J(F) \equiv a \pmod{m}} \phi_T(F)}{n_T^{-2k} \cdot \#|\{(I,J) \in V(\mathbb{Z}/n_T^k) : I \equiv 0 \pmod{n_T}, J \equiv a \pmod{m}\}|}.
\end{aligned}$$

Now we note that

$$\begin{aligned}
&\lim_{k \to \infty} n_T^{-5k} \cdot \sum_{F \in V(\mathbb{Z}/n_T^k) : I(F) \equiv 0 \pmod{n_T}, J(F) \equiv a \pmod{m}} \phi_T(F) \\
&= \int_{F \in V\left(\prod_{p \leq T} \mathbb{Z}_p\right) : (I(F), J(F)) \in \prod_{p \leq T} \mathrm{Inv}_p^{(T,\mathcal{B})}} dF \, \phi_T(F),
\end{aligned}$$

and so, by dominated convergence and Fubini (note that we are implicitly using Proposition 3.18 of Bhargava-Shankar's [21], and indeed arguing as in their proof of Proposition 2.21 of their [21]), it follows that[8]

$$\begin{aligned}
&n_T^{-5k} \cdot \sum_{F \in V(\mathbb{Z}/n_T) : I(F) \equiv 0 \pmod{n_T}, J(F) \equiv a \pmod{m}} \phi_T(F) \\
&= (1 + o_{T \to \infty}(1)) \cdot (1 + O_T(o_{k \to \infty}(1))) \\
&\qquad \cdot \prod_p \int_{v \in V(\mathbb{Z}_p) : (I(v), J(v)) \in \mathrm{Inv}_p^{(T,\mathcal{B})}} dv \, \phi_*^{(p)}(v).
\end{aligned}$$

Similarly, we of course have:

$$\begin{aligned}
&n_T^{-2k} \cdot \#|\{(I,J) \in V(\mathbb{Z}/n_T^k) : I \equiv 0 \pmod{n_T}, J \equiv a \pmod{m}\}| \\
&= \prod_{p \leq T} \int_{(I,J) \in \mathrm{Inv}_p^{(T,\mathcal{B})}} dI dJ \\
&= (1 + o_{T \to \infty}(1)) \cdot \prod_p \int_{(I,J) \in \mathrm{Inv}_p^{(T,\mathcal{B})}} dI dJ.
\end{aligned}$$

[8] See the previous footnote (about factoring out $(1 + o_{k \to \infty}(1))$ and $(1 + o_{T \to \infty}(1))$).

Combining all these, we find that:

$$\begin{aligned}
&\operatorname*{Avg}_{B\in\mathcal{B}:|B|\leq X} \#|\mathrm{Sel}_2(E_{0,B}/\mathbb{Q})-\{0\}| \\
&\leq (1+O_T(o_{X\to\infty}(1)))\cdot(1+o_{T\to\infty}(1))\cdot(1+O_T(o_{k\to\infty}(1))) \\
&\qquad \cdot \frac{\int_{\coprod_i \mathcal{R}^{(i)}(X^2/4)} dv}{\int_{\coprod_i R^{(i)}(X^2/4)} dIdJ} \cdot \frac{\prod_p \int_{v\in V(\mathbb{Z}_p):(I(v),J(v))\in\mathrm{Inv}_p^{(T,\mathcal{B})}} dv\, \phi_*^{(p)}(v)}{\prod_p \int_{(I,J)\in\mathrm{Inv}_p^{(T,\mathcal{B})}} dIdJ}.
\end{aligned}$$

But the calculations in Section 3.6 of Bhargava-Shankar's [21] amount to the statement that

$$\frac{\int_{\coprod_i \mathcal{R}^{(i)}(X^2/4)} dv}{\int_{\coprod_i R^{(i)}(X^2/4)} dIdJ} \cdot \frac{\prod_p \int_{v\in V(\mathbb{Z}_p):(I(v),J(v))\in\mathrm{Inv}_p^{(T,\mathcal{B})}} dv\, \phi_*^{(p)}(v)}{\prod_p \int_{(I,J)\in\mathrm{Inv}_p^{(T,\mathcal{B})}} dIdJ} = 2.$$

Taking $k\to\infty$, then $X\to\infty$, and finally $T\to\infty$, we deduce Theorem 3.1.1. □

3.3.2 The uniformity estimate.

In order to prove the matching lower bound (recall that we are in the special case where $\mathcal{B}\subseteq\mathbb{Z}-\{0\}$ is defined by finitely many congruence conditions) we simply run the above argument with minorants instead — that is to say, we instead take $0 =: \phi_0^{(p)}(F)\leq\phi_1^{(p)}(F)\leq\cdots\leq\phi_n^{(p)}(F)\leq\cdots\leq\phi_*^{(p)}(F)$ for all $F\in V(\mathbb{Z}_p)^{\Delta\neq 0}$, with $\phi_n : V(\mathbb{Z}_p)^{\Delta\neq 0}\to[0,1]$ factoring through $V(\mathbb{Z}/p^n)$ (and not $V(\mathbb{Z}/p^{n-1})$) and such that $\phi_n^{(p)}(F)\to\phi_*^{(p)}(F)$ as $n\to\infty$. The argument is precisely the same, with the exception of the first step.

Specifically, when proving the upper bound we implicitly used that the binary quartics $F\in V(\mathbb{Z})^{\text{nontriv.}}$ with $I(F)=0$ representing 2-Selmer classes of $E_{0,J(F)}/\mathbb{Q}$ — i.e. such that $Z^2=F(X,Y)$ is nontrivially soluble in all completions of $\mathbb{Q}$ — are in particular locally soluble at those $p\leq T$. However of course the converse does not hold. So, just as in Bhargava-Shankar's [21], we need only prove that the number of binary quartics F in e.g. $B(u,t,\lambda,X) \cap Y(\mathbb{Z})$ which are *not* locally soluble at some prime p with $p > \Pi$ is

$$\ll \frac{\lambda^{12+o(1)}}{\Pi \log \Pi}$$

when e.g. $\Pi \leq \lambda^{10^{-10}}$ and we are in the "bulk", so that $t \ll \lambda^{o(1)}$.

Now, just as in the proof of Proposition 3.18 of Bhargava-Shankar's [21], if $F \in V(\mathbb{Z})^{\mathrm{nontriv.}}$ is a binary quartic that is not locally soluble at p, then $F \pmod{p}$ has splitting type one of $(1^2 1^2)$, (2^2), or (1^4). But if moreover $I(F) = 0$, and thus $I(F) \equiv 0 \pmod{p}$, one gets much more: the splitting types $(1^2 1^2)$ and (2^2) are not possible, as one can see by e.g. explicit calculation.[9] Thus $F \pmod{p}$ must be a fourth power of a linear form, which is to say that $F \pmod{p}$ lies on the codimension 3 subvariety $Z \subseteq V$ given by fourth powers of linear forms, namely the affine cone on a rational normal curve.

So it follows that we may bound the number of $F \in B(u,t,\lambda,X) \cap Y(\mathbb{Z})$ which are not locally soluble at some p with $p > \Pi$ by the number of $F \in B(u,t,\lambda,X) \cap Y(\mathbb{Z})$ for which the reduction $F \pmod{p} \in Z(\mathbb{F}_p)$ for some $p > \Pi$, and then the desired bound follows from invoking Theorem 1.1 of Browning-Heath-Brown's [33].

3.3.3 Point counting.

Let us now prove Lemmas 3.3.1 and 3.3.2. We note that we will give an essentially one-line proof of Lemma 3.3.1 (namely, "use the divisor bound to determine a, e

[9] Working over $\overline{\mathbb{F}}_p$ and changing variables suitably, this amounts to the assertion that

$$I(X^2 \cdot (X - n \cdot Y)^2) = I(X^4 - 2n \cdot X^3 Y + n^2 \cdot X^2 Y^2) = n^4.$$

from b, c, d"), which subsumes the entirety of Ruth's Section 3.5 (pages $29 - 37$ of [87]).

Proof of Lemma 3.3.1. Let $F \in \lambda \cdot n_u \cdot a_t \cdot G_0 \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}}$. Write $F(X, Y) =: a \cdot X^4 + b \cdot X^3 Y + c \cdot X^2 Y^2 + d \cdot XY^3 + e \cdot Y^4$. Evidently (by e.g. compactness of G_0 and L) we have that:

$$\begin{aligned} 0 \neq |a| &\ll \frac{\lambda^4}{t^4}, \\ |b| &\ll \frac{\lambda^4}{t^2}, \\ |c| &\ll \lambda^4, \\ |d| &\ll t^2 \cdot \lambda^4, \\ 0 \neq |e| &\ll t^4 \cdot \lambda^4. \end{aligned}$$

Therefore the number of tuples $(b, c, d) \in \mathbb{Z}^3$ among $F \in \lambda \cdot n_u \cdot a_t \cdot G_0 \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}}$ is $\ll \lambda^{12}$.

Moreover, by hypothesis $12ae - 3bd + c^2 = 0$, i.e. $0 \neq 12ae = 3bd - c^2 \ll \lambda^8$. Thus (b, c, d) determine (a, e) up to $\ll \lambda^{o(1)}$ choices. In other words, the map

$$\lambda \cdot n_u \cdot a_t \cdot G_0 \cdot L \cap V(\mathbb{Z})^{\text{nontriv.}} \to \mathbb{Z}^3$$

via

$$a \cdot X^4 + b \cdot X^3 Y + c \cdot X^2 Y^2 + d \cdot XY^3 + e \cdot Y^4 \mapsto (b, c, d)$$

has image of size $\ll \lambda^{12}$ and fibres of size $\ll \lambda^{o(1)}$. The lemma follows. □

As for Lemma 3.3.2, we first note that it is essentially identical to Ruth's Proposition $3.4.1$ (modulo the small inaccuracies in his treatment that we have already mentioned), which he states without proof (since the range $t \leq \lambda^{\delta}$ with $\delta \ll 1$ is easily dealt with using the smoothed delta symbol method).

For the reader's convenience we will give a full proof of Lemma 3.3.2 anyway.

Proof of Lemma 3.3.2. Before we begin we note once again that it is not necessary to use the smoothed delta symbol method, since we are asking about zeroes of a quadric in five variables, something easily handled by the classical form of the circle method.

We follow the notation of Heath-Brown's [54]. Let $w_0 \in C_c^\infty(\mathbb{R})$ via

$$w_0(x) := \begin{cases} \exp\left(-\frac{1}{(1-x^2)}\right) & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}.$$

Let

$$\omega(x) := \frac{4}{\int_{\mathbb{R}} w_0(t)dt} \cdot w_0(x).$$

Let

$$h(x,y) := \sum_{q\geq 1} \frac{\omega(qx) - \omega\left(\frac{|y|}{qx}\right)}{qx}.$$

Note that $h(x,y) = 0$ when $x \gg 1 + |y|$ and that $h(x,y) \ll x^{-1}$.

We will first detail the argument in the case of $M = 1$ (i.e. no congruence condition) and then comment on necessary modifications to more general M and $F_0 \in V(\mathbb{Z}/M)$ as above.

Applying Theorem 2 of Heath-Brown's [54] with his $n = 5$ and his $Q = \lambda^4$, we find that:

$$\begin{aligned}
&\sum_{F\in V(\mathbb{Z})} \varphi(a_t^{-1} \cdot n_u^{-1} \cdot (\lambda \cdot \mathrm{id})^{-1} \cdot F) \\
&= (\lambda^{-8} + O_N(\lambda^{-N})) \cdot \sum_{q\geq 1} q^{-5} \sum_{\vec{c}\in V(\mathbb{Z})^*} \left(\sum_{u\in(\mathbb{Z}/q)^\times} \sum_{G\in V(\mathbb{Z}/q)} e_q(u \cdot I(G) + G \cdot \vec{c}) \right) \\
&\qquad \cdot \int_{F\in V(\mathbb{R})} dF\, \varphi(a_t^{-1} \cdot n_u^{-1} \cdot (\lambda \cdot \mathrm{id})^{-1} \cdot F) \cdot h\left(\frac{q}{\lambda^4}, \frac{I(F)}{\lambda^8}\right) \cdot e_q(-F \cdot \vec{c}),
\end{aligned}$$

where we have written $e_q(z) := e\left(\frac{z}{q}\right) := e^{\frac{2\pi i z}{q}}$.

For us the error term will consist of those terms where $\vec{c} \neq \vec{0}$, and the terms with $\vec{c} = \vec{0}$ will comprise the main term (in the end we will simply cite Heath-Brown's [54] for the analysis of the main term, which in any case is considerably simpler).

Via the change of variable $F \mapsto n_u \cdot a_t \cdot (\lambda \cdot \mathrm{id}) \cdot F$ (note that $(\lambda \cdot \mathrm{id}) \cdot F = \lambda^4 \cdot F$ since F is homogeneous of degree 4 — here $(\lambda \cdot \mathrm{id}) \cdot F$ on the left-hand side indicates the action of $\lambda \cdot \mathrm{id} \in G$ on $F \in V$ via $G \curvearrowright V$, and the $\cdot$ on the right-hand side denotes multiplication), we find that:

$$\begin{aligned}&\int_{F \in V(\mathbb{R})} dF\, \varphi(a_t^{-1} \cdot n_u^{-1} \cdot (\lambda \cdot \mathrm{id})^{-1} \cdot F) \cdot h\left(\frac{q}{\lambda^4}, \frac{I(F)}{\lambda^8}\right) \cdot e_q(-F \cdot \vec{c}) \\ &= \lambda^{20} \cdot \int_{F \in V(\mathbb{R})} dF\, \varphi(F) \cdot h\left(\frac{q}{\lambda^4}, I(F)\right) \cdot e_q(-\lambda^4 \cdot F \cdot ((n_u \cdot a_t)^\dagger \cdot \vec{c})).\end{aligned}$$

Therefore we see that the error term is:

$$\begin{aligned}(\lambda^{12} + O_N(\lambda^{-N})) \cdot \sum_{q \geq 1} q^{-5} \sum_{\vec{0} \neq \vec{c} \in V(\mathbb{Z})^*} \left(\sum_{u \in (\mathbb{Z}/q)^\times} \sum_{G \in V(\mathbb{Z}/q)} e_q(u \cdot I(G) + G \cdot \vec{c}) \right) \\ \cdot \int_{F \in V(\mathbb{R})} dF\, \varphi(F) \cdot h\left(\frac{q}{\lambda^4}, I(F)\right) \cdot e_q(-\lambda^4 \cdot F \cdot ((n_u \cdot a_t)^\dagger \cdot \vec{c})).\end{aligned}$$

Note that the $\varphi(F)$ term in the integral forces $||F||_\infty \ll_\varphi 1$ if the integrand is to be nonzero, and then our observation that $h(x, y) = 0$ if $x \gg 1 + |y|$ forces $q \ll_\varphi \lambda^4$ as well.

Note also that if $q \ll \lambda^{4-\varepsilon} \cdot ||(n_u \cdot a_t)^\dagger \cdot \vec{c}||_\infty$ — i.e. if

$$||(n_u \cdot a_t)^\dagger \cdot \vec{c}||_\infty \gg \frac{q}{\lambda^{4-\varepsilon}}$$

— we find, by repeated integration by parts, that such terms contribute $O_{\varepsilon,\varphi,N}(t^N \cdot \lambda^{O(1)-N})$.

Now the complete exponential sum, which is just

$$\sum_{u\in(\mathbb{Z}/q)^\times}\sum_{G_0,\ldots,G_4\in\mathbb{Z}/q} e_q(u\cdot(12G_0G_4-3G_1G_3+G_2^2)+(c_0G_0+c_1G_1+c_2G_2+c_3G_3+c_4G_4)),$$

is very easy to calculate (for a general quadric the same calculation works, but let us just focus on our particular case). Let us calculate it in detail anyway. Because the exponential sum is complete, we first reduce to treating the case $q=p^e$ a prime power. For simplicity of notation, let us deal only with the case $p\neq 2,3$. Write

$$I^*(\vec{G}) := -(12)^{-1}\cdot G_0G_4 + 3^{-1}\cdot G_1G_3 - 4^{-1}\cdot G_2^2.$$

We first average over changes of variable $G\mapsto v\cdot G$ (and $u\mapsto v^{-2}\cdot u$) with $v\in(\mathbb{Z}/q)^\times$, getting:

$$\sum_{u\in(\mathbb{Z}/q)^\times}\sum_{G_0,\ldots,G_4\in\mathbb{Z}/q} e_q(u\cdot(12G_0G_4-3G_1G_3+G_2^2)+(c_0G_0+c_1G_1+c_2G_2+c_3G_3+c_4G_4))$$
$$=\frac{1}{\varphi(q)}\cdot\sum_{v\in(\mathbb{Z}/q)^\times}\sum_{u\in(\mathbb{Z}/q)^\times}\sum_{\vec{G}\in(\mathbb{Z}/q)^5} e_q(u\cdot I(\vec{G})+(v\cdot\vec{c})\cdot\vec{G}).$$

Completing the square (and scaling to remove the coefficients 12 and -3) in the evident way produces:

$$\sum_{G_0,\ldots,G_4\in\mathbb{Z}/q} e_q(u\cdot(12G_0G_4-3G_1G_3+G_2^2)+v\cdot(c_0G_0+c_1G_1+c_2G_2+c_3G_3+c_4G_4))$$
$$=e_q\left((-(12)^{-1}\cdot c_0c_4+3^{-1}\cdot c_1c_3-4^{-1}\cdot c_2^2)\cdot v^2\right)\cdot\sum_{G_0,\ldots,G_4\in\mathbb{Z}/q} e_q(u\cdot(G_0G_4+G_1G_3+G_2^2)).$$

Thus

$$\frac{1}{\varphi(q)}\cdot\sum_{v\in(\mathbb{Z}/q)^\times}\sum_{u\in(\mathbb{Z}/q)^\times}\sum_{\vec{G}\in(\mathbb{Z}/q)^5} e_q(u\cdot I(\vec{G})+(v\cdot\vec{c})\cdot\vec{G})$$

$$= \frac{1}{\varphi(q)} \cdot \left(\sum_{v \in (\mathbb{Z}/q)^\times} e_q\left(I^*(\vec{c}) \cdot v^2\right) \right) \cdot \left(\sum_{G_0, \ldots, G_4 \in \mathbb{Z}/q} e_q(u \cdot (G_0 G_4 + G_1 G_3 + G_2^2)) \right).$$

Writing $I^*(\vec{c}) =: p^{v_p(I^*(\vec{c}))} \cdot \widetilde{I^*(\vec{c})}$, the first term is:

$$\sum_{v \in (\mathbb{Z}/q)^\times} e_q\left(I^*(\vec{c}) \cdot v^2\right)$$
$$= \begin{cases} 0 & v_p(I^*(\vec{c})) < e-1, \\ \varphi(q) = p^{e-1}(p-1) & v_p(I^*(\vec{c})) \geq e, \\ \left(\left(\frac{\widetilde{I^*(\vec{c})}}{p} \right) \cdot \sqrt{p} - 1 \right) \cdot p^{e-1} & v_p(I^*(\vec{c})) = e-1, p \equiv 1 \pmod 4, \\ \left(\left(\frac{\widetilde{I^*(\vec{c})}}{p} \right) \cdot i\sqrt{p} - 1 \right) \cdot p^{e-1} & v_p(I^*(\vec{c})) = e-1, p \equiv 3 \pmod 4, \end{cases}$$

by Hensel's lemma.

As for the second term, summing over G_3 and G_4 gives 0 unless $G_0 = G_1 = 0 \in \mathbb{Z}/q$, so that:

$$\sum_{u \in (\mathbb{Z}/q)^\times} \sum_{G_0, \ldots, G_4 \in \mathbb{Z}/q} e_q(u \cdot (G_0 G_4 + G_1 G_3 + G_2^2)) = q^2 \cdot \sum_{u \in (\mathbb{Z}/q)^\times} \sum_{G_2 \in \mathbb{Z}/q} e_q(u \cdot G_2^2).$$

Now we switch sums and sum instead over $u \in (\mathbb{Z}/q)^\times$, obtaining a Ramanujan sum: 0 if $v_p(G_2) < \frac{e-1}{2}$, $-p^{e-1}$ if $v_p(G_2) = \frac{e-1}{2}$, and $\varphi(q) = (p-1) \cdot p^{e-1}$ if $v_p(G_2) \geq \frac{e}{2}$. Thus the full sum is:

$$\sum_{u \in (\mathbb{Z}/q)^\times} \sum_{G_0, \ldots, G_4 \in \mathbb{Z}/q} e_q(u \cdot (G_0 G_4 + G_1 G_3 + G_2^2))$$
$$= q^2 \cdot \left(-p^{e-1} \cdot \# \left| \left\{ x \in \mathbb{Z}/p^e : v_p(x) = \frac{e-1}{2} \right\} \right| + (p-1) \cdot p^{e-1} \cdot \# \left| \left\{ x \in \mathbb{Z}/p^e : v_p(x) \geq \frac{e}{2} \right\} \right| \right)$$
$$= p^{2e} \cdot \begin{cases} (p-1) \cdot p^{\frac{3e}{2} - 1} & e \equiv 0 \pmod 2, \\ 0 & e \equiv 1 \pmod 2. \end{cases}$$

In any case it follows that (the cases $p = 2$ or 3 being slightly more notationally cumbersome), for general $q \in \mathbb{Z}^+$:

$$\sum_{u \in (\mathbb{Z}/q)^\times} \sum_{\vec{G} \in V(\mathbb{Z}/q)} e_q(u \cdot I(G) + \vec{G} \cdot \vec{c}) \ll \begin{cases} 0 & \exists p > 3 : v_p(q) = 1 \\ q^{\frac{7}{2}+o(1)} & \forall p | q : p > 3, v_p(q) \geq 2, \end{cases}$$

and indeed one can sharpen the bound significantly — e.g. the sum is zero unless $p|q$ and $p > 3$ implies $v_p(q) \geq 2$ and $v_p(I^*(\vec{c})) \geq v_p(q) - 1$, but we will not use this.

Now we return to the smoothed delta method. Again, the integral is nonnegligible only for

$$||(n_u \cdot a_t)^\dagger \cdot \vec{c}||_\infty \ll \frac{q}{\lambda^{4-\varepsilon}}$$

(in which case it is $\ll_\varphi \frac{\lambda^4}{q}$). Thus either $\vec{c} = \vec{0}$, in which case the corresponding summand contributes to the main term dealt with by Ruth (and by Heath-Brown in [54]), or else $\vec{c} \neq \vec{0}$, in which case we must have that $q \gg \frac{\lambda^{4-\varepsilon}}{t^4}$ since by inspection $||(n_u \cdot a_t)^\dagger \cdot \vec{c}||_\infty \gg t^{-4} \cdot ||\vec{c}||_\infty \gg t^{-4}$.

The error term is therefore:

$$\ll \lambda^{12+o(1)} \cdot \sum_{\frac{\lambda^{4-\varepsilon}}{t^4} \ll q \ll \lambda^4,\, q \text{ powerful}} q^{-\frac{3}{2}} \sum_{\vec{0} \neq \vec{c} \in V(\mathbb{Z})^* : ||\vec{c}||_\infty \ll \frac{q}{\lambda^{4-\varepsilon}}} \int_{F \in V(\mathbb{R})} dF\, \varphi(F) \cdot h\left(\frac{q}{\lambda^4}, I(F)\right) \cdot e_q(-\lambda^4 \cdot F \cdot ((n_u \cdot a_t)^\dagger \cdot \vec{c})).$$

Bounding the integrals trivially (i.e. by $\ll_\varphi \frac{q}{\lambda^4}$) and noting that the number of $0 \neq \vec{c} \in \mathbb{Z}^5$ such that $||(n_u \cdot a_t)^\dagger \cdot \vec{c}||_\infty \ll \frac{q}{\lambda^{4-\varepsilon}}$ is

$$\ll \frac{t^4 \cdot q}{\lambda^{4-\varepsilon}} \cdot \left(1 + \frac{t^2 \cdot q}{\lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{q}{\lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{q}{t^2 \cdot \lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{q}{t^4 \cdot \lambda^{4-\varepsilon}}\right)$$

when $\frac{\lambda^{4-\varepsilon}}{t^4} \ll q \ll \lambda^4$, we get that the error term is:

$$\ll t^4 \cdot \lambda^{4+\varepsilon+o(1)} \cdot \sum_{\frac{\lambda^{4-\varepsilon}}{t^4} \ll q \ll \lambda^4,\, q \text{ powerful}} q^{\frac{1}{2}} \cdot \left(1 + \frac{t^2 \cdot q}{\lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{q}{\lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{t^{-2} \cdot q}{\lambda^{4-\varepsilon}}\right) \cdot \left(1 + \frac{t^{-4} \cdot q}{\lambda^{4-\varepsilon}}\right)$$

$$\ll t^6 \cdot \lambda^{8+5\varepsilon},$$

as desired.

Thus we have bounded the error term suitably. The required analysis of the main term is already done in Heath-Brown's [54] (see e.g. the bottom of page 51, i.e. the end of the proof of his Theorems 4 and 5), at least in the case $M = 1$.

We now discuss the modifications necessary for general $M \in \mathbb{Z}^+$ and $F_0 \in V(\mathbb{Z}/M)$. First, in the application of the smoothed delta symbol method, instead of summing over $F \in V(\mathbb{Z})$, we sum instead over the $F \in V(\mathbb{Z})$ for which $F \equiv F_0 \pmod{M}$ by summing over $\tilde{F} \in V(\mathbb{Z})$ and writing $F := F_0 + M \cdot \tilde{F}$ (we implicitly choose a representative of F_0 in $V(\mathbb{Z})$ and abuse notation by writing it as $F_0 \in V(\mathbb{Z})$). We then change variables from $\tilde{F}$ back to F in the integral and incur a factor of M^{-5}. The rest of the analysis of the error term is precisely the same (except that the primes one has to treat separately in the complete exponential sum calculation are now those $p|6M$, and the error terms now depend on M).

It remains to treat the main term, i.e. the local densities. We note that, by definition, we find local densities

$$\sigma_p^{(F_0 \pmod{M})}(Y(\mathbb{Z})) := \lim_{k\to\infty} p^{-4k} \cdot \#|\{\tilde{F} \in V(\mathbb{Z}/p^k) : I(F_0 + M \cdot \tilde{F}) \equiv 0 \pmod{p^k}\}|.$$

We therefore are reduced to the claim that

$$M^{-5} \cdot \prod_p \sigma_p^{(F_0 \pmod{M})}(Y(\mathbb{Z})) = \prod_p \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})).$$

Of course for $p \nmid M$ we have that

$$\sigma_p^{(F_0 \pmod{M})}(Y(\mathbb{Z})) = \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) = \sigma_p(Y(\mathbb{Z})),$$

via the evident change of variables $\tilde{F} \mapsto M^{-1} \cdot (\tilde{F} - F_0)$.

However, it is also evident that

$$M^{-5} \cdot \prod_{p \mid M} \sigma_p^{(F_0 \pmod{M})}(Y(\mathbb{Z})) = \prod_{p \mid M} \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})),$$

for the following reason. For $k \in \mathbb{Z}^+$ with $k \gg 1$ we have that:

$$\begin{aligned}
&M^{-5} \cdot \prod_{p \mid M} \sigma_p(Y_{F_0 \pmod{M}}(\mathbb{Z})) \\
&= M^{-5} \cdot \prod_{p \mid M} \left(\begin{array}{c} p^{-4k \cdot v_p(M)} \cdot \#|\{\tilde{F} \in V(\mathbb{Z}/p^{k \cdot v_p(M)}) : I(F_0 + M \cdot \tilde{F}) \equiv 0 \pmod{p^{k \cdot v_p(M)}}\}| \\ +O(p^{-k \cdot v_p(M)}) \end{array} \right) \\
&= \left(1 + O(e^{-\Omega_M(k)})\right) \cdot M^{-4k-5} \cdot \#|\{\tilde{F} \in V(\mathbb{Z}/M^k) : I(F_0 + M \cdot \tilde{F}) \equiv 0 \pmod{M^k}\}|,
\end{aligned}$$

by the Chinese remainder theorem and that fact that all $p \geq 2$.

Because the condition $I(F_0 + M \cdot \tilde{F}) \equiv 0 \pmod{M^k}$ only depends on $\tilde{F} \pmod{M^{k-1}}$, we find that:

$$\begin{aligned}
&M^{-4k-5} \cdot \#|\{\tilde{F} \in V(\mathbb{Z}/M^k) : I(F_0 + M \cdot \tilde{F}) \equiv 0 \pmod{M^k}\}| \\
&= M^{-4k} \cdot \#|\{F \in V(\mathbb{Z}/M^k) : I(F) \equiv 0 \pmod{M^k}, F \equiv F_0 \pmod{M}\}| \\
&= \prod_{p \mid M} \left(\begin{array}{c} p^{-4k \cdot v_p(M)} \cdot \#|\{F \in V(\mathbb{Z}/p^{k \cdot v_p(M)}) : I(F) \equiv 0 \pmod{p^{k \cdot v_p(M)}}, F \equiv F_0 \pmod{M}\}| \\ +O(p^{-k \cdot v_p(M)}). \end{array} \right)
\end{aligned}$$

Thus taking $k \to \infty$ we find that

$$M^{-5} \cdot \prod_{p|M} \sigma_p^{(F_0 \text{ (mod } M))}(Y(\mathbb{Z})) = \prod_{p|M} \sigma_p(Y_{F_0 \text{ (mod } M)}(\mathbb{Z})),$$

as desired. □

3.4 Proof of Theorem 3.1.2.

Now to the proof of Theorem 3.1.2.

3.4.1 Reduction to point counting.

We run the same argument as above, except our notation follows Bhargava-Ho's [18] rather than Ruth's [87], and we appeal in the end to part (f) of Theorem 1.1 of Bhargava-Ho's [18] (rather than Theorem 3.1 of Bhargava-Shankar's [21]) to calculate the product of local densities. Because otherwise the argument is essentially the same as in the previous section (in fact it is easier, since in the circle method argument we deal with a quadric in eight variables instead of five) we will be a significantly more terse in this section.

Again, we follow the notation in Bhargava-Ho's [18] (the relevant parametrization is by triply symmetric hypercubes). Let $V := 2 \otimes \mathrm{Sym}_3(2)$, the space of pairs of "threes-in" binary cubic forms. Let G be the image in $\mathrm{GL}(V)$ of $\{(g, h) \in \mathrm{GL}_2 \times \mathrm{GL}_2 : \det g \cdot (\det h)^3 = 1\}$, acting in the evident way on $2 \otimes \mathrm{Sym}_3(2)$ (the first GL_2 on the first factor via the standard representation, and the second GL_2 on the second via the induced action on Sym_3 of the standard representation). Note that $G \cong (\mathrm{SL}_2 \times \mathrm{SL}_2)/\mu_2$.

We write, for $v \in V$,

$$H(v) := \max\left(|I_2(v)|^{\frac{1}{2}}, |I_6(v)|^{\frac{1}{6}}\right)^{24},$$

where I_2 and I_6 are the invariants a_2 and a_6 of Section 6.3.2 of Bhargava-Ho's [17] and a_1 and a_3 of line 6 (corresponding to the family $F_1(3)$) of Table 1 in Bhargava-Ho's [18].

Let R be a fundamental domain for $G(\mathbb{R}) \curvearrowright V(\mathbb{R})^{\Delta\neq 0}$ (note that Bhargava-Ho write $V(\mathbb{R})^{\mathrm{stab}} := V(\mathbb{R})^{\Delta\neq 0}$), as constructed in Section 5 of Bhargava-Ho's [18] (via, in their notation, $R := \coprod_i R^{(i)}$). Let $L := \{v(\vec{a}) : \vec{a} \in (\mathbb{R}^m)^{\Delta\neq 0}_{H=1}\}$ (here in their notation $m = 2$ and $v(\vec{a})$ is as defined in Section 4 of Bhargava-Ho's [18]) and $\Lambda := \{(\lambda \cdot \mathrm{id}, \mathrm{id}) \in \mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R}) : \lambda \in \mathbb{R}^+\} \subseteq \mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R})$. Note that $R = \Lambda \cdot L$. Let $R(X) := \{v \in R : H(v) \leq X\}$. Let $\vec{F}_\pm := v((0, \pm 1)) \in L$ be the two points in L with $I_2 = 0$.

Note that, by construction, since $H((\lambda, \mathrm{id}) \cdot v) = \lambda^{24} \cdot H(v)$, the coefficients of a $v \in \lambda \cdot L \subseteq R(X)$ are all $\ll \lambda \ll X^{\frac{1}{24}}$, and hence, for $G_0 \subseteq G(\mathbb{R})$ compact, the coefficients of a $v \in \lambda \cdot G_0 \cdot L \subseteq R(X)$ are all $\ll_{G_0} \lambda \ll X^{\frac{1}{24}}$.

Let $\mathcal{F}$ be a fundamental domain for $G(\mathbb{Z}) \curvearrowright G(\mathbb{R})$, as constructed in Section 5.2 of Bhargava-Ho's [18]. Note that $\mathcal{F}$ lies inside the following Siegel set:

$$\mathcal{F} \subseteq N \cdot A \cdot K,$$

where

$$N := \left\{(n_{u_1}, n_{u_2}) \in G(\mathbb{R}) : |u_i| \leq \frac{1}{2}\right\},$$
$$A := \left\{(a_{t_1}, a_{t_2}) \in G(\mathbb{R}) : t_i \geq \sqrt{\frac{\sqrt{3}}{2}}\right\},$$
$$K := \mathrm{SO}_2(\mathbb{R}) \times \mathrm{SO}_2(\mathbb{R}) \subseteq G(\mathbb{R}),$$

with notation as before: $n_u := \left(\begin{smallmatrix} 1 & 0 \\ u & 1 \end{smallmatrix}\right)$ and $a_t := \left(\begin{smallmatrix} t^{-1} & 0 \\ 0 & t \end{smallmatrix}\right)$.

As before, let $\alpha \in C_c^\infty(G(\mathbb{R}))$ and $\beta \in C_c^\infty(L)$ be compactly supported smooth functions such that: α is K-invariant, $\int_{G(\mathbb{R})} \alpha = 1$, $\beta(\vec{F}_\pm) = 1$, and $\operatorname{supp} \beta, \beta^{-1}(\{1\}) \subseteq L^{(1)}$ are both unions of two small compact intervals respectively containing $\vec{F}_\pm$. Let

$$\varphi(v) := \sum_{g \cdot v_L = v} \alpha(g) \cdot \beta(v_L).$$

Let $V(\mathbb{Z})^{\text{nontriv.}} := \{(F_1, F_2) \in V(\mathbb{Z}) : 0 \notin F_i(\mathbb{P}^1(\mathbb{Q}))\}$. Let $Y_{\mathcal{B}}(\mathbb{Z}) := \{v \in V(\mathbb{Z}) : I_2(v) = 0, I_6(v) \in \mathcal{B}\}$ and $Y_{\mathcal{B}}(\mathbb{Z})^{\text{nontriv.}} := Y_{\mathcal{B}}(\mathbb{Z}) \cap V(\mathbb{Z})^{\text{nontriv.}}$.

Write $n_{(u_1,u_2)} := (n_{u_1}, n_{u_2})$, and similarly $a_{(t_1,t_2)} := (a_{t_1}, a_{t_2})$.

Let, for $S \subseteq V(\mathbb{Z})$,

$$\#_\varphi |B(\vec{u}, \vec{t}, \lambda, X) \cap S| := \sum_{\vec{F} \in S : ||I_6(\vec{F})||_\infty \leq X} \varphi(a_{\vec{t}}^{-1} \cdot n_{\vec{u}}^{-1} \cdot (\lambda \cdot \text{id}, \text{id}) \cdot (F_1, F_2)).$$

Let also $Y(\mathbb{Z}) := Y_{\mathbb{Z}-\{0\}}(\mathbb{Z})$ and $Y(\mathbb{Z})^{\text{nontriv.}} := Y_{\mathbb{Z}-\{0\}}(\mathbb{Z})^{\text{nontriv.}}$.

We again see that it suffices to prove the following two lemmas. The proof that these lemmas suffice, including reducing the evaluation of the resulting product of local densities by using the same trick to reduce to the same evaluation done (for the larger family $F_1(3)$) in the proof of part (f) of Theorem 1.1 of Bhargava-Ho's [18], is entirely the same as in the previous section, so we omit it.

Lemma 3.4.1. *Let $\lambda \in \mathbb{R}^+, u_i \in \left[-\frac{1}{2}, \frac{1}{2}\right], \sqrt{\frac{\sqrt{3}}{2}} \leq t_i \ll \lambda$. Then:*

$$\#_\varphi |B(\vec{u}, \vec{t}, \lambda, X) \cap Y(\mathbb{Z})^{\text{nontriv.}}| \ll_\varphi \lambda^{6+o(1)}.$$

Lemma 3.4.2. *Let $\emptyset \neq \mathcal{B} \subseteq \mathbb{Z} - \{0\}$ be a set defined by finitely many congruence conditions. Let*

$$\lambda \in \mathbb{R}^+, u_i \in \left[-\frac{1}{2}, \frac{1}{2}\right], \sqrt{\frac{\sqrt{3}}{2}} \leq t_i \ll \lambda.$$

Then:

$$\#_\varphi|B(\vec{u},\vec{t},\lambda,X)\cap Y_{\mathcal{B}}(\mathbb{Z})| = \sigma_\infty(\vec{u},\vec{t},\lambda,X)\cdot\prod_p \sigma_p(\mathcal{B})+O_\varphi(||\vec{t}||_\infty^8\cdot\lambda^{4+o(1)})+O_{\varphi,N}(||\vec{t}||_\infty^N\cdot\lambda^{O(1)-N}),$$

where

$$\sigma_\infty(\vec{u},\vec{t},\lambda,X) := \lim_{\varepsilon\to 0}\frac{\int_{v\in V(\mathbb{R}):|I_2(v)|\leq\varepsilon,|I_6(v)|\leq X} dv\,\varphi(a_{\vec{t}}^{-1}\cdot n_{\vec{u}}^{-1}\cdot(\lambda\cdot\mathrm{id},\mathrm{id})^{-1}\cdot v)}{2\varepsilon}$$

and

$$\sigma_p(\mathcal{B}) := \lim_{n\to\infty} p^{-4n}\cdot\#|\{\vec{F}\in V(\mathbb{Z}/p^n) : I_2(\vec{F})\equiv 0 \pmod{p^n}, I_6(\vec{F})\equiv a \pmod{m}\}.$$

3.4.2 The uniformity estimate.

In fact the proof of the uniformity estimate is identical to the one proven in Section 3.3.2, for the following reason. Recall that, given a pair of binary cubic forms $\vec{F} =: (F_1, F_2)$ with each $F_i \in \mathbb{Z}[X,Y]$, one produces a binary quartic form via

$$G_{\vec{F}}(x,y) := \mathrm{disc}_{X,Y}(x\cdot F_1(X,Y) + y\cdot F_2(X,Y)) \in \mathbb{Z}[x,y].$$

Note that, as one can see by e.g. explicit calculation, $I_2(\vec{F})|I(G_{\vec{F}})$.

Now in fact one has by definition that the pair $\vec{F} = (F_1, F_2)$ is locally soluble at p if and only if $G_{\vec{F}}$ is locally soluble at p. Therefore to bound the number of $\vec{F}\in B(\vec{u},\vec{t},\lambda,X)\cap Y(\mathbb{Z})$ which are not locally soluble at a p with $p > \Pi$, it suffices to observe that the statement that $\vec{F}$ is not locally soluble at p implies that $G_{\vec{F}} \pmod{p}$

lies on a codimension 3 subvariety of the space of binary quartics, so that (after checking the independence of the resulting three equations in the coefficients of $\vec{F}$, which in fact imply that either F_2 is proportional to F_1 or else $F_1 = 0$) $\vec{F} \pmod{p}$ lies on a codimension 3 subvariety of the space of pairs of binary cubics, in which case we may again apply Theorem 1.1 of Browning-Heath-Brown's [33] to conclude.

3.4.3 Point counting.

The proof of Lemma 3.4.1 is very much the same as the proof of Lemma 3.3.1. We give it here anyway.

Proof of Lemma 3.4.1. As mentioned, each coefficient of a $(F^{(1)}, F^{(2)}) =: \vec{F} \in B(\vec{u}, \vec{t}, \lambda, X) \cap Y(\mathbb{Z})^{\mathrm{nontriv.}}$ is $\ll_\varphi \lambda$. Applying the divisor bound, we determine $(F_0^{(1)}, F_3^{(2)})$, up to $\ll_\varphi \lambda^{o(1)}$ choices, from $(F_1^{(1)}, F_2^{(1)}, F_3^{(1)}, F_0^{(2)}, F_1^{(2)}, F_2^{(2)})$, and there are $\ll_\varphi \lambda^6$ choices for the latter. □

The proof of Lemma 3.4.2 is also much the same as the proof of Lemma 3.3.2. We comment on the only two differences.

Proof of Lemma 3.4.2. The only differences are the following. In applying the smoothed delta symbol method (i.e. Theorem 2 of Heath-Brown's [54]), we take $n = 8$ and $Q = \lambda$. In calculating the complete exponential sum, we note that, because we have an even number of variables and thus the mod-q complete exponential sums no longer vanish for q prime, we instead use the bound $\ll q^{5+o(1)}$ for all q. For the same reason we no longer reduce to a sum over only powerful q, but rather we sum over all $q \ll \lambda$ in the bounding of the error term.

Otherwise the proof is mutatis mutandis the same. □

Chapter 4

$\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S})$: S-integral points on hyperelliptic curves.

Abstract.

Let K be a number field, S a finite set of places of K, and $f \in \mathfrak{o}_K[t]$ of degree $d \geq 3$ a polynomial with discriminant $\Delta_f \neq 0$. Write $\mathcal{C}_f^{\text{aff.}} : y^2 = f(x)$ for the affine Weierstrass model of the hyperelliptic curve C_f/K associated to f. Write $J_f := \operatorname{Jac} C_f$. Let $g \in \mathfrak{o}_K[t]$ be a divisor of f of degree at least 3. Let $K_g := K[t]/(g(t))$. In this chapter we prove:

$$\#|\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S})| \ll \min\left(2^{\operatorname{rank} J_f(K)}, \#|\mathrm{Cl}(\mathfrak{o}_{K_g,S})[2]|\right)\cdot O(1)^{d^3\cdot\#|S|}\cdot|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|^{o(d^3\cdot[K:\mathbb{Q}])}.$$

This improves the best-known bound, due to Evertse-Silverman [46] (we note also work of Bombieri-Gubler [29]), from 1986. The proof is an exercise in optimizing a technique introduced by the famous mathematician X in 1926 [108].

4.1 Introduction.

4.1.1 Main theorem.

In this chapter we prove the following theorem. The proof is again quite straightforward: the technique can be summarized as applying a 2-descent and following one's (or Siegel's) nose.

Theorem 4.1.1. *Let K be a number field. Let S be a finite set of places of K. Let $f \in \mathfrak{o}_K[t]$ of degree $d \geq 3$ with discriminant $\Delta_f \neq 0$. Let $\mathcal{C}_f^{\text{aff.}} : y^2 = f(x)$ be the Weierstrass model of the hyperelliptic curve C_f/K corresponding to f. Let g be a divisor of f over K of degree at least 3. Let $K_g := K[t]/(g(t))$. Then:*

$$\#|\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S})| \ll \min\left(2^{\operatorname{rank} J_f(K)}, \#|\mathrm{Cl}(\mathfrak{o}_{K_g,S})[2]|\right) \cdot O(1)^{d^3 \cdot \#|S|} \cdot O(1)^{d^3 \cdot [K:\mathbb{Q}] \cdot \frac{\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}.$$

(Here, to be clear, by $\mathfrak{o}_{K_g,S}$ we mean the localization $S^{-1}\mathfrak{o}_{K_g}$ as an $\mathfrak{o}_K$-module, where $\mathfrak{o}_{K_g}$ is the product of rings of integers of the various number fields whose product is the semisimple K-algebra K_g.)

Remark 4.1.2. *We note that the bound improves when one applies the same technique to a superelliptic curve $y^k = f(x)$ with $k > 2$. The point is that one immediately gets to a degree k Thue equation upon performing the 2-descent.*

It is worth noting that standard conjectures imply that

$$\#|\mathrm{Cl}(\mathfrak{o}_{K_g,S})[2]| \ll |\Delta_{K_g}|^{o(1)} \ll |\Delta_K|^{o(d^3 \cdot [K:\mathbb{Q}])} \cdot |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|^{o(d^3 \cdot [K:\mathbb{Q}])}.$$

4.1.2 A corollary.

An obvious corollary (apply Bhargava-Gross [16] or Shankar-Wang [92]) is the following. We write, for $f \in \mathfrak{o}_K[t]$, $H(f) := \max_i |a_i|^{\frac{1}{i}}$, where $f(t) =: \sum_i a_{\deg f - i} \cdot t^i$.

Corollary 4.1.3. *Let K be a number field. Let S be a finite set of places of K. Let $d \geq 3$. Then:*

$$\operatorname*{Avg}_{f:H(f)\leq X, \Delta_f \neq 0} \#|\mathcal{C}_f^{\mathrm{aff.}}(\mathfrak{o}_{K,S})| \ll_{K,\#|S|,d,\varepsilon} X^{\varepsilon}.$$

We expect that one can do better with not much more effort (at least for odd d), but that is beyond the scope of this thesis. Of course the bound is known[1] to be much better when $d = 3$.

4.1.3 Class groups versus Selmer groups.

We also include in Section 4.3 an unrelated lemma inspired by Schaefer's Berkeley PhD thesis [88, 89], which implies in our situation that

$$\#|\mathrm{Sel}_2(J_f/K)| = \#|\mathrm{Cl}(\mathfrak{o}_{K_f})[2]| \cdot \Theta_d(1)^{[K:\mathbb{Q}]+\#|\{\mathfrak{p}|\Delta_f\}|}$$

— the point is that at all but the obvious primes the local conditions defining the Selmer group and the class group in $H^1(K, J_f[2])$ are the same.

4.1.4 Main technique.

The method of proof is to simply perform a 2-descent and find ourselves faced with a Thue-Mahler equation, to which we apply Evertse's bound. The descent procedure is exactly the same as Baker [10] uses in his effective upper bound on the heights of integral points on hyperelliptic curves, and is original to Siegel, under the pseudonym X [108]. Evertse and Silverman [46] derive very similar bounds using much the same method (namely, "just do a 2-descent"), but our bound is stronger than theirs — crucially, we only have a factor of $\#|\mathrm{Cl}(\mathfrak{o}_{K_g,S})[2]|$, rather than $\#|\mathrm{Cl}(\mathfrak{o}_F)[2]|^2$ for F/K a field containing three roots of f, which could a priori

[1] This problem was first treated in my senior thesis [4], though see Corollary 2.1.2 of Chapter 2 as well.

be much larger — because we are more efficient. (An improvement of Evertse-Silverman's result by Bombieri-Gubler (see Theorem $5.3.5$ of Bombieri-Gubler's [29]) suffers from the same weakness.) We note that intuitively one expects a factor involving only the 2-torsion of the class group of K_f when doing a 2-descent since one passes to $K_f^\times/(K_f^\times)^2$ in the course of the descent. The key reason we are able to improve on Evertse-Silverman and Bombieri-Gubler can essentially be summarized by the observation that one can use a Galois action to produce more equations from one. It is also crucial in the argument that we represent ideal classes by primes so that, in the eventual application of a bound due to Evertse on solutions to a Thue-Mahler equation, the auxiliary set of primes is not too large (we have already seen this trick in Chapter 2).

4.2 Proof of Theorem 4.1.1.

Proof of Theorem 4.1.1. Factorize

$$g =: \prod_i g_i$$

into K-irreducible factors $g_i \in \mathfrak{o}_K[t]$ with

$$1 \leq \deg g_1 \leq \deg g_2 \leq \cdots .$$

Note that $(g_i, g_j) = (1)$ for each $i \neq j$ because $\Delta_f \neq 0$. Thus $K_g \simeq \oplus_i K_{g_i}$, with $K_{g_i} := K[t]/(g_i(t))$.

We simply do a classical 2-descent on the curve to reduce to a Thue equation over a small extension of K_g, to which we apply bounds on the number of solutions due to Evertse. This reduction to a Thue equation is *exactly* the technique used in Baker's [10] first applications of his bound on linear forms in logarithms (we note that he follows Siegel [108]), except that we sharpen the bounds that arise somewhat by using a trick to lower the cardinalities of the finite sets of primes in play — precisely, we represent ideal classes of a number field by (not-too-large) primes by using Chebotarev's density theorem applied to the Hilbert class field of said number field.

In any case, let us prove the bound.

4.2.1 Preparing for the descent.

4.2.1.1 Standard observations.

We note the following observation which is used freely throughout the argument. By the Chebotarev density theorem (and its explicit error term) applied to the Hilbert class field of K_{g_i}, each ideal class of K_{g_i} contains a prime of norm

$$\ll |\Delta_{K_{g_i}}|^{O(1)}.$$

We will also use the fact that

$$|\Delta_{K_{g_i}}| \ll |\Delta_K|^{O(d^3)} \cdot |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|^{O(d^3)}.$$

Finally we will use the following standard observation. Let $W \subseteq C_f(\overline{\mathbb{Q}})$ be the set of Weierstrass points of C_f — that is, the points $(\rho, 0) \in C_f(\overline{\mathbb{Q}})$ for each root ρ of f, and the one or two points at infinity, depending on whether d is odd or even, respectively. Let $\infty \in W$ be a point at infinity. We embed $C_f \hookrightarrow J_f = \mathrm{Pic}^0(C_f)$ via $P \mapsto P - \infty$. Then, using this notation, we have the following isomorphism of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-modules:

$$J_f[2] \simeq \mathbb{F}_2[W - \{\infty\}]/\mathbb{F}_2 \cdot \left(\sum_{P \in W - \{\infty\}} P\right).$$

The single relation arises from the divisor of y, and the fact that the points $P \in W$ are 2-torsion is evident using the divisor of $x - x(P)$. In other words, we find that

$$\mathrm{Ind}_K^{K_f} \mathbb{F}_2 \simeq \mathbb{F}_2 \oplus J_f[2],$$

where we have written $\mathrm{Ind}_K^{K_f}(\bullet) := \bigoplus_i \mathrm{Ind}_K^{K_{f_i}}(\bullet)$ with $f =: \prod_i f_i$ the factorization of f into irreducibles in $\mathfrak{o}_K[t]$. Thus it follows that

$$H^1(K_f, \mathbb{F}_2) \simeq H^1(K, \mathrm{Ind}_K^{K_f} \mathbb{F}_2) \simeq H^1(K, \mathbb{F}_2) \oplus H^1(K, J_f[2]),$$

where the first isomorphism follows by Shapiro's lemma (and $H^1(K_f, \bullet) := \bigoplus_i H^1(K_{f_i}, \bullet)$). By Kummer it follows that

$$H^1(K, J_f[2]) \simeq (K_f^\times/2)_{\mathrm{Nm}=\square},$$

where, to be clear, $L^\times/2 := L^\times/(L^\times)^2$.

Thus by taking invariants of $0 \to J_f[2] \to J_f \to J_f \to 0$ we obtain

$$J_f(K)/2 \hookrightarrow H^1(K, J_f[2]) \simeq (K_f^\times/2)_{\mathrm{Nm}=\square}.$$

We write $G \subseteq H^1(K, J_f[2])$ for the image of this map. Note that the restriction of this map to $C_f(\overline{\mathbb{Q}}) - W$ is simply $(x, y) \mapsto x - \rho$, where ρ is the image of t in $K_f := K[t]/(f(t))$. It is evident that, for $\mathfrak{p} \nmid \Delta_f$ a prime of K_f, $v_\mathfrak{p}(x - \rho)$ is even. We note also that, writing

$$C := \{\beta \in (K_f^\times/2)_{\mathrm{Nm}=\square} : \forall \mathfrak{p} \subseteq \mathfrak{o}_{K_f}, v_\mathfrak{p}(\beta) \in 2\mathbb{Z}\} \hookrightarrow H^1(K, J_f[2]) \simeq (K_f^\times/2)_{\mathrm{Nm}=\square}$$

for the $\beta \in (K_f^\times/2)_{\mathrm{Nm}=\square}$ with $v_{\mathfrak{p}}(\beta)$ even for all $\mathfrak{p}$, we have that[2]

$$0 \to \mathfrak{o}_{K_f}^\times/2 \to C \to \mathrm{Cl}(\mathfrak{o}_{K_f})[2] \to 0$$

via

$$\beta \mapsto \prod_{\mathfrak{p}} \mathfrak{p}^{\frac{v_{\mathfrak{p}}(\beta)}{2}}.$$

4.2.1.2 Choices.

In order to be clear about the order of quantifiers, we fix ahead of time a few choices. Let $1 \leq m \leq 3$ be minimal such that

$$\sum_{i=1}^{m} \deg g_i \geq 3.$$

Let L/K be the splitting field of $\prod_{i=1}^m g_i$. For each $e : \{\mathfrak{p} \subseteq \mathfrak{o}_{K_{g_i}} : \mathfrak{p}^2 | \Delta_f\} \to \{0,1\}$ for which $\prod_{\mathfrak{p}^2|\Delta_f} \mathfrak{p}^{e(\mathfrak{p})}$ has ideal class in $2 \cdot \mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})$, we let $\mathfrak{q}_e$ be a prime of K_{g_i} of norm $\ll |\Delta_{K_{g_i}}|^{O(1)}$ for which

$$\mathfrak{q}_e^2 \equiv \prod_{\mathfrak{p}^2|\Delta_f} \mathfrak{p}^{e(\mathfrak{p})}$$

modulo principal ideals of $\mathfrak{o}_{K_{g_i},S}$. (One exists by our Chebotarev argument.) Let, for such e, $\beta_e \in K_{g_i}^\times$ be a generator of the principal fractional ideal

$$\prod_{\mathfrak{p}^2|\Delta_f} \mathfrak{p}^{e(\mathfrak{p})} \cdot \mathfrak{q}_e^{-2}.$$

Note that β_e can be written as a quotient of elements of $\mathfrak{o}_{K_{g_i}}$ of norm $\ll |\Delta_{K_{g_i}}|^{O(1)}$, by Minkowski.

[2] In other words, since $\mathfrak{o}_{K_f}^\times/2$ is small, we may think of the 2-torsion of the class group of K_f as the elements of $H^1(K, J_f[2])$ satisfying certain (in fact, simply unramified) local conditions everywhere. Of course the 2-Selmer group is defined by matching local conditions at finite primes away from $2\Delta_f$, which accounts for the closeness in size of $\mathrm{Cl}(\mathfrak{o}_{K_f})[2]$ and $\mathrm{Sel}_2(J_f/K)$ alluded to earlier.

Let $P^{(i)}$ be a set of prime representatives of $\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]$ in K_{g_i} of norm $\ll |\Delta_{K_{g_i}}|^{O(1)}$ (this exists by our Chebotarev argument). Because the primes of $P^{(i)}$ represent all elements of $\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]$, the square roots of $\prod_{\mathfrak{p}^2|\Delta_f} \mathfrak{p}^{e(\mathfrak{p})}$ are of the form $\mathfrak{q}_e \cdot \mathfrak{p}$ (modulo principal ideals) for $\mathfrak{p} \in P^{(i)}$. (Let us take $(1) \in P^{(i)}$ as the representative of $0 \in \mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]$.) This finishes our choices.

4.2.2 The 2-descent.

Write $f =: gh$ with $h \in \mathfrak{o}_K[t]$. Write $\rho^{(i)}$ for the image of t in $K_{g_i} = K[t]/(g_i(t))$. We will only ever take $1 \leq i \leq m$ in what follows.

For $(x, y) \in C_f(\mathfrak{o}_{K,S})$, we of course have

$$(y)^2 = (x - \rho^{(i)}) \cdot \left(\frac{g_i(x)}{x - \rho^{(i)}}\right) \cdot \prod_{j \neq i} (g_j(x)) \cdot (h(x))$$

as ideals in $\mathfrak{o}_{K_{g_i},S}$.

Now, if $\mathfrak{p}$ divides two factors in this product, then $\mathfrak{p}^2 | \Delta_f$. It follows therefore that, by unique factorization into prime ideals, there is an

$$e^{(i)} : \{\mathfrak{p} \subseteq \mathfrak{o}_{K_{g_i},S} : \mathfrak{p}^2 | \Delta_f\} \to \{0, 1\}$$

such that

$$(x - \rho^{(i)}) = \prod_{\mathfrak{p}^2 | \Delta_f} \mathfrak{p}^{e^{(i)}(\mathfrak{p})} \cdot \square,$$

where $\square$ denotes the square of an ideal of $\mathfrak{o}_{K_{g_i},S}$. It moreover follows that, for such $e^{(i)}$, $\prod_{\mathfrak{p}^2|\Delta_f} \mathfrak{p}^{e^{(i)}(\mathfrak{p})}$ is a square in the class group.

4.2.3 Preparing the Thue equation.

It follows — recall our definitions of $\beta_{e^{(i)}}$ and $\mathfrak{q}_{e^{(i)}}$ above — that there is an $\mathfrak{a}_i \subseteq \mathfrak{o}_{K_{g_i},S}$ for which

$$(x - \rho^{(i)}) = (\beta_{e^{(i)}}) \cdot \mathfrak{a}_i^2.$$

By virtue of this equality it follows that there is a $\mathfrak{p}_i \in P^{(i)}$ for which

$$\mathfrak{a}_i \equiv \mathfrak{p}_i$$

modulo principal ideals of K_{g_i}. Hence $\mathfrak{a}_i$ is principal in $\mathfrak{o}_{K_{g_i},\{\mathfrak{p}_i\}\cup S}$. So let $\alpha_i \in \mathfrak{o}_{K_{g_i},S}$ be such that

$$x - \rho^{(i)} = \beta_{e_i} \cdot \alpha_i^2 \cdot (\in \mathfrak{o}^{\times}_{K_{g_i},\{\mathfrak{p}_i\}\cup S}),$$

where $(\in \mathfrak{o}^{\times}_{K_{g_i},\{\mathfrak{p}_i\}\cup S})$ denotes an element of the $(\{\mathfrak{p}_i\} \cup S)$-units of K_{g_i}.

Let $U^{(i)}$ be a minimal set of representatives of $\mathfrak{o}^{\times}_{K_{g_i},\{\mathfrak{p}_i\}\cup S}/2$ (that is, modulo squares). By Dirichlet's unit theorem,

$$\#|U^{(i)}| \ll 2^{\deg g_i \cdot (\#|S| + [K:\mathbb{Q}])}.$$

It follows that there are $\varepsilon^{(i)} \in U^{(i)}$ and $u^{(i)} \in \mathfrak{o}^{\times}_{K_{g_i},\{\mathfrak{p}_i\}\cup S}$ for which:

$$x - \rho^{(i)} = \beta_{e^{(i)}} \cdot \varepsilon^{(i)} \cdot (\alpha_i \cdot u^{(i)})^2.$$

Let then $\gamma_i := \beta_{e^{(i)}} \cdot \varepsilon^{(i)}$ and $\eta_i := \alpha_i \cdot u^{(i)}$. Thus

$$x - \rho^{(i)} = \gamma_i \cdot \eta_i^2.$$

Therefore we find that, for each embedding $\sigma : K_{g_i} \to L$ extending our chosen $K \subseteq L$, we have:

$$x - \sigma(\rho^{(i)}) = \sigma(\gamma_i) \cdot \sigma(\eta_i)^2.$$

Let then, for each j, $\tau_{1+\sum_{i=1}^{j-1} \deg g_i}, \ldots, \tau_{\sum_{i=1}^{j} \deg g_i}$ be the K-embeddings $K_{g_j} \to L$. Similarly let, for each j and $\sum_{i=1}^{j-1} \deg g_i < k \leq \sum_{i=1}^{j} \deg g_i$,

$$\begin{aligned} \kappa_k &:= \tau_k(\rho^{(j)}), \\ \lambda_k &:= \tau_k(\gamma_i), \\ \mu_k &:= \tau_k(\eta_i). \end{aligned}$$

In particular, for $1 \leq k \leq 3$,

$$x - \kappa_k = \lambda_k \cdot \mu_k^2$$

as elements of L. Hence, for $1 \leq k \neq \ell \leq 3$, we find that:

$$\kappa_\ell - \kappa_k = \lambda_k \cdot \mu_k^2 - \lambda_\ell \cdot \mu_\ell^2.$$

Let

$$L_{k\ell} := L\left(\sqrt{\lambda_k}, \sqrt{\lambda_\ell}\right).$$

Then we find:

$$(\kappa_\ell - \kappa_k) = \left(\sqrt{\lambda_k}\mu_k - \sqrt{\lambda_\ell}\mu_\ell\right) \cdot \left(\sqrt{\lambda_k}\mu_k + \sqrt{\lambda_\ell}\mu_\ell\right)$$

as ideals of $\mathfrak{o}_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}$, where localizing $\mathfrak{o}_{L_{k\ell}}$ at $\{\mathfrak{p}_k, \mathfrak{p}_\ell\} \cup S$ has the obvious interpretation — we localize at all primes of $\mathfrak{o}_{L_{k\ell}}$ that lie over either primes of S in K, or over $\mathfrak{p}_k$ in K_{g_k}, or $\mathfrak{p}_\ell$ in K_{g_ℓ}.

Let $D_{k\ell}$ be a minimal set of representatives of $\{\xi \in \mathfrak{o}_{L_{k\ell}} : \xi | (\kappa_k - \kappa_\ell)\} / \mathfrak{o}_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}^\times$. This is of course a finite set (since the number of ideals dividing a given ideal is finite as well), of size

$$\#|D_{k\ell}| \leq \prod_{\mathfrak{q}|(\kappa_k - \kappa_\ell)} (v_{\mathfrak{p}}(\kappa_k - \kappa_\ell) + 1),$$

the product taken over the primes $\mathfrak{q}$ of $\mathfrak{o}_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}$ that divide $\kappa_k - \kappa_\ell$ (and hence $\mathfrak{q}^2|\Delta_f$).

Let also $V_{k\ell}$ be a minimal set of representatives of $\mathfrak{o}^\times_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}/3$ (that is, modulo cubes). Note that

$$\#|V_{k\ell}| \ll 3^{4\cdot[L:\mathbb{Q}]+4\cdot[L:K]\cdot\#|S|}.$$

It follows that there are $v_{k\ell,\pm} \in V_{k\ell}, \xi_{k\ell,\pm} \in D_{k\ell}, \zeta_{k\ell,\pm} \in \mathfrak{o}^\times_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}$ for which:

$$\sqrt{\lambda_k}\mu_k \pm \sqrt{\lambda_\ell}\lambda_\ell = (v_{k\ell,\pm} \cdot \xi_{k\ell,\pm}) \cdot \zeta^3_{k\ell,\pm}.$$

Combining this information for the tuples $((1,2),\pm), ((2,3),\mp), ((3,1),-)$, we get the following equality:

$$\begin{aligned} 0 &= (\sqrt{\lambda_1}\mu_1 \pm \sqrt{\lambda_2}\mu_2) \mp (\sqrt{\lambda_2}\mu_2 \pm \sqrt{\lambda_3}\mu_3) + (\sqrt{\lambda_3}\mu_3 - \sqrt{\lambda_1}\mu_1) \\ &= (v_{12,\pm} \cdot \xi_{12,\pm}) \cdot \zeta^3_{12,\pm} \mp (v_{23,\pm} \cdot \xi_{23,\pm}) \cdot \zeta^3_{23,\pm} + (v_{31,-} \cdot \xi_{31,-}) \cdot \zeta^3_{31,-}. \end{aligned}$$

We may of course rearrange this as:

$$(v_{23,\pm} \cdot \xi_{23,\pm}) \cdot (\zeta_{23,\pm} \cdot \zeta^{-1}_{12,\pm})^3 + (v_{31,-} \cdot \xi_{31,-}) \cdot (\zeta_{31,-} \cdot \zeta^{-1}_{12,\pm})^3 = -v_{12,\pm} \cdot \xi_{12,\pm}.$$

4.2.4 Evertse's bound.

The form

$$F_\pm(X,Y) = (v_{23,\pm} \cdot \xi_{23,\pm}) \cdot X^3 + (v_{31,-} \cdot \xi_{31,-}) \cdot Y^3 \in \mathfrak{o}_{L_{123}}[X,Y],$$

where we have written $L_{123} := L(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3})$, is a cubic form with nonzero discriminant (by virtue of all $v_{k\ell,\pm}, \xi_{k\ell,\pm} \neq 0$ since they are divisors of $\kappa_k - \kappa_\ell$ and hence of $\Delta_f \neq 0$). Note that it suffices to take $\mathfrak{o}_{L_{123}}^\times$-cosets of the solutions of $F_\pm(X,Y) = -v_{12,\pm} \cdot \xi_{12,\pm}$ because for both (X,Y) and (tX,tY) to be solutions we would have to have $t^3 = 1$, so that the number of such t is ≤ 3. Now we apply Theorem 3 of Evertse's [45] (which we have already reproduced as Theorem 2.2.6 of Chapter 2), which implies that, by this coset observation, the number of solutions of

$$F_\pm(X,Y) = -v_{12,\pm} \cdot \xi_{12,\pm},$$

with $(X,Y) \in \mathfrak{o}_{L_{123},\{\mathfrak{p}_1,\mathfrak{p}_2,\mathfrak{p}_3\}\cup S}$, is:

$$\ll O(1)^{[L:\mathbb{Q}]+[L:K]\cdot\#|S|+\#|\{\mathfrak{p}\subseteq\mathfrak{o}_{L_{123}}:\mathfrak{p}|\xi_{12,\pm}\}|},$$

where we note that the number of primes dividing $-v_{12,\pm} \cdot \xi_{12,\pm}$ is the same as the count for $\xi_{12,\pm}$ because $-v_{12,\pm}$ is a unit. Because $\xi_{12,\pm}$ is a divisor of $\kappa_1 - \kappa_2$, it follows that this number of primes is

$$\ll \{\mathfrak{p} \subseteq \mathfrak{o}_L : \mathfrak{p}^2 | \Delta_f\}.$$

It thus suffices to show that our original point (x,y) can be recovered from the solutions

$$F_+(\zeta_{23,+} \cdot \zeta_{12,+}^{-1}, \zeta_{31,-} \cdot \zeta_{12,+}^{-1}) = -v_{12,+} \cdot \xi_{12,+}$$

and

$$F_-(\zeta_{23,-} \cdot \zeta_{12,-}^{-1}, \zeta_{31,-} \cdot \zeta_{12,-}^{-1}) = -v_{12,-} \cdot \xi_{12,-}$$

up to $O(1)$ many choices. To do this we simply multiply the two Y-coordinates together to form

$$\frac{\zeta_{31,-}^2}{\zeta_{12,+}\cdot\zeta_{12,-}}.$$

Note that:

$$\begin{aligned}(v_{12,+}\cdot\xi_{12,+})\cdot(v_{12,-}\cdot\xi_{12,-})\cdot(\zeta_{12,+}\cdot\zeta_{12,-}) &= (\sqrt{\lambda_1}\mu_1+\sqrt{\lambda_2}\mu_2)\cdot(\sqrt{\lambda_1}\mu_1-\sqrt{\lambda_2}\mu_2)\\ &= \lambda_1\mu_1^2-\lambda_2\mu_2^2\\ &= \kappa_2-\kappa_1.\end{aligned}$$

Thus it follows that

$$\zeta_{12,+}\cdot\zeta_{12,-}=\frac{\kappa_2-\kappa_1}{v_{12,+}\cdot v_{12,-}\cdot\xi_{12,+}\cdot\xi_{12,-}}.$$

Hence the product of the two Y-coordinates is:

$$\zeta_{31,-}^2\cdot\left(\frac{v_{12,+}\cdot v_{12,-}\cdot\xi_{12,+}\cdot\xi_{12,-}}{\kappa_2-\kappa_1}\right).$$

We note that the term in parentheses is fixed (in terms of our choices up til now). Thus we may recover $\zeta_{31,-}^2$, and hence $\zeta_{31,-}$ up to at most two choices. Having done so we return to the Y-coordinates of both solutions and recover $\zeta_{12,+}$ and $\zeta_{12,-}$. Then the equality

$$\begin{aligned}2\sqrt{\lambda_1}\mu_1 &= (\sqrt{\lambda_1}\mu_1+\sqrt{\lambda_2}\mu_2)+(\sqrt{\lambda_1}\mu_1-\sqrt{\lambda_2}\mu_2)\\ &= (v_{12,+}\cdot\xi_{12,+}\cdot\zeta_{12,+}^3)+(v_{12,-}\cdot\xi_{12,-}\cdot\zeta_{12,-}^3)\end{aligned}$$

implies that we can recover $2\sqrt{\lambda_1}\mu_1$. Squaring this we find that we can recover

$$4\lambda_1\mu_1^2=4(x-\kappa_1).$$

Since κ_1 is fixed we can recover x, and then there are at most two choices for y given x, so we can recover the point up to $O(1)$ many choices.

4.2.5 Final bookkeeping.

From the above analysis it follows a point $(x, y) \in C_f(\mathfrak{o}_{K,S})$ is determined up to $O(1)$ many choices by the data

$$(e^{(1)}, \ldots, e^{(m)}, \mathfrak{p}_1, \ldots, \mathfrak{p}_m, \varepsilon^{(1)}, \ldots, \varepsilon^{(m)}, v_{12,+}, v_{12,-}, v_{23,+}, v_{23,-}, v_{31,-}, \xi_{12,+}, \xi_{12,-}, \xi_{23,+}, \xi_{23,-}, \xi_{31,-}).$$

We will now insert the bounds

$$[L : \mathbb{Q}] \leq d^3 \cdot [K : \mathbb{Q}]$$

and

$$\#|\{\mathfrak{p} \subseteq \mathfrak{o}_L : \mathfrak{p}^2 | \Delta_f\}| \leq d^3 \cdot \#|\{\mathfrak{p} \subseteq \mathfrak{o}_K : \mathfrak{p} | \Delta_f\}|.$$

We will write

$$\omega_K(\Delta_f) := \#|\{\mathfrak{p} \subseteq \mathfrak{o}_K : \mathfrak{p} | \Delta_f\}|.$$

The number of choices for each e_i is $\ll 2^{d^3 \cdot \omega_K(\Delta_f)}$, thus all of them together contribute

$$\ll 8^{d^3 \cdot \omega_K(\Delta_f)}.$$

Recall that

$$\#|P^{(i)}| = \#|\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]|.$$

Therefore a crude bound for the number of choices for each $\mathfrak{p}_i$ is simply $\leq \#|\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]|$, whence a similar crude bound for the number of choices for $(\mathfrak{p}_1, \ldots, \mathfrak{p}_m)$ is

$$\leq \prod_{i=1}^{m} \#|\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]|.$$

The number of choices for each $\varepsilon^{(i)}$ is $\ll 2^{d^3 \cdot (\#|S|+1)}$. It follows that the number of choices for $(\varepsilon^{(1)}, \varepsilon^{(2)}, \varepsilon^{(3)})$ is

$$\ll 8^{d^3 \cdot (\#|S|+1)}.$$

The number of choices for each $v_{k\ell,\pm}$ is $\ll O(1)^{d^3 \cdot ([K:\mathbb{Q}]+\#|S|)}$, so the number of choices for $(v_{12,+}, v_{12,-}, v_{23,+}, v_{23,-}, v_{31,-})$ is

$$\ll O(1)^{d^3 \cdot ([K:\mathbb{Q}]+\#|S|)}.$$

Finally, the number of choices for each $\xi_{k\ell,\pm}$ is $\leq \prod_{\mathfrak{p}|(\kappa_i - \kappa_j)}(v_{\mathfrak{p}}(\kappa_k - \kappa_\ell) + 1)$, the product taken over the primes $\mathfrak{p} \subseteq \mathfrak{o}_{L_{k\ell},\{\mathfrak{p}_k,\mathfrak{p}_\ell\}\cup S}$ that divide $\kappa_i - \kappa_j$ (and hence $\mathfrak{p}^2|\Delta_f$). We use the usual divisor bound to bound this by simply $\ll O(1)^{d^3 \cdot [K:\mathbb{Q}] \cdot \frac{\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}$, since when $\sum_i e_i f_i = n$ one has $\prod_i (e_i k + 1) \leq (k+1)^n$. Therefore the number of choices for $(\xi_{12,+}, \xi_{12,-}, \xi_{23,+}, \xi_{23,-}, \xi_{31,-})$ is

$$\ll O(1)^{d^3 \cdot [K:\mathbb{Q}] \cdot \frac{\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}.$$

Therefore the total number of tuples

$$(e^{(1)}, \ldots, e^{(m)}, \mathfrak{p}_1, \ldots, \mathfrak{p}_m, \varepsilon^{(1)}, \ldots, \varepsilon^{(m)}, v_{12,+}, v_{12,-}, v_{23,+}, v_{23,-}, v_{31,-}, \xi_{12,+}, \xi_{12,-}, \xi_{23,+}, \xi_{23,-}, \xi_{31,-})$$

is:

$$\ll \left(\prod_{i=1}^{m} \#|\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]| \right) \cdot O(1)^{d^3 \cdot \#|S|} \cdot O(1)^{d^3 \cdot [K:\mathbb{Q}] \cdot \frac{\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log |\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}.$$

Since we have seen that there are at most $O(1)$ points corresponding to any given such tuple, we find that:

$$\#|\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S})| \ll \left(\prod_{i=1}^{m} \#|\mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]|\right) \cdot O(1)^{d^3\cdot\#|S|} \cdot O(1)^{d^3\cdot[K:\mathbb{Q}]\cdot\frac{\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}.$$

But of course $\mathfrak{o}_{K_f,S} \simeq \prod_i \mathfrak{o}_{K_{g_i},S}$, so that

$$\mathrm{Cl}(\mathfrak{o}_{K_f,S})[2] \simeq \bigoplus_i \mathrm{Cl}(\mathfrak{o}_{K_{g_i},S})[2]$$

— i.e. we also have that

$$\#|\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S})| \ll \#|\mathrm{Cl}(\mathfrak{o}_{K_g,S})[2]| \cdot O(1)^{d^3\cdot\#|S|} \cdot O(1)^{d^3\cdot[K:\mathbb{Q}]\cdot\frac{\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}{\log\log|\mathrm{Nm}_{K/\mathbb{Q}}\Delta_f|}}.$$

To complete the proof of the theorem it remains only to note that (ignoring Weierstrass points) the number of tuples $(\mathfrak{p}_1, \ldots, \mathfrak{p}_m)$ actually mapped to by S-integral points of C_f is also

$$\leq \#|G| \leq 2^{d-1} \cdot 2^{\operatorname{rank} J_f(K)}.$$

This amounts to saying that the image of $\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S}) - W$ in $(K_f^\times/2)_{\mathrm{Nm}=\square}$ via

$$(x, y) \mapsto x - \rho$$

is of size $\leq \#|G|$. This is evident because the map factors as

$$\mathcal{C}_f^{\text{aff.}}(\mathfrak{o}_{K,S}) \hookrightarrow C_f(K) \hookrightarrow J_f(K) \twoheadrightarrow J_f(K)/2 \hookrightarrow H^1(K, J_f[2]) \simeq (K_f^\times/2)_{\mathrm{Nm}=\square},$$

as already noted. □

4.3 Class groups versus Selmer groups.

This section is unrelated to the main theorem of this chapter but was in fact the main inspiration for the chapter. We note that the following lemma is very much inspired by Schaefer's thesis [89], where we learned about the very close *quantitative* relation between class and Selmer groups (and which led us to think that existing bounds on the number of integral points on hyperelliptic curves were therefore clearly suboptimal).

The lemma is as follows.

Lemma 4.3.1. *Let K be a number field and $f \in \mathfrak{o}_K[t]$ of degree $d \geq 3$ with nonzero discriminant $\Delta_f \neq 0$. Let C_f/K be the hyperelliptic curve $y^2 = f(x)$ associated to f and $J_f := \operatorname{Jac} C_f$. Let $K_f := K[t]/(f)$. Then:*

$$\#|\mathrm{Sel}_2(J_f/K)| = \#|\mathrm{Cl}(\mathfrak{o}_{K_f})[2]| \cdot \Theta_d(1)^{[K:\mathbb{Q}]+\#|\{\mathfrak{p}:\mathfrak{p}|\Delta_f\}|}.$$

In fact Schaefer essentially proves this in his Berkeley PhD thesis [88,89], except that he is after a far more precise bound, which leads him to impose additional hypotheses (that do not necessarily hold in general) on f.

Proof. We have already seen that $H^1(K, J_f[2]) \simeq (K_f^\times/2)_{\mathrm{Nm}=\square}$ canonically. Write $S := \mathrm{Sel}_2(J_f/K) \hookrightarrow H^1(K, J_f[2])$ and C for the image of

$$\{\beta \in (K_f^\times/2)_{\mathrm{Nm}=\square} : \forall \mathfrak{p} \subseteq \mathfrak{o}_{K_f}, v_\mathfrak{p}(\beta) \in 2\mathbb{Z}\} \subseteq (K_f^\times/2)_{\mathrm{Nm}=\square}$$

under this canonical isomorphism. Again, we have already seen that

$$0 \to \mathfrak{o}_{K_f}^\times/2 \to C \to \mathrm{Cl}(\mathfrak{o}_{K_f})[2] \to 0.$$

Let $L := K(J_f[2])$. Evidently L/K is Galois. Note that C may also be described as the subgroup of elements of $H^1(K, J_f[2])$ which map, under the inflation-restriction sequence

$$0 \to H^1(\mathrm{Gal}(L/K), J_f[2]) \to H^1(K, J_f[2]) \to \mathrm{Hom}_{\mathrm{Gal}(L/K)}(\mathrm{Gal}(\overline{\mathbb{Q}}/L), J_f[2]),$$

to everywhere unramified homomorphisms.

Write, for each place v of K, S_v for the image in $H^1(K_v, J_f[2])$ of $J_f(K_v)/2$ under the boundary map arising from multiplication-by-2, i.e. from

$$0 \to J_f[2] \to J_f \to J_f \to 0.$$

Write C_v for the subgroup of elements of $H^1(K_v, J_f[2])$ that map, under the inflation-restriction sequence

$$0 \to H^1(\mathrm{Gal}(L_w/K_v), J_f[2]) \to H^1(K_v, J_f[2]) \to \mathrm{Hom}_{\mathrm{Gal}(L_w/K_v)}(\mathrm{Gal}(\overline{L}_w/L_w), J_f[2]),$$

to unramified homomorphisms.

Then, since C and S are defined by local conditions only, it follows that

$$C/(C \cap S) \hookrightarrow \prod_v C_v/(C_v \cap S_v)$$

and

$$S/(C \cap S) \hookrightarrow \prod_v S_v/(C_v \cap S_v).$$

For $v \nmid 2\Delta_f \cdot \infty$ we have already seen that $C_v = S_v$, so that these are finite products. Otherwise we easily have (see e.g. Schaefer's Lemma 3.1 [88] for C_v) that $\#|C_v|, \#|S_v| \ll_d 1$. This completes the proof. $\square$

We note that we have implicitly used the bounds $\#|H^1(G, J_f[2])|, \#|H^2(G, J_f[2])| \ll_d 1$ for $G \subseteq \mathrm{Gal}(L/K)$, which follow e.g. via bounding the number of inhomogeneous i-cocycles $G^{\times i} \to J_f[2]$ trivially. Implicit in our improvement on Schaefer's thesis is that he uses hypotheses of the form $H^1(G, J_f[2]) = H^2(G, J_f[2]) = 0$ for various $G \subseteq \mathrm{Gal}(L/K)$ when he could just multiply various bounds by $\#|H^1(G, J_f[2])|$ or $\#|H^2(G, J_f[2])|$. (Of course, he is trying for sharp bounds, which significantly complicates matters.)

Chapter 5

$C_f(\mathbb{Q})$: rational points on genus two curves.

This chapter is based on my [5].

Abstract.

We prove that, when genus two curves $C/\mathbb{Q}$ with a marked Weierstrass point are ordered by height, the average number of rational points $\#|C(\mathbb{Q})|$ is bounded.

5.1 Introduction.

5.1.1 Main theorem.

In this chapter we prove the following theorem.

Theorem 5.1.1.

$$\operatorname*{Avg}_{f\in\mathcal{F}_{\text{univ.}}:H(f)\leq T} \#|C_f(\mathbb{Q})| \ll 1.$$

Here

$$\mathcal{F}_{\text{univ.}} := \{f(x) =: \sum_{i=0}^{5} a_i\cdot x^{5-i} \in \mathbb{Z}[x] : a_0 = 1, a_1 = 0, \Delta_f \neq 0\},$$

and, for $f \in \mathbb{Z}[x]$ with $f(x) =: \sum_{i=0}^{5} a_i \cdot x^{5-i}$, we write $C_f : y^2 = f(x)$ and

$$H(f) := \max_i |a_i|^{\frac{1}{i}}.$$

5.1.2 Main technique.

The key point is to bound the number of rational points on each C_f sufficiently well as to be able to apply the following theorem of Bhargava-Gross.

Theorem 5.1.2 (Bhargava-Gross, see Theorem 1 in their [16].)**.**

$$\operatorname*{Avg}_{f \in \mathcal{F}_{\text{univ.}} : H(f) \leq T} \#|\mathrm{Sel}_2(J_f/\mathbb{Q})| = 3 + o_{T\to\infty}(1).$$

Here $J_f := \mathrm{Jac}\, C_f/\mathbb{Q}$.

To do this we split $C_f(\mathbb{Q}) =: C_f(\mathbb{Q})^{\text{small}} \cup C_f(\mathbb{Q})^{\text{medium}} \cup C_f(\mathbb{Q})^{\text{large}}$. We prove that

$$\#|C_f(\mathbb{Q})^{\text{medium}}| \ll 2^{\operatorname{rank} J_f(\mathbb{Q})}$$

(see Proposition 5.2.36 for a slightly more optimized bound) using repulsion, and prove by hand (see Proposition 5.2.3) that

$$\operatorname*{Avg}_{f \in \mathcal{F}_{\text{univ.}} : H(f) \leq T} \#|C_f(\mathbb{Q})^{\text{small}}| \leq o_{T\to\infty}(1).$$

The latter generalizes to larger degrees (i.e. to larger genus hyperelliptic curves). The inequality

$$\#|C_f(\mathbb{Q})^{\text{large}}| \ll 2^{\operatorname{rank} J_f(\mathbb{Q})}$$

follows from the more general Theorem 6.1.1 of Chapter 6. Combining the three bounds (and the usual $2^{\operatorname{rank} A(K)} \leq \#|\mathrm{Sel}_2(A/K)|$) gives the theorem.

Unfortunately we will have to use techniques that do not generalize to genus $g > 3$ to prove Proposition 5.2.36 (specifically, known explicit formulas/equations for the addition law/Kummer variety, and good control on local height differences). For example, our proof of an explicit form of the Mumford gap principle depends on said explicit formulas: if $P =: (x, y), Q =: (X, Y) \in C_f(\mathbb{Q})$, then the point on the Kummer surface $J_f/\{\pm 1\} =: K_f \subseteq \mathbb{P}^3$ (with embedding given by $2 \cdot \Theta$, and theta divisor given by the image of the Abel-Jacobi map corresponding to the point at ∞ — i.e. the image of $C_f \hookrightarrow J_f$ via $P \mapsto P - \infty$) corresponding to $P + Q$ is:

$$\left[1, X + x, Xx, \frac{2a_5 + a_4 \cdot (X + x) + 2a_3 \cdot Xx + a_2 \cdot Xx(X + x) + X^2x^2(X + x) - 2Yy}{(X - x)^2}\right].$$

From this it is easy to read off an upper bound on $h(P + Q)$, and thus (ignoring the height difference) an explicit form of the Mumford gap principle.

5.2 Proof of Theorem 5.1.1

Let us now prove Theorem 5.1.1. We write $\mathcal{F}^{\leq T}_{\text{univ.}} := \{f \in \mathcal{F}_{\text{univ.}} : H(f) \leq T\}$.

The first thing to note is that $|\mathcal{F}^{\leq T}_{\text{univ.}}| \asymp T^{14}$. The second thing to note is that if $\left(\frac{a}{b}, \frac{c}{d}\right) \in C_f(\mathbb{Q})$ is in lowest terms, then since (by clearing denominators of the defining equation) $d^2 | c^2 b^5$ and $b^5 | d^2 \cdot (a^5 + (\in b\mathbb{Z}))$, it follows that $d^2 = b^5$ — i.e., that all rational points on C_f are of the form $\left(\frac{a}{e^2}, \frac{b}{e^5}\right)$ in lowest terms.

Now, given $f \in \mathcal{F}^{\leq T}_{\text{univ.}}$, write $C_f(\mathbb{Q}) =: \mathrm{I}_f \cup \mathrm{II}_f \cup \mathrm{III}_f$, with[1]:

$$\mathrm{I}^{\uparrow}_f := \{P \in C_f(\mathbb{Q}) : |x(P)| \geq \delta^{-\delta^{-1}} H(f), h(P) < (c_{\uparrow} - \delta)h(f)\},$$

$$\mathrm{I}^{\downarrow}_f := \{P \in C_f(\mathbb{Q}) : |x(P)| < \delta^{-\delta^{-1}} H(f), h(P) < (c_{\downarrow} - \delta)h(f)\},$$

[1] In the general degree d case, we have that $|\mathcal{F}^{\leq T}_{\text{univ.}}| \asymp T^{\frac{d(d+1)}{2} - 1}$, that $c_{\uparrow} = \frac{d(d+1)-5}{3}$, and that $c_{\downarrow} = \frac{d(d+1)}{3} - 2$, via precisely the same methods. (These values of $c_{\uparrow}$ and $c_{\downarrow}$ are easily improved upon in the general case.)

$$\begin{aligned}
\mathrm{I}_f &:= \mathrm{I}_f^{\uparrow} \cup \mathrm{I}_f^{\downarrow}, \\
\mathrm{II}_f^{\uparrow} &:= \{P \in C_f(\mathbb{Q}) : |x(P)| \geq \delta^{-\delta^{-1}} H(f), P \notin \mathrm{I}_f, h(P) < \delta^{-\delta^{-1}} h(f)\}, \\
\mathrm{II}_f^{\downarrow} &:= \{P \in C_f(\mathbb{Q}) : |x(P)| \leq \delta^{\delta^{-1}} H(f), P \notin \mathrm{I}_f, h(P) < \delta^{-\delta^{-1}} h(f)\}, \\
\mathrm{II}_f^{\bullet} &:= \{P \in C_f(\mathbb{Q}) : \delta^{\delta^{-1}} H(f) < |x(P)| < \delta^{-\delta^{-1}} H(f), P \notin \mathrm{I}_f, h(P) < \delta^{-\delta^{-1}} h(f)\}, \\
\mathrm{II}_f &:= \mathrm{II}_f^{\uparrow} \cup \mathrm{II}_f^{\bullet} \cup \mathrm{II}_f^{\downarrow}, \\
\mathrm{III}_f &:= C_f(\mathbb{Q}) - (\mathrm{I}_f \cup \mathrm{II}_f), \\
c_{\uparrow} &:= \frac{25}{3}, \\
c_{\downarrow} &:= 8.
\end{aligned}$$

We will call points in I_f *small*, points in II_f *medium*-sized, and points in III_f *large*. In words, the decorations $\uparrow, \bullet, \downarrow$ indicate the size of the x-coordinates of such points, and we have broken into: large points, and small/medium-sized points with surprisingly large/surprisingly small/otherwise (the second occurring only in the case of medium points) x-coordinates. Let us first handle the small points.

5.2.1 Small points.

We first turn to those small points with large x-coordinate, i.e. the points in $\mathrm{I}_f^{\uparrow}$.

Lemma 5.2.1.

$$\sum_{f \in \mathcal{F}_{\mathrm{univ.}}^{\leq T}} \#|\mathrm{I}_f^{\uparrow}| \ll T^{14-\Omega(\delta)}.$$

Proof. Certainly

$$\begin{aligned}
&\sum_{f \in \mathcal{F}_{\mathrm{univ.}}^{\leq T}, H(f) \asymp T} \#|\mathrm{I}_f^{\uparrow}| \\
&\leq \# \left| \left\{ (s, d, a_2, t, a_3, a_4, a_5) \;\middle|\; \begin{array}{c} t^2 = s^5 + a_2 d^4 s^3 + \cdots + a_5 d^{10}, |a_i| \leq T^i, \Delta_f \neq 0, \\ (s, d) = 1, (t, d) = 1, T \ll |s| \leq T^{c_{\uparrow} - \delta}, d^2 \leq \delta^{\delta^{-1}} |s| T^{-1} \end{array} \right\} \right|,
\end{aligned}$$

and it suffices to prove the claimed bound for this sum (via a dyadic partition of size $\log T$ — note that we have restricted $H(f) \asymp T$, rather than just $\leq T$).

We will say that an integer z *extends* the tuple $(\tilde{s}, \tilde{d}, \ldots)$ if there is an element $(s, d, a_2, t, a_3, a_4, a_5)$ in the above set agreeing in the respective entries — i.e., $s = \tilde{s}, d = \tilde{d}, \ldots$ — and with value z in the next entry. (The tuple can be length one, or even empty, of course.) For example, given an element $(s, d, a_2, t, a_3, a_4, a_5)$ of the above set, a_5 extends (s, d, a_2, t, a_3, a_4), a_4 extends (s, d, a_2, t, a_3), and so on.

Given a tuple (s, d, a_2, t, a_3, a_4), of course there is at most one a_5 that extends this tuple — indeed, the defining equation lets us solve for $a_5 d^{10}$, and hence for a_5. Next, given (s, d, a_2, t, a_3), if both a_4 and a_4' extend $(s, \ldots, a_3)$ to $(s, \ldots, a_3, a_4, a_5)$ and $(s, \ldots, a_3, a_4', a_5')$, respectively, then, on taking differences of the respective defining equations, we find that

$$0 = (a_4 - a_4')d^8 s + (a_5 - a_5')d^{10},$$

and so

$$|a_4 - a_4'| \ll \frac{T^5 d^2}{|s|}$$

since $|a_5|, |a_5'| \ll T^5$. Moreover we also have that $0 \equiv (a_4 - a_4')sd^8 \pmod{d^{10}}$, i.e. that (since $(s, d) = 1$)

$$a_4 \equiv a_4' \pmod{d^2}.$$

Hence the number of a_4 extending the tuple (s, d, a_2, t, a_3) is

$$\ll 1 + \frac{T^5}{|s|}.$$

Similarly, if a_3 and a_3' extend (s, d, a_2, t), then, with similar notation as before,

$$0 = (a_3 - a_3')s^2 d^6 + (a_4 - a_4')sd^8 + (a_5 - a_5')d^{10},$$

whence (here we use that $|s| \gg Td^2$)

$$|a_3 - a_3'| \ll \frac{T^4 d^2}{|s|},$$

and, again, on reducing the equation modulo d^8 it follows that

$$a_3 \equiv a_3' \pmod{d^2}.$$

Thus the number of a_3 extending (s, d, a_2, t) is

$$\ll 1 + \frac{T^4}{|s|}.$$

Next, if both $t, t' > 0$ extend (s, d, a_2), then

$$(t - t')(t + t') = (a_3 - a_3')s^2 d^6 + (a_4 - a_4')sd^8 + (a_5 - a_5')d^{10},$$

and so

$$|t - t'| \ll \frac{T^3 |s|^2 d^6}{t + t'}.$$

Since $t^2 = s^5 + a_2 s^3 d^4 + a_3 s^2 d^6 + a_4 s d^8 + a_5 d^{10}$ and $|s| \geq \delta^{-\delta^{-1}} T d^2$, it follows that, once $\delta \ll 1$ (i.e. once δ is sufficiently small), $t \asymp |s|^{\frac{5}{2}}$. Thus we have found that

$$|t - t'| \ll T^3 |s|^{-\frac{1}{2}} d^6.$$

Moreover, since

$$t^2 \equiv s^5 + a_2 s^3 d^4 \pmod{d^6},$$

and there are $\ll_\varepsilon d^\varepsilon$ many such square-roots of $s^5 + a_2 s^3 d^4$ modulo d^6 (there are $\ll 2^{\omega(d)}$, to be precise, since $(s, d) = 1$), it follows that any such t falls into $\ll_\varepsilon d^\varepsilon \ll_\varepsilon T^\varepsilon$ congruence classes modulo d^6 and inside one of two (depending on sign) intervals of length $\ll T^3|s|^{-\frac{1}{2}}d^6$, whence the number of t extending (s, d, a_2) is

$$\ll_\varepsilon T^\varepsilon \left(1 + \frac{T^3}{|s|^{\frac{1}{2}}}\right).$$

Finally, since $|a_2| \ll T^2$, of course there are only $\ll T^2$ many a_2 extending (s, d).

Hence, in sum, we have shown that (on multiplying these bounds together and summing over the possible s and d)

$$\begin{aligned}\sum_{f \in \mathcal{F}^{\leq T}_{\text{univ.}}} \#|\mathrm{I}^{\uparrow}_f| &\ll_\varepsilon T^{2+\varepsilon} \sum_{T \ll |s| \ll T^{c_\uparrow - \delta}} \sum_{d \ll |s|^{\frac{1}{2}} T^{-\frac{1}{2}}} \left(1 + \frac{T^3}{|s|^{\frac{1}{2}}}\right)\left(1 + \frac{T^4}{|s|}\right)\left(1 + \frac{T^5}{|s|}\right) \\ &\ll_\varepsilon T^{2+\varepsilon} \sum_{T \ll |s| \ll T^{c_\uparrow - \delta}} |s|^{\frac{1}{2}} T^{-\frac{1}{2}} + T^{\frac{5}{2}} + \frac{T^{\frac{9}{2}}}{|s|^{\frac{1}{2}}} + \frac{T^{\frac{15}{2}}}{|s|} + \frac{T^{\frac{17}{2}}}{|s|^{\frac{3}{2}}} + \frac{T^{\frac{23}{2}}}{|s|^2} \\ &\ll_\varepsilon T^\varepsilon \left(T^{\frac{3}{2}(c_\uparrow - \delta) + \frac{3}{2}} + T^{(c_\uparrow - \delta) + \frac{9}{2}} + T^{\frac{1}{2}(c_\uparrow - \delta) + \frac{13}{2}} + T^{\frac{25}{2}}\right),\end{aligned}$$

and this is admissible (remember that the size of the family is T^{14}) since $c_\uparrow = \frac{25}{3}$. □

Now we turn to those small points that do *not* have such large x-coordinates. The ideas for this case are much the same as the ideas for the previous one.

Lemma 5.2.2.

$$\sum_{f \in \mathcal{F}^{\leq T}_{\text{univ.}}} \#|\mathrm{I}^{\downarrow}_f| \ll T^{14 - \Omega(\delta)}.$$

Proof. In the same way as Lemma 5.2.1, we write

$$\begin{aligned}\sum_{f \in \mathcal{F}^{\leq T}_{\text{univ.}}, H(f) \asymp T} \#|\mathrm{I}^{\downarrow}_f| \leq \#|\{(s, d, a_2, t, a_3, a_4, a_5) : t^2 = s^5 + a_2 d^4 s^3 + \cdots + a_5 d^{10}, \\ |a_i| \leq T^i, \Delta_f \neq 0, (s, d) = 1, (t, d) = 1, |s| \leq T^{c_\downarrow - \delta}, d^2 > \delta^{\delta^{-1}} |s| T^{-1}\}|.\end{aligned}$$

And, again, we argue by iteratively bounding the number of integers extending a tuple, where our definition of *extending* is much the same as before (except the ambient set has changed). First, the number of a_5 extending (s, d, a_2, t, a_3, a_4) is ≤ 1, again because we can solve for $a_5 d^{10}$ and hence a_5 in the defining equation.

Now, if both a_4 and a_4' extend (s, d, a_2, t, a_3) to $(s, d, a_2, t, a_3, a_4, a_5)$ and $(s, d, a_2, t, a_3, a_4', a_5')$, then, on taking differences of the defining equations, we find that

$$0 = (a_4 - a_4')sd^8 + (a_5 - a_5')d^{10},$$

and so $0 \equiv (a_4 - a_4')sd^8 \pmod{d^{10}}$ — i.e.,

$$a_4 \equiv a_4' \pmod{d^2},$$

in much the same way as before. This time we simply use that $|a_4| \ll T^4$ to get that the number of a_4 extending (s, d, a_2, t, a_3) is

$$\ll 1 + \frac{T^4}{d^2}$$

since if there is at least one solution, any other must lie in an interval of length T^4 intersected with a congruence class modulo d^2.

In the same way, if a_3 and a_3' extend (s, d, a_2, t), we find that

$$a_3 \equiv a_3' \pmod{d^2},$$

and since $|a_3| \ll T^3$ it follows that the number of a_3 extending (s, d, a_2, t) is

$$\ll 1 + \frac{T^3}{d^2}.$$

Now if t extends (s, d, a_2), then since

$$t^2 \equiv s^5 + a_2 s^3 d^4 \pmod{d^6}$$

and there are $\ll_\varepsilon d^\varepsilon \ll_\varepsilon T^\varepsilon$ many square roots of an element of $(\mathbb{Z}/d^6)^\times$, we find that t must fall in one of $\ll_\varepsilon T^\varepsilon$ congruence classes modulo d^6. Also, since

$$t^2 \ll T^5 d^{10}$$

(since $d^2 \gg T^{-1}|s|$) via the defining equation, it follows that

$$|t| \ll T^{\frac{5}{2}} d^5.$$

Hence the number of t extending (s, d, a_2) is

$$\ll_\varepsilon T^\varepsilon \left(1 + \frac{T^{\frac{5}{2}}}{d}\right).$$

Finally, $|a_2| \ll T^2$ implies the number of a_2 extending (s, d) is of course $\ll T^2$.

In sum, we have found that (here we use that we will choose $\delta \asymp 1$ in the end)

$$\begin{aligned}
\sum_{f \in \mathcal{F}^{\leq T}_{\text{univ.}}} \#|\mathrm{I}^{\downarrow}_f| &\ll_\varepsilon T^{2+\varepsilon} \sum_{d \leq T^{\frac{c_\downarrow - \delta}{2}}} \sum_{|s| \ll \min(T^{c_\downarrow - \delta}, \delta^{-\delta^{-1}} T d^2)} \left(1 + \frac{T^{\frac{5}{2}}}{d}\right)\left(1 + \frac{T^3}{d^2}\right)\left(1 + \frac{T^4}{d^2}\right) \\
&\ll_\varepsilon T^{2+\varepsilon} \sum_{T^{\frac{c_\downarrow - 1 - \delta}{2}} \ll d \leq T^{\frac{c_\downarrow - \delta}{2}}} \sum_{|s| \ll T^{c_\downarrow - \delta}} \left(1 + \frac{T^{\frac{5}{2}}}{d}\right)\left(1 + \frac{T^3}{d^2}\right)\left(1 + \frac{T^4}{d^2}\right) \\
&\quad + T^{2+\varepsilon} \sum_{d \ll T^{\frac{c_\downarrow - 1 - \delta}{2}}} \sum_{|s| \ll T d^2} \left(1 + \frac{T^{\frac{5}{2}}}{d}\right)\left(1 + \frac{T^3}{d^2}\right)\left(1 + \frac{T^4}{d^2}\right) \\
&\ll_\varepsilon T^\varepsilon \left(T^{\frac{3(c_\downarrow - \delta)}{2} + 2} + T^{c_\downarrow + \frac{9}{2} - \delta} + T^{\frac{19}{2}} + T^{\frac{c_\downarrow - 1 - \delta}{2} + 7} + T^{10} + T^{\frac{25}{2}}\right),
\end{aligned}$$

and this is admissible since $c_\downarrow = 8$. □

Lemmas 5.2.1 and 5.2.2 together prove the following.

Proposition 5.2.3.

$$\operatorname*{Avg}_{f \in \mathcal{F}^{\leq T}_{\text{univ.}}} \#|\mathrm{I}_f| \ll T^{-\Omega(1)}.$$

This completes the small point counting.

5.2.2 Medium points.

We recall the notation $C_f : y^2 = f(x), J_f := \mathrm{Jac}(C_f), K_f := J_f/\{\pm 1\}$ for the curve, Jacobian, and Kummer variety associated to f. Write $\infty \in C_f(\mathbb{Q})$ for the marked rational Weierstrass point on $C_f \subseteq \mathbb{P}^2$. Write $j : C_f \to J_f$ for the Abel-Jacobi map associated to ∞, and write $\kappa : C_f \to K_f$ for $j : C_f \to J_f \twoheadrightarrow K_f$, i.e. j composed with the canonical projection $J_f \to K_f$. We will embed $K_f \subseteq \mathbb{P}^3$ as in Cassels-Flynn [36]. Explicitly, $C_f \to K_f \to \mathbb{P}^3$ is the map $(x, y) \mapsto [0, 1, x, x^2]$. We define, for $P \in C_f(\overline{\mathbb{Q}})$, $h(P) := h(x(P))$. We define h_K to be the pullback of the logarithmic Weil height on $\mathbb{P}^3$ to K_f via the Cassels-Flynn embedding, and, for $P \in K_f(\overline{\mathbb{Q}})$, $\hat{h}(P) := \lim_{n\to\infty} \frac{h_K(2^n P)}{4^n}$. We will omit j and κ in expressions like $h_K(P + Q)$ or $h_K(P)$ for $P, Q \in C_f(\mathbb{Q})$ — of course these mean $h_K(j(P) + j(Q))$ (with the inner expression projected to K_f) and $h_K(\kappa(P))$, respectively, but since there will always be a unique sensible interpretation there will be no ambiguity in dropping the various embeddings.

Note that $\kappa(P) = [0, 1, x(P), x(P)^2]$ implies $h_K(\kappa(P)) = 2h(P)$ for all $P \in C_f(\overline{\mathbb{Q}})$. Thus e.g. our small point counting allows us to focus only on those $P \in C_f(\mathbb{Q})$ with $h_K(\kappa(P)) \geq (16 - 2\delta)h(f)$, since $c_\uparrow \geq c_\downarrow = 8$.

Having set all this notation, let us turn to bounding the number of medium points on these curves. To do this we will establish an explicit Mumford gap principle via the explicit addition law on $J_f \hookrightarrow \mathbb{P}^{15}$ provided by Flynn [49] on his website, and then use the usual sphere packing argument to conclude the section.

5.2.2.1 Upper bounds on $\hat{h}(P+Q)$.

First let us deal with the gap principle. We will first prove upper bounds on expressions of the form $\hat{h}(P+Q)$. We split into two cases, depending on whether or not the given points have unusually (Archimedeanly) large x-coordinate or not.

Lemma 5.2.4. *Let $P \neq \pm Q \in C_f(\mathbb{Q})$ with $|x(P)|, |x(Q)| \ll \delta^{-\delta^{-1}} H(f)$ and $h(P) \geq h(Q)$. Then:*

$$\hat{h}(P+Q) \leq 3h_K(P) + (5 + O(\delta))h(f).$$

Proof. Write $P =: (X, Y) =: \left(\frac{S}{D^2}, \frac{U}{D^5}\right), Q =: (x, y) =: \left(\frac{s}{d^2}, \frac{u}{d^5}\right)$, with $(S, D) = (s, d) = 1$. By Flynn's explicit formulas, we find that:

$$\kappa(P+Q) = \begin{bmatrix} (X-x)^2, (X-x)^2(X+x), (X-x)^2Xx, \\ 2a_5 + a_4(X+x) + 2a_3Xx + a_2Xx(X+x) + X^2x^2(X+x) - 2Yy \end{bmatrix}.$$

Clearing denominators, we find that:

$$\begin{aligned} &\kappa(P+Q) \\ &= \begin{bmatrix} D^2d^2(Sd^2 - sD^2)^2, (Sd^2 - sD^2)^2(Sd^2 + sD^2), Ss(Sd^2 - sD^2)^2, \\ 2a_5D^6d^6 + a_4D^4d^4(Sd^2 + sD^2), \\ 2a_3D^4d^4Ss + a_2D^2d^2Ss(Sd^2 + sD^2) + S^2s^2(Sd^2 + sD^2) - 2DdUu \end{bmatrix}. \end{aligned}$$

It therefore follows that, with these affine coordinates,

$$\begin{aligned} |\kappa(P+Q)_1| &\ll H_K(P)^3, \\ |\kappa(P+Q)_2| &\ll H_K(P)^3, \\ |\kappa(P+Q)_3| &\ll H_K(P)^3, \\ |\kappa(P+Q)_4| &\ll H(f)^5 H_K(P)^3, \end{aligned}$$

since $|S|, |s|, D^2, d^2 \ll H_K(P)^{\frac{1}{2}}$ and $|X|, |x| \ll H(f)$ by hypothesis.

Next, for $R = [k_1, k_2, k_3, k_4] \in K_f$, we have that $2R = [\delta_1(R), \delta_2(R), \delta_3(R), \delta_4(R)]$, where:

$$
\begin{aligned}
&\delta_1(k_1, \ldots, k_4) \\
&\quad := 4a_2^2a_5k_1^4 + 8a_4^2k_1^3k_2 - 32a_3a_5k_1^3k_2 - 8a_2a_5k_1^2k_2^2 + 4a_5k_2^4 - 16a_2a_5k_1^3k_3 \\
&\quad - 4a_2a_4k_1^2k_2k_3 - 16a_5k_1k_2^2k_3 + 4a_4k_2^3k_3 + 16a_5k_1^2k_3^2 - 8a_4k_1k_2k_3^2 + 4a_3k_2^2k_3^2 + 4a_2k_2k_3^3 \\
&\quad + 4a_2a_4k_1^3k_4 - 32a_5k_1^2k_2k_4 - 4a_4k_1k_2^2k_4 - 8a_4k_1^2k_3k_4 - 8a_3k_1k_2k_3k_4 - 4a_2k_1k_3^2k_4 + 8k_3^3k_4 \\
&\quad + 4a_3k_1^2k_4^2 - 4k_2k_3k_4^2 + 4k_1k_4^3, \\
&\delta_2(k_1, \ldots, k_4) \\
&\quad := a_2a_4^2k_1^4 - 4a_2a_3a_5k_1^4 + 16a_5^2k_1^4 - 4a_2^2a_5k_1^3k_2 \\
&\quad + 16a_4a_5k_1^3k_2 + 4a_4^2k_1^2k_2^2 - 4a_2a_5k_1k_2^3 - 6a_2^2a_4k_1^3k_3 + 16a_4^2k_1^3k_3 - 32a_3a_5k_1^3k_3 \\
&\quad + 16a_3a_4k_1^2k_2k_3 - 20a_2a_5k_1^2k_2k_3 - 8a_2a_4k_1k_2^2k_3 + 8a_5k_2^3k_3 + 5a_2^3k_1^2k_3^2 + 16a_3^2k_1^2k_3^2 \\
&\quad - 14a_2a_4k_1^2k_3^2 - 12a_2a_3k_1k_2k_3^2 + 32a_5k_1k_2k_3^2 + 4a_4k_2^2k_3^2 - 6a_2^2k_1k_3^3 + 16a_4k_1k_3^3 + a_2k_3^4 \\
&\quad + 4a_2a_5k_1^3k_4 + 2a_2a_4k_1^2k_2k_4 + 8a_5k_1k_2^2k_4 + 4a_4k_2^3k_4 - 12a_2a_3k_1^2k_3k_4 - 16a_5k_1^2k_3k_4 \\
&\quad - 10a_2^2k_1k_2k_3k_4 + 16a_4k_1k_2k_3k_4 + 8a_3k_2^2k_3k_4 + 16a_3k_1k_3^2k_4 + 2a_2k_2k_3^2k_4 + 4a_4k_1^2k_4^2 \\
&\quad + 8a_3k_1k_2k_4^2 + 5a_2k_2^2k_4^2 - 8a_2k_1k_3k_4^2 + 4k_3^2k_4^2 + 4k_2k_4^3, \\
&\delta_3(k_1, \ldots, k_4) \\
&\quad := 4a_3a_4^2k_1^4 - 16a_3^2a_5k_1^4 + 8a_2a_4a_5k_1^4 + 4a_2a_4^2k_1^3k_2 - 16a_2a_3a_5k_1^3k_2 - 32a_5^2k_1^3k_2 \\
&\quad - 8a_4a_5k_1^2k_2^2 + 4a_4^2k_1k_2^3 - 16a_3a_5k_1k_2^3 - 8a_2^2a_5k_1^3k_3 - 16a_4a_5k_1^3k_3 - 8a_4^2k_1^2k_2k_3 \\
&\quad - 24a_2a_5k_1k_2^2k_3 - 8a_3a_4k_1^2k_3^2 + 24a_2a_5k_1^2k_3^2 - 4a_2a_4k_1k_2k_3^2 + 12a_5k_2^2k_3^2 - 16a_5k_1k_3^3 \\
&\quad + 8a_4k_2k_3^3 + 4a_3k_3^4 + 8a_4^2k_1^3k_4 - 32a_3a_5k_1^3k_4 - 24a_2a_5k_1^2k_2k_4 - 8a_5k_2^3k_4 - 4a_2a_4k_1^2k_3k_4 \\
&\quad - 24a_5k_1k_2k_3k_4 - 4a_4k_2^2k_3k_4 - 8a_4k_1k_3^2k_4 + 4a_2k_3^3k_4 - 12a_5k_1^2k_4^2 - 4a_4k_1k_2k_4^2 + 4k_3k_4^3, \\
&\delta_4(k_1, \ldots, k_4) \\
&\quad := a_2^2a_4^2k_1^4 - 2a_4^3k_1^4 - 4a_2^2a_3a_5k_1^4 + 8a_3a_4a_5k_1^4
\end{aligned}
$$

$$
\begin{aligned}
&- 16a_2a_5^2k_1^4 - 8a_3a_4^2k_1^3k_2 + 32a_3^2a_5k_1^3k_2 - 16a_2a_4a_5k_1^3k_2 - 2a_2a_4^2k_1^2k_2^2 + 8a_2a_3a_5k_1^2k_2^2 \\
&+ 16a_5^2k_1^2k_2^2 + a_4^2k_2^4 - 4a_3a_5k_2^4 - 4a_2a_4^2k_1^3k_3 + 16a_2a_3a_5k_1^3k_3 + 32a_5^2k_1^3k_3 + 8a_2^2a_5k_1^2k_2k_3 \\
&+ 16a_4a_5k_1^2k_2k_3 - 8a_2a_5k_2^3k_3 + 12a_4^2k_1^2k_3^2 - 32a_3a_5k_1^2k_3^2 - 8a_2a_5k_1k_2k_3^2 - 2a_2a_4k_2^2k_3^2 \\
&- 4a_2a_4k_1k_3^3 - 8a_5k_2k_3^3 + a_2^2k_3^4 - 2a_4k_3^4 - 8a_2^2a_5k_1^3k_4 - 12a_4^2k_1^2k_2k_4 + 48a_3a_5k_1^2k_2k_4 \\
&+ 8a_2a_5k_1k_2^2k_4 + 16a_2a_5k_1^2k_3k_4 + 4a_2a_4k_1k_2k_3k_4 - 16a_5k_2^2k_3k_4 - 8a_5k_1k_3^2k_4 - 12a_4k_2k_3^2k_4 \\
&- 8a_3k_3^3k_4 - 2a_2a_4k_1^2k_4^2 + 8a_5k_1k_2k_4^2 - 2a_2k_3^2k_4^2 + k_4^4.
\end{aligned}
$$

In particular, $(2R)_i = \delta_i(R)$ is a homogenous polynomial (with $\mathbb{Z}$ coefficients) of degree $(12+i, 4)$, where we give the a_i degree $(i, 0)$ and the k_i degree $(i, 1)$. Hence $(2^N R)_i$ is homogeneous of degree $(4^{N+1} - 4 + i, 4^N)$ under this grading, with integral coefficients (thought of as a polynomial in $\mathbb{Z}[a_2, \ldots, a_5, k_1, \ldots, k_4]$) of size $O_N(1)$.

It therefore follows that:

$$
\begin{aligned}
|\kappa(2^N(P+Q))_i| &\ll_N \max_{\alpha_1+\alpha_2+\alpha_3+\alpha_4=4^N} H(f)^{4^{N+1}-4+i-\alpha_1-2\alpha_2-3\alpha_3-4\alpha_4} H_K(P)^{3\cdot 4^N} H(f)^{5\alpha_4} \\
&\ll H(f)^{4^{N+1}+4^N} H_K(P)^{3\cdot 4^N},
\end{aligned}
$$

which is to say that (since $H_K(R) \leq \max_i |R_i|$ when all the coordinates $R_i \in \mathbb{Z}$)

$$
\frac{1}{4^N} h_K(2^N(P+Q)) \leq 3h_K(P) + 5h(f) + O_N(1).
$$

Now since we'd like a result about the canonical height, observe that:

$$
\begin{aligned}
\hat{h}(P+Q) - \frac{1}{4^N} h_K(2^N(P+Q)) &= \sum_{n \geq N} 4^{-n-1} \left(h_K(2^{n+1}(P+Q)) - 4h_K(2^n(P+Q)) \right) \\
&= \sum_{n \geq N} 4^{-n-1} \left(\lambda_\infty(2^{n+1}(P+Q)) - 4\lambda_\infty(2^n(P+Q)) \right)
\end{aligned}
$$

$$+\sum_{p}\sum_{n\geq N}4^{-n-1}\left(\lambda_p(2^{n+1}(P+Q))-4\lambda_p(2^n(P+Q))\right),$$

where $\lambda_v(R):=\max_i\log|R_i|_v$ are the local naïve heights.

Now, by Stoll [97] (see Corollary 5.2.8), we have that

$$|\lambda_p(2R)-4\lambda_p(R)|\ll v_p(\Delta_f)\log p$$

(and the difference is 0 if $v_p(\Delta_f)\leq 1$), whence this expression simplifies to

$$\begin{aligned}&\hat{h}(P+Q)-\frac{1}{4^N}h_K(2^N(P+Q))\\&=\sum_{n\geq N}4^{-n-1}\left(\lambda_\infty(2^{n+1}(P+Q))-4\lambda_\infty(2^n(P+Q))\right)+O(4^{-N}\log|\Delta_f|).\end{aligned}$$

Moreover, certainly (by examining the explicit expressions for $2R$ for $R\in K_f(\mathbb{Q})$)

$$\lambda_\infty(2R)\leq 4\lambda_\infty(R)+O(h(f)).$$

Thus we find that the full expression is

$$\ll 4^{-N}h(f),$$

since

$$|\Delta_f|\ll H(f)^{20}.$$

Hence we have that

$$\begin{aligned}\hat{h}(P+Q)&\leq\frac{1}{4^N}h_K(2^N(P+Q))+O(4^{-N}h(f))\\&\leq 3h_K(P)+(5+O(4^{-N}))h(f)+O_N(1).\end{aligned}$$

Taking $N\asymp\log(\delta^{-1})$ (which is $\asymp 1$) gives the result. □

Now let us handle the case of points with large x-coordinate.

Lemma 5.2.5. *Let $P \neq \pm Q \in C_f(\mathbb{Q})$ with $|x(P)|, |x(Q)| \gg \delta^{-\delta^{-1}} H(f)$ and $h(P) \geq h(Q)$. Then:*

$$\hat{h}(P+Q) \leq 3h_K(P) - h(f) + O(\delta h(f)).$$

Proof. The proof follows in the same way as the previous Lemma, except now from

$$\kappa(P+Q) = \begin{bmatrix} (X-x)^2, (X-x)^2(X+x), (X-x)^2Xx, \\ 2a_5 + a_4(X+x) + 2a_3Xx + a_2Xx(X+x) + X^2x^2(X+x) - 2Yy \end{bmatrix}$$

we see that

$$|\kappa(P+Q)_i| \ll \max(|X|,|x|)^{i+1},$$

without a factor of $H(f)^5$ for $i=4$ (as we had last time), since both $|x|$ and $|X|$ are so large. Thus, in the same way as the previous Lemma, we find that

$$|\kappa(2^N(P+Q))_i| \ll_N \max_{\alpha_1+\cdots+\alpha_4=4^N} H(f)^{4^{N+1}-4+i-\alpha_1-2\alpha_2-3\alpha_3-4\alpha_4} \cdot \max(|X|,|x|)^{2\alpha_1+3\alpha_2+4\alpha_3+5\alpha_4+4^N}.$$

Now since $\max(|X|,|x|) \gg H(f)$, we find that therefore

$$|\kappa(2^N(P+Q))_i| \ll_N \max(|X|,|x|)^{5\cdot 4^N} \cdot H(f)^{-4+i}.$$

Now since $D^6d^6\kappa(P+Q)_i \in \mathbb{Z}$, we see that $D^{6\cdot 4^N}d^{6\cdot 4^N}\kappa(2^N(P+Q))_i \in \mathbb{Z}$, and so

$$\frac{1}{4^N}h_K(2^N(P+Q)) \leq 6\log D + 6\log d + 5\log\max(|X|,|x|) + O_N(1) + O(4^{-N}h(f)).$$

Since $h(P) = 2\log D + \log|X|$, we see that $h_K(P) = 4\log D + 2\log|X|$, and similarly for $h_K(Q)$. Thus we find that

$$\frac{1}{4^N}h_K(2^N(P+Q)) \leq 3h_K(P) - \log\max(|X|,|x|) + O_N(1) + O(4^{-N}h(f))$$

$$\leq 3h_K(P) - h(f) + O_N(1) + O(4^{-N}h(f)).$$

The rest of the argument (to turn this into a bound on the canonical height) is the same. □

This completes the upper bounding of $\hat{h}(P+Q)$ required for the gap principle. We will also need lower bounds on $\hat{h}(P-Q)$, which will require a detailed study of what we will call $\hat{\lambda}_\infty$.

5.2.2.2 Definition of local 'heights' and Stoll's bound.

To begin with, let us define the local canonical 'height' functions.

Definition 5.2.6. *Let*

$$\hat{\lambda}_v(\bullet) := \lambda_v(\bullet) + \sum_{n\geq 0} 4^{-n-1}\left(\lambda_v(2^{n+1}\bullet) - \lambda_v(2^n\bullet)\right),$$

the local canonical "height" *at a place v.*

(Here $\lambda_v(\bullet) := \max(\log|\bullet|_v, 0)$.) Note that this sum converges by the inequality given in (7.1) (i.e., the treatment of the Archimedean case — at finite primes Stoll's explicit bounds on the local height difference have already been mentioned) in [96]. Indeed, evidently one has the upper bound $\lambda_v(2P) - \lambda_v(P) \leq O(h(f))$, and the corresponding lower bound (and thus a two-sided bound uniform in P) is given by (7.1). For completeness, note that the uniformity in P follows from e.g. Formulas 10.2 and 10.3 in [96] — the bound only depends on the roots of f. It follows therefore that each term in the sum is bounded in absolute value by a constant depending only on f (in fact, the constant is $\ll h(f)$ as well), whence convergence.

We have written the word height in quotes because these $\hat{\lambda}_v$ are not functions on $K_f \subseteq \mathbb{P}^3$, but rather on a lift (via the canonical projection $\mathbb{A}^4 - \{0\} \twoheadrightarrow \mathbb{P}^3$) that we will call $\tilde{K}_f$, the cone on K_f in $\mathbb{A}^4$ — i.e. the subvariety of $\mathbb{A}^4$ defined by the same defining quartic. That is, these functions $\hat{\lambda}_v$ *do* change under scaling homogeneous coordinates, but they do so in a controlled fashion — indeed, so that $\sum_v \hat{\lambda}_v = \hat{h}$ remains invariant.

Now we may state Stoll's [97] bound on the local height differences at finite primes.

Theorem 5.2.7 (Stoll, [97])**.**

$$|\hat{\lambda}_p - \lambda_p| \leq \begin{cases} \frac{1}{3}v_p(2^4\Delta_f)\log p & v_p(\Delta_f) \geq 2 \\ 0 & v_p(\Delta_f) \leq 1. \end{cases}$$

Note that this will be all we use to handle the local heights at finite places. Note also that it immediately follows that:

Corollary 5.2.8.

$$\sum_p |\hat{\lambda}_p - \lambda_p| \leq \frac{1}{3}\log|\Delta_f| + O(1).$$

5.2.2.3 Analysis of $\hat{\lambda}_\infty$ and the partition at ∞.

Thus we are left with studying $\hat{\lambda}_\infty$. Following Pazuki [79], we will relate $\hat{\lambda}_\infty$ to a Riemann theta function plus the logarithm of a certain linear form (which is implicit in his normalization of Kummer coordinates), and then prove that the theta function term is harmless and may be ignored. Thus $\hat{\lambda}_\infty$ will be related to a linear form involving roots of our original quintic polynomial f, except that we will be able to choose which roots we consider. (This corresponds to translating our original point by a suitable two-torsion point.) We will then prove that there is at least one choice for which the resulting height is as desired, completing the argument.

So let

$$\tau_f =: \begin{pmatrix} \tau_1 & \tau_{12} \\ \tau_{12} & \tau_2 \end{pmatrix} \in \mathrm{Sym}^2(\mathbb{C}^2)$$

be the Riemann matrix corresponding to J_f in the Siegel fundamental domain $\mathcal{F}_2$ — that is, so that τ_f is a symmetric 2×2 complex matrix whose imaginary part is positive-definite, and so that $||\,\mathfrak{Re}\,\tau|| \ll 1$,

$$\mathfrak{Im}\,\tau_2 \geq \mathfrak{Im}\,\tau_1 \geq 2\,\mathfrak{Im}\,\tau_{12} > 0,$$

and

$$\mathfrak{Im}\,\tau_1 \geq \frac{\sqrt{3}}{2}.$$

Let

$$\Psi_f : \mathbb{C}^2/(\mathbb{Z}^2 + \tau_f\mathbb{Z}^2) \simeq J_f(\mathbb{C})$$

be a complex uniformization. Let $(\vec{a}, \vec{b})$ be a theta characteristic (i.e., simply $\vec{a}, \vec{b} \in \frac{1}{2}\mathbb{Z}^2/\mathbb{Z}^2$). Let $\theta_{\vec{a},\vec{b}} : \mathbb{C}^2 \to \mathbb{C}^2$ via:

$$\theta_{\vec{a},\vec{b}}(Z) := \sum_{\vec{n}\in\mathbb{Z}^2} e\left(\frac{1}{2}\left\langle \vec{n}+\vec{a}, \tau_f \cdot (\vec{n}+\vec{a})\right\rangle + \left\langle \vec{n}+\vec{a}, Z+\vec{b}\right\rangle\right),$$

where $e(z) := e^{2\pi i z}$. This is the usual Riemann theta function associated to the theta characteristic $(\vec{a}, \vec{b})$.

Notice that $\theta_{\vec{a},\vec{b}}$ is an even function if and only if $(2\vec{a}) \cdot (2\vec{b}) \equiv 0 \pmod 2$ — indeed, one easily sees (via the change of variable $\vec{n} \mapsto -\vec{n} - 2\vec{a}$) that $\theta_{\vec{a},\vec{b}}(-Z) = (-1)^{4\langle\vec{a},\vec{b}\rangle}\theta_{\vec{a},\vec{b}}(Z)$. As usual we call characteristics $(\vec{a}, \vec{b})$ for which $(2\vec{a}) \cdot (2\vec{b}) \equiv 0 \pmod 2$ even, and those for which this fails odd. Evidently there are exactly ten even theta characteristics, namely $\vec{a} = \vec{b} = 0$, the six more with either $\vec{a} = 0$ or $\vec{b} = 0$, the two with $\{\vec{a}, \vec{b}\} = \{(\frac{1}{2}, 0), (0, \frac{1}{2})\}$, and $\vec{a} = \vec{b} = (\frac{1}{2}, \frac{1}{2})$. The remaining six characteristics are odd.

Given a root α of f, we will write $R_\alpha := (\alpha, 0) \in C_f(\overline{\mathbb{Q}})$ for the corresponding point on C_f, whose image in J_f is two-torsion. If $\alpha = \infty$ we will interpret $R_\alpha = \infty$. Note that

$$J_f(\mathbb{C})[2] = \{0\} \cup \{[R_\alpha] - [\infty] : f(\alpha) = 0\} \cup \{[R_\alpha] + [R_\beta] - 2[\infty] : \alpha \neq \beta, f(\alpha) = f(\beta) = 0\}.$$

For ease of notation we will write

$$Q_{\alpha,\beta} := [R_\alpha] + [R_\beta] - 2[\infty] \in J_f(\overline{\mathbb{Q}}).$$

Let now

$$\Theta := \mathrm{im}\,(j) = \mathrm{im}\,(C_f \to J_f) = \{[P] - [\infty] : P \in C_f(\mathbb{C})\} \subseteq J_f(\mathbb{C}),$$

the theta divisor of C_f in J_f. Regarding this theta divisor and these R_α we have the following famous theorem of Riemann [84]:

Theorem 5.2.9 (Riemann). *Let $(\vec{a}, \vec{b})$ be a theta characteristic. Then: there exists a unique $Q \in J_f(\mathbb{C})[2]$ such that $\mathrm{div}_0(\theta_{\vec{a},\vec{b}}) = \Theta + Q$, where div_0 is the divisor of zeroes. Moreover, the odd theta characteristics are exactly those that correspond to a point in the image of C_f — i.e., either 0 or one of the form $[R_\alpha] - [\infty]$.*[2]

Here we have used the abuse of notation $\Theta + Q := \{([P] - [\infty]) + Q : P \in C_f(\mathbb{C})\}$ for the translate of the theta divisor by the two-torsion point Q. (Note in particular that, for an even theta characteristic $(\vec{a}, \vec{b})$, $\theta_{\vec{a},\vec{b}}(0) \neq 0$ since $0 \notin \Theta + Q_{\alpha,\beta}$! Indeed, otherwise we would have some P for which $h^0(P + \infty) = 2$, which would be a contradiction by Riemann-Roch.)

[2] Indeed these are the only two-torsion points in the image of C_f, since otherwise the existence of a nontrivial linear equivalence $R_\alpha + R_\beta \sim P + \infty$ would imply that there is a nonconstant meromorphic function f on $C_f(\mathbb{C})$ with $\mathrm{div}_\infty(f) \leq P + \infty$, which is contradicted by Riemann-Roch ($h^0(\infty - P) = 0$ since otherwise C_f would have a degree 1 map to $\mathbb{P}^1$, thus have genus 0).

Of course by considering cardinalities we see that the six odd characteristics are in bijection with the five roots of f plus the point at infinity, and the ten even characteristics are likewise in bijection with the ten pairs of finite roots of f.

Note also that a natural question is which characteristic $(\vec{a}, \vec{b})$ has $\operatorname{div}_0(\theta_{\vec{a},\vec{b}}) = \Theta$ — i.e., which characteristic corresponds to $0 \in J_f(\mathbb{C})$. The answer is given by a theorem of Mumford (see e.g. Theorem 5.3 in [74]).[3]

Theorem 5.2.10 (Mumford)**.**

$$\operatorname{div}_0(\theta_{\left(\frac{1}{2},\frac{1}{2}\right),\left(0,\frac{1}{2}\right)}) = \Theta.$$

Henceforth we will write

$$\chi_\infty := \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(0, \frac{1}{2}\right)\right),$$

call our characteristics χ, and, for $\chi = (\vec{a}, \vec{b})$, we will write $\chi_a := \vec{a}$ and $\chi_b := \vec{b}$. We will also write χ_ρ for the odd characteristic corresponding to the root ρ of f. We will further write $P_\rho := j(R_\rho)$, and $\tilde{P}_\rho := (0, 1, \rho, \rho^2) \in \tilde{K}_f(\mathbb{C})$ for a lift of P_ρ, regarded as a point of $K_f(\mathbb{C})$, to $\mathbb{A}^4$. Let also ℓ_ρ be the following linear form:

$$\ell_\rho(w, x, y, z) := \rho^2 w - \rho x + y.$$

Let us identify a lift of P_ρ along our uniformization $\Psi_f : \mathbb{C}^2/(\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2) \simeq J_f(\mathbb{C})$. Write, for χ a theta characteristic,

$$\tilde{\chi} := \chi_b + \tau_f \cdot \chi_a \in \mathbb{C}^2.$$

[3] This is not strictly necessary for our arguments, though it is convenient to use the form of χ_∞ below.

Lemma 5.2.11.

$$P_\rho = \Psi_f(\tilde{\chi}_\infty + \tilde{\chi}_\rho).$$

Proof. Since $K_f(\mathbb{C})[2] \cap \mathrm{div}_0(\theta_{\chi_\rho}) \cap \mathrm{div}_0(\theta_{\chi_\infty}) = \{0, P_\rho\}$, it suffices to show that both these theta functions vanish at this point, since were $\tilde{\chi}_\infty + \tilde{\chi}_\rho \in \mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2$, it would follow first that $(\chi_\infty)_a = (\chi_\rho)_a$, and then that $(\chi_\infty)_b = (\chi_\rho)_b$, a contradiction. But, by explicit calculation, for χ, η theta characteristics,

$$\theta_\chi(\tilde{\eta}) = \theta_{\chi+\eta}(0) \cdot e\left(-\frac{1}{2}\langle \eta_a, \tau_f \cdot \eta_a \rangle - \langle \eta_a, \chi_b + \eta_b \rangle\right).$$

Taking $\eta = \chi_\infty + \chi_\rho$ and $\chi = \chi_\rho$ or χ_∞ gives us the vanishing of both theta functions, as desired. $\square$

We note here that, as is e.g. easily verified case-by-case, for χ, χ', χ'' distinct odd characteristics, $\chi + \chi' + \chi''$ is even.

Next let us analyze the vanishing locus of ℓ_ρ.

Lemma 5.2.12.

$$\mathrm{div}_0(\ell_\rho) \cap K_f(\mathbb{C}) = 2\mathrm{div}_0(\theta_{\chi_\rho}) = 2(\Theta + P_\rho) \pmod{\pm 1}.$$

Proof. Notice that, from the addition law, if $P \in K_f(\mathbb{C})$ is not ∞ or P_ρ, then

$$\ell_\rho(\kappa(P + P_\rho)_1, \ldots, \kappa(P + P_\rho)_4) = 0.$$

The statement is also true for $P = \infty$ and $P = P_\rho$ by construction. This determines ℓ_ρ up to constants (and thus determines its zero divisor uniquely, since e.g. $\ell_\rho(P_\gamma) \neq 0$ for $\gamma \neq \rho$ a root of f) inside $\kappa^*\mathcal{O}_{\mathbb{P}^3}(1)(K_f)$. Now since the embedding is via $\mathcal{L}_\Theta^{\otimes 2}$, and since $\theta_{\chi_\rho}^2$, a section of this bundle, also satisfies these criteria (up to scaling by a nonzero constant), the conclusion follows from Riemann's theorem. Alternatively, one could see this by explicit computation. □

Now we have enough information to study $\hat{\lambda}_\infty$. For notational ease we will write

$$\Xi_\chi(Z) := \theta_\chi(Z) \cdot e^{-\pi \left\langle \mathfrak{Im}\, Z, (\mathfrak{Im}\, \tau_f)^{-1} \cdot \mathfrak{Im}\, Z \right\rangle}.$$

Note that $|\Xi_\chi|$ is invariant under translation by $\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2$, and thus descends to a function on $J_f(\mathbb{C}) - \mathrm{Supp}(\Theta + P_\rho)$, and even further to $K_f(\mathbb{C}) - \mathrm{Supp}(\Theta + P_\rho)$ as well, thanks to the absolute value.

Lemma 5.2.13. *Let ρ be a root of f. Let $\ell_\rho(w, x, y, z) := \rho^2 w - \rho x + y$.*[4] *Then there exists a constant $c_\rho \in \mathbb{R}$ such that the following formula holds. Let $R \in K_f(\mathbb{Q}) - \mathrm{Supp}(\Theta + P_\rho)$. Let $\tilde{R} \in \tilde{K}_f \subseteq \mathbb{A}^4$ be a lift of R under the canonical projection $\mathbb{A}^4 - \{0\} \twoheadrightarrow \mathbb{P}^3$. Let $Z \in [-\frac{1}{2}, \frac{1}{2}]^{\times 2} + \tau_f \cdot [-\frac{1}{2}, \frac{1}{2}]^{\times 2} \subseteq \mathbb{C}^2$ (a fundamental domain of $\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2$) be such that $\Psi_f(Z) \pmod{\pm 1} = R$ under the map $\mathbb{C}^2/(\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2) \simeq J_f(\mathbb{C}) \twoheadrightarrow K_f(\mathbb{C})$. Then:*

$$\hat{\lambda}_\infty(\tilde{R}) = -\log\left|\Xi_{\chi_\rho}(Z)^2\right| + \log|\ell_\rho(\tilde{R})| + c_\rho.$$

Proof. Note that $\hat{\lambda}_\infty(\tilde{R}) - \log|\ell_\rho(\tilde{R})|$ is invariant under scaling, whence it descends to a function on $K_f(\mathbb{C}) - \mathrm{Supp}(\Theta + P_\rho)$. Now note that

$$\left[\hat{\lambda}_\infty(2\tilde{R}) - \log|\ell_\rho(2\tilde{R})|\right] - 4\left[\hat{\lambda}_\infty(\tilde{R}) - \log|\ell_\rho(\tilde{R})|\right] = -\log\left|\frac{\ell_\rho(2\tilde{R})}{\ell_\rho(\tilde{R})^4}\right|$$

by definition of $\hat{\lambda}_\infty$.

[4] For the same theorem for other theta characteristics χ, one has to modify the linear form ℓ_χ — for χ_∞,

$$\ell_\infty(w, x, y, z) := w,$$

and, for $\chi_{\alpha,\beta}$,

$$\ell_{\alpha,\beta}(w, x, y, z) := \frac{2a_5 + a_4(\alpha+\beta) + 2a_3\alpha\beta + a_2\alpha\beta(\alpha+\beta) + \alpha^2\beta^2(\alpha+\beta)}{(\alpha-\beta)^2} \cdot w + \alpha\beta \cdot x - (\alpha+\beta) \cdot y + z.$$

The modification of the rest of the theorem (replacing P_ρ with $P_\alpha + P_\beta$, etc.) is straightforward.

Now since

$$\left[-\log\left|\Xi_{\chi_\rho}(2Z)^2\right|\right] - 4\left[-\log\left|\Xi_{\chi_\rho}(Z)^2\right|\right] = -\log\left|\frac{\theta_{\chi_\rho}(2Z)^2}{\theta_{\chi_\rho}(Z)^8}\right|,$$

and in both cases the right-hand sides are of the form $-\log|G|$ with G a function on K_f with $\operatorname{div}_0(G) = [2\cdot]^*(2(\Theta + P_\rho)) - 4\cdot(2(\Theta + P_\rho))$, and both $\hat{\lambda}_\infty(\tilde{R}) - \log|\ell_\rho(\tilde{R})|$ and $-\log\left|\Xi_{\chi_\rho}(Z)^2\right|$ are defined on $K_f(\mathbb{C}) - \operatorname{Supp}(\Theta + P_\rho)$, it follows that the two functions must differ by a constant.[5] Indeed, since e.g. by the Baire category theorem $K_f(\mathbb{C}) - \bigcup_{k\in\mathbb{Z}}[2^k\cdot]^{-1}(\operatorname{Supp}(\Theta + P_\rho))$ is topologically dense (in the Archimedean topology), for P any element of this set we have seen that, writing $F(P) := \hat{\lambda}_\infty(\tilde{P}) - \log|\ell_\rho(\tilde{P})| + \log\left|\Xi_{\chi_\rho}(\Psi_f^{-1}(P))^2\right|$,

$$F(2^{n+1}P) - 4F(2^nP)$$

is independent of P and n, since we have seen that this difference must be constant (since two meromorphic functions with the same divisor of zeroes must differ by a multiplicative constant).

Hence $F(P) = \frac{1}{4^n}F(2^nP) + O_F(4^{-n})$ for all P in this set. By choosing a sequence n_i such that $2^{n_i}P$ converges to a point in $K_f(\mathbb{C}) - \bigcup_{k\in\mathbb{Z}}[2^k\cdot]^{-1}(\operatorname{Supp}(\Theta + P_\rho))$ in the Archimedean topology[6], we see that $F(P) = 0$, and so, by continuity, we see that in fact $F = 0$ on the whole of $K_f(\mathbb{C}) - \operatorname{Supp}(\Theta + P_\rho)$, as desired. $\square$

[5] For a more explicit way to see this, see the computer algebra calculations included in [5]. The point is that, via Yoshitomi's [110] formulas (stated in Grant [53] and originally from H.F. Baker's 1907 book, [11]), one can express the quotient of theta functions considered above in terms of the x- and y-coordinates of the corresponding points in the corresponding divisor in the Jacobian, and now one is comparing two rational functions of x- and y-coordinates. Said compute algebra calculations do this in the case of $\chi = \chi_\infty$ — the other cases are obtained by translating by the corresponding two-torsion point.

[6] One exists since otherwise P must be a torsion point, else its multiples would be dense in a nontrivial abelian subvariety of $K_f(\mathbb{C})$, whence in particular there would be a subsequence converging to some 2^NP with N sufficiently large. But all two-power torsion points are excluded, and any other torsion points will become, after multiplying by a suitably high power of 2, odd order. But an odd order torsion point has periodic orbit under the multiplication by 2 map.

Let us apply this to the case of χ_ρ. We find then that:

Corollary 5.2.14. *Let $\alpha \neq \beta$ be roots of f. Then:*

$$c_\beta = 2\log|\theta_{\chi_\beta+\chi_\infty+\chi_\alpha}(0)| + \log\frac{|f'(\alpha)|^{\frac{1}{2}}}{|\alpha-\beta|}.$$

Proof. Apply Lemma 5.2.13 to the point P_α and note that

$$\hat{\lambda}_\infty(\tilde{P}_\alpha) = \frac{1}{4}\hat{\lambda}_\infty(2\tilde{P}_\alpha) = \frac{1}{2}\log|f'(\alpha)|,$$

since

$$(\delta_1(\tilde{P}_\alpha), \ldots, \delta_4(\tilde{P}_\alpha)) = (0,0,0,f'(\alpha)^2).$$

Finally, we use the already-used fact that, for χ, η theta characteristics,

$$\theta_\chi(\tilde{\eta}) = \theta_{\chi+\eta}(0) \cdot e\left(-\frac{1}{2}\langle \eta_a, \tau_f \cdot \eta_a\rangle - \langle \eta_a, \chi_b + \eta_b\rangle\right).$$

□

We note here that the constancy of the right-hand side in α amounts essentially to Thomae's formula for the theta constants of this curve(!).

Thus we have an expression for the canonical local height at infinity. We will only use a crude lower bound[7] for the last term, so let us get rid of it now:

Lemma 5.2.15. *Let $\alpha \neq \beta$ be roots of f. Then:*

$$c_\beta \geq 2\log|\theta_{\chi_\beta+\chi_\infty+\chi_\alpha}(0)| + \frac{1}{4}\log|\Delta_f| - 4h(f) - O(1).$$

[7] In fact $\max_{\alpha\neq\beta: f(\alpha)=f(\beta)=0} \frac{|f'(\alpha)|^{\frac{1}{2}}}{|\alpha-\beta|} \gg H(f)$. To see this, take α to be a root with maximal absolute value, and β to be the root closest to α. Since $5\alpha = \sum_{\rho\neq\alpha: f(\rho)=0} \alpha - \rho$, it follows that $5|\alpha| \leq \sum_{\rho\neq\alpha: f(\rho)=0} |\alpha-\rho|$. Since each $|\alpha - \rho| \leq 2|\alpha|$, at least three of the ρ (namely, all the other roots besides α and β) must satisfy $|\alpha-\rho| \gg |\alpha|$. Since $|\alpha| \asymp H(f)$, it follows that $\frac{|f'(\alpha)|^{\frac{1}{2}}}{|\alpha-\beta|} \gg H(f)^{\frac{3}{2}} \cdot |\alpha-\beta|^{-\frac{1}{2}}$, and the claim follows.

Proof. Since $|\rho| \ll H(f)$ for all roots ρ of f (see Lemma 5.2.27), it follows that, for all roots ρ, ρ' of f,

$$|\rho - \rho'| \ll H(f)$$

as well. (Thus e.g. $|\alpha - \beta| \ll H(f)$.)

Now since

$$|f'(\alpha)| = \prod_{\rho \neq \alpha} |\rho - \alpha|,$$

it follows that

$$|\Delta_f| = \prod_{\rho \neq \rho'} |\rho - \rho'|^2 = |f'(\alpha)|^2 \cdot \prod_{\rho, \rho' \neq \alpha, \rho \neq \rho'} |\rho - \rho'|^2 \ll |f'(\alpha)|^2 \cdot H(f)^{12}.$$

That is to say,

$$|f'(\alpha)| \gg \frac{|\Delta_f|^{\frac{1}{2}}}{H(f)^6}.$$

Combining all these with Lemma 5.2.14 gives the claim. □

Next we will show that we may ignore the contribution of the theta function. Let us first upper bound $\Xi_\chi(Z)$ — at first uniformly, and then in the special case of $Z = A + \tau_f \cdot B$ with $A, B \in [-\varepsilon, \varepsilon]^{\times 2}$ we will obtain a significantly stronger bound.

Lemma 5.2.16. *Let χ be a theta characteristic. Then: for any $Z \in \mathbb{C}^2$,*

$$|\Xi_\chi(Z)| \ll 1.$$

Proof. By double periodicity, it suffices to take $Z \in [-\frac{1}{2}, \frac{1}{2}]^{\times 2} + \tau_f \cdot [-\frac{1}{2}, \frac{1}{2}]^{\times 2}$, a fundamental domain for $\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2$ in $\mathbb{C}^2$. We will bound Ξ "trivially" — i.e., via the triangle inequality. Note that, for such Z,

$$|\Xi_\chi(Z)| \leq \sum_{\vec{n} \in \mathbb{Z}^2} e^{-\pi\left(\left\langle \vec{n}+\vec{a}, \mathfrak{Im}\,\tau_f \cdot (\vec{n}+\vec{a})\right\rangle + 2\langle \vec{n}+\vec{a}, \mathfrak{Im}\,Z\rangle + \left\langle \mathfrak{Im}\,Z, (\mathfrak{Im}\,\tau_f)^{-1} \cdot \mathfrak{Im}\,Z\right\rangle\right)}.$$

Now the term in the exponent is just

$$\begin{aligned}&\langle \vec{n}+\vec{a}, \mathfrak{Im}\,\tau_f\cdot(\vec{n}+\vec{a})\rangle + 2\,\langle \vec{n}+\vec{a}, \mathfrak{Im}\,Z\rangle + \left\langle \mathfrak{Im}\,Z, (\mathfrak{Im}\,\tau_f)^{-1}\cdot \mathfrak{Im}\,Z\right\rangle\\ &= \left\langle \vec{n}+\vec{a}+(\mathfrak{Im}\,\tau_f)^{-1}\cdot\mathfrak{Im}\,Z, \mathfrak{Im}\,\tau_f\cdot\left(\vec{n}+\vec{a}+(\mathfrak{Im}\,\tau_f)^{-1}\cdot\mathfrak{Im}\,Z\right)\right\rangle,\end{aligned}$$

i.e. we have completed the square. Now since $\mathfrak{Im}\,\tau_2 \geq \mathfrak{Im}\,\tau_1 \geq 2\,\mathfrak{Im}\,\tau_{12} > 0$ and $\mathfrak{Im}\,\tau_1 \geq \frac{\sqrt{3}}{2}$, we have that (by considering $\frac{\det \mathfrak{Im}\,\tau}{\mathrm{Tr}\,\mathfrak{Im}\,\tau}$) the eigenvalues of $\mathfrak{Im}\,\tau$ are both $\gg 1$. It follows that

$$\langle v, (\mathfrak{Im}\,\tau)\cdot v\rangle \gg ||v||_2^2$$

for any $v \in \mathbb{C}^2$. Applying this to $\vec{n}+\vec{a}+(\mathfrak{Im}\,\tau)^{-1}\cdot\mathfrak{Im}\,Z$, we find that this term is

$$\gg ||\vec{n}+\vec{a}+(\mathfrak{Im}\,\tau)^{-1}\cdot\mathfrak{Im}\,Z||_2^2.$$

Now since $Z \in [-\frac{1}{2},\frac{1}{2}]^{\times 2} + \tau_f\cdot[-\frac{1}{2},\frac{1}{2}]^{\times 2}$, it follows that $(\mathfrak{Im}\,\tau)^{-1}\cdot\mathfrak{Im}\,Z \in [-\frac{1}{2},\frac{1}{2}]$, so that

$$||\vec{n}+\vec{a}+(\mathfrak{Im}\,\tau)^{-1}\cdot\mathfrak{Im}\,Z||_2^2 \gg ||\vec{n}||_2^2 - O(1).$$

Therefore we have found that

$$|\Xi_\chi(Z)| \ll \sum_{\vec{n}\in\mathbb{Z}^2} e^{-\Omega(||\vec{n}||_2^2)} \ll 1,$$

as desired. □

Having uniformly upper bounded the size of $\Xi_\chi(Z)$, we will now determine the size of $\Xi_\chi(Z)$ when $Z = A + \tau_f\cdot B$ and $A, B \in [-\varepsilon,\varepsilon]^{\times 2}$ — in particular, we will also determine the size of the theta constants $\Xi_{\alpha,\beta}(0)$. To do this, we will simply use Proposition 7.6 of [98] (originally from [63]), though we will modify it slightly by allowing the argument of the theta function to range in a very small neighbourhood about 0.[8] By running the same analysis as is done in [98] (thus in [63] — just factor out the relevant exponential and observe that the negative-definite quadratic form in the exponent is strictly smaller away from the closest points to the origin, and then compute explicitly for those points), we find:

Proposition 5.2.17 (Cf. Proposition 7.6 in [98].). *Let $Z = A + \tau_f \cdot B \in \mathbb{C}^2$ be such that $||A||, ||B|| \ll \varepsilon$, with $\varepsilon \ll 1$ sufficiently small. Then*[9]*:*

$$\begin{aligned}
&\theta_{0,0,0,0}(Z) \asymp 1,\\
&\theta_{0,0,\frac{1}{2},0}(Z) \asymp 1,\\
&\theta_{0,0,0,\frac{1}{2}}(Z) \asymp 1,\\
&\theta_{0,0,\frac{1}{2},\frac{1}{2}}(Z) \asymp 1,\\
&|\theta_{\frac{1}{2},0,0,0}(Z)| \asymp e^{-\frac{\pi}{4}\,\mathfrak{Im}\,\tau_1 + O(\varepsilon\,\mathfrak{Im}\,\tau_2)},\\
&|\theta_{\frac{1}{2},0,0,\frac{1}{2}}(Z)| \asymp e^{-\frac{\pi}{4}\,\mathfrak{Im}\,\tau_1 + O(\varepsilon\,\mathfrak{Im}\,\tau_2)},\\
&|\theta_{0,\frac{1}{2},0,0}(Z)| \asymp e^{-\frac{\pi}{4}\,\mathfrak{Im}\,\tau_2 + O(\varepsilon\,\mathfrak{Im}\,\tau_2)},\\
&|\theta_{0,\frac{1}{2},\frac{1}{2},0}(Z)| \asymp e^{-\frac{\pi}{4}\,\mathfrak{Im}\,\tau_2 + O(\varepsilon\,\mathfrak{Im}\,\tau_2)},\\
&|\theta_{\frac{1}{2},\frac{1}{2},0,0}(Z)| \asymp e^{-\frac{\pi}{4}(\mathfrak{Im}\,\tau_1 + \mathfrak{Im}\,\tau_2 - 2\,\mathfrak{Im}\,\tau_{12}) + O(\varepsilon\,\mathfrak{Im}\,\tau_2)},\\
&|\theta_{\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}}(Z)| \asymp \left|\cos\left(\pi(Z_1+Z_2)\right)\cdot e\left(\frac{\tau_{12}}{2}\right) - \cos\left(\pi(Z_1-Z_2)\right)\right| \cdot e^{-\frac{\pi}{4}(\mathfrak{Im}\,\tau_1 + \mathfrak{Im}\,\tau_2 - 2\,\mathfrak{Im}\,\tau_{12})}.
\end{aligned}$$

Note, of course, that for any Z we also have that

$$-\log|\Xi_\chi(Z)| = -\log|\theta_\chi(Z)| + \pi\left\langle \mathfrak{Im}\,Z, (\mathfrak{Im}\,\tau_f)^{-1}\cdot\mathfrak{Im}\,Z\right\rangle \geq -\log|\theta_\chi(Z)|$$

[8] While the extension to $Z \neq 0$ sufficiently close to 0 is not necessary for our argument, if one is using the canonical height with an even characteristic (and partitioning the fundamental domain as we do in our argument) this generalized proposition is quite useful, and thus we include it.

[9] Note that we have omitted the absolute values in the first four asymptotics: here, by $C \asymp 1$ we mean that there are positive absolute constants $\kappa > \kappa' > 0$ such that $|C - \kappa| \leq \kappa'$. This technically clashes with our definition of the symbol $\asymp$, hence this explanation. Note that the condition for C implies it for $\mathfrak{Re}\,C$.

by positive-definiteness of $\mathfrak{Im}\,\tau_f$, so for the purposes of lower bounds it suffices to just deal with the asymptotics of θ_χ near 0.

Now let us analyze the constants c_ρ.

Lemma 5.2.18. *There is a root ρ_* of f such that, for all roots $\rho \neq \rho_*$ of f,*

$$c_\rho \geq \frac{1}{4}\log|\Delta_f| - 4h(f) - O(1).$$

Proof. Observe that, for each α such that $(\chi_\alpha)_a \neq \left(\frac{1}{2}, \frac{1}{2}\right)$, there is a β such that $(\chi_\beta + \chi_\infty + \chi_\alpha)_a = (0, 0)$. Indeed, $(\chi_\infty + \chi_\alpha)_a \neq (0, 0) \in \frac{1}{2}\mathbb{Z}^2/\mathbb{Z}^2$, and so choosing an odd characteristic $\chi \neq \chi_\alpha, \chi_\infty$ with $\chi_a = (\chi_\infty + \chi_\alpha)_a$, the claim follows from Lemma 5.2.15 and Proposition 5.2.17. □

We note here that this immediately implies a lower bound on the canonical local height at infinity for points very close to $[0, 0, 0, 1] \in K_f(\mathbb{C})$ in the Archimedean topology.

Lemma 5.2.19. *Let $\tilde{P} =: (\tilde{P}_1, \ldots, \tilde{P}_4) \in \tilde{K}_f(\mathbb{Q})$ be such that $|\tilde{P}_{i+1}| \gg \delta^{-\delta^{-1}} H(f)|\tilde{P}_i|$ for $1 \leq i \leq 3$. Then:*

$$\hat{\lambda}_\infty(\tilde{P}) = \lambda_\infty(\tilde{P}) + O(\delta h(f)).$$

Proof. We note that $2^N\tilde{P}$ has the same property for $N \ll \delta^{-1}$, where we define, for $\tilde{R} \in \tilde{K}_f(\mathbb{C})$,

$$2\tilde{R} := (\delta_1(\tilde{R}), \ldots, \delta_4(\tilde{R})).$$

Indeed, by the explicit formulas and the dominance of P_4 in all expressions, we find that, if $|\tilde{R}_{i+1}| \geq C \cdot \delta^{-\delta^{-1}} H(f)|\tilde{R}_i|$, then

$$(2\tilde{R})_i \asymp R_i \cdot R_4^3,$$

whence, for $1 \leq i \leq 3$,

$$|(2\tilde{R})_{i+1}| \geq \delta \cdot C \cdot \delta^{-\delta^{-1}} H(f)|(2\tilde{R})_i|.$$

It follows that, for $N \asymp \delta^{-1}$,

$$\frac{1}{4^N}\lambda_\infty(2^N\tilde{P}) = \lambda_\infty(\tilde{P}) + O(1).$$

(Here we have dropped the N in $O_N(1)$ since δ is a (very, very small) constant.) Note also that, for ρ a root of f, since $|\rho| \ll H(f)$ (see Lemma 5.2.27),

$$\ell_\rho(2^N\tilde{P}) \asymp (2^N\tilde{P})_3.$$

Now using Lemmas 5.2.13, 5.2.16, and 5.2.18 on $2^N\tilde{P}$, we find that

$$\hat{\lambda}_\infty(\tilde{P}) \geq \frac{1}{4^N}\lambda_\infty(2^N\tilde{P}) + O(4^{-N}h(f)),$$

whence the lower bound.

As for the upper bound, observe instead that $\lambda_\infty(2R) \leq 4\lambda_\infty(R) + O(h(f))$ by the explicit formulas, and then apply the Tate telescoping series to get that $\hat{\lambda}_\infty(R) \leq \lambda_\infty(R)+O(h(f))$. Now use the above argument, except with this upper bound. □

We find as corollaries the case of $\kappa(P)$ with $P \in C_f(\mathbb{Q})$ with large x-coordinate, as well as the case of $\kappa(P-Q)$ with $P \neq \pm Q \in C_f(\mathbb{Q})$ both with large x-coordinates.

Corollary 5.2.20. *Let $P \in C_f(\mathbb{Q})$ with $|x(P)| \gg \delta^{-\delta^{-1}}H(f)$. Then:*

$$\hat{\lambda}_\infty(\kappa(P)) = \lambda_\infty(\kappa(P)) - O(\delta h(f)).$$

Proof. Since $\kappa(P) = [0, 1, x(P), x(P)^2]$, the hypothesis of Lemma 5.2.19 follows. □

We also get a similarly strong statement for differences of two points with large x-coordinate.[10]

Lemma 5.2.21. *Let $P \neq \pm Q \in C_f(\mathbb{Q})$ with $|x(P)|, |x(Q)| \gg \delta^{-\delta^{-1}} H(f)$ with $y(P), y(Q) \geq 0$. Then:*

$$\hat{\lambda}_\infty(P - Q) \geq \frac{1}{2}\lambda_\infty(\kappa(P)) + \frac{1}{2}\lambda_\infty(\kappa(Q)) + h(f) - O(\delta h(f)).$$

Proof. Write $P =: (X, Y)$ and $Q =: (x, y)$. By switching, we may assume without loss of generality that $|X| \geq |x|$. Note that, by hypothesis, since $X^5 \sim f(X) = Y^2 \geq 0$, it follows that $X \geq 0$, and similarly for x. Now let us examine the image of $P - Q$ in $K_f(\mathbb{C})$ — that is, the coordinates of

$$\left(1, X + x, Xx, \frac{2a_5 + a_4(X + x) + 2a_3Xx + a_2Xx(X + x) + X^2x^2(X + x) + 2Yy}{(X - x)^2}\right).$$

The first three coordinates certainly satisfy the hypotheses of Lemma 5.2.19, so let us show that the fourth coordinate is $\gg x^2X$, which suffices since $|x| \gg \delta^{-\delta^{-1}} H(f)$. We bound the denominator by $|(X - x)^2| \ll X^2$ and note that the numerator is $X^2x^2(X + x) + 2Yy + O(H(f)^2X^2x)$. Now since $Y^2 = f(X) = X^5 \cdot (1 + O(\delta))$, it follows that $Y \geq X^{\frac{5}{2}} \cdot (1 + O(\delta))$, and similarly for y. Thus the numerator is $\gg X^3x^2$, as desired. Now Lemma 5.2.19 applies and we are done. □

It remains to analyze c_β when $\chi_\beta = \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, 0\right)\right)$.

[10] In fact we have the slightly stronger bound

$$\hat{\lambda}_\infty(P - Q) \geq \frac{1}{2}\max(\lambda_\infty(\kappa(P)), \lambda_\infty(\kappa(Q))) + \min(\lambda_\infty(\kappa(P)), \lambda_\infty(\kappa(Q))) - O(\delta h(f)),$$

which plays a role in bounding the number of *integral* points on these curves — e.g. for large points it results in a gap principle of shape $\cos\theta \leq \frac{1}{4}$, matching the Mumford gap principle for integral points observed by Helfgott and Helfgott-Venkatesh: one expects a right-hand side of $\frac{1}{g}$ for rational points, and $\frac{1}{2g}$ for integral points.

Lemma 5.2.22. *Let β be such that $\chi_\beta = \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, 0\right)\right)$. Then:*

$$c_\beta \geq -\frac{\pi}{2} \mathfrak{Im}\, \tau_1 + \frac{1}{4} \log |\Delta_f| - 4h(f) - O(1).$$

Proof. We will take α such that $\chi_\alpha = \left(\left(\frac{1}{2}, 0\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right)$, so that

$$\chi_\beta + \chi_\infty + \chi_\alpha = \left(\left(\frac{1}{2}, 0\right), (0, 0)\right).$$

It follows that, by Proposition 5.2.17,

$$|\theta_{\chi_\beta + \chi_\infty + \chi_\alpha}(0)| \asymp e^{-\frac{\pi}{4} \mathfrak{Im}\, \tau_1},$$

as desired. □

We will next show that, when $Z = A + \tau_f \cdot B$ and $||A||, ||B|| \ll \varepsilon$, the extra $\frac{\pi}{2} \mathfrak{Im}\, \tau_1$ term in the lower bound for c_β will be cancelled by an improved upper bound on θ_{χ_β}.

Lemma 5.2.23. *Let β be such that $\chi_\beta = \left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, 0\right)\right)$. Let $||A||, ||B|| \ll \varepsilon$ and $Z := A + \tau_f \cdot B$. Then:*

$$|\Xi_{\chi_\beta}(Z)| \ll e^{-(1 - O(\varepsilon)) \frac{\pi}{4} \mathfrak{Im}\, \tau_1}.$$

Proof. Observe that

$$\begin{aligned} |\Xi_{\chi_\beta}(Z)| &\leq \sum_{(n_1, n_2) \in \mathbb{Z}^2} e^{-\pi \left\langle \begin{pmatrix} n_1 + \frac{1}{2} \\ n_2 + \frac{1}{2} \end{pmatrix}, \mathfrak{Im}\, \tau_f \cdot \begin{pmatrix} n_1 + \frac{1}{2} \\ n_2 + \frac{1}{2} \end{pmatrix} \right\rangle - 2\pi \left\langle \begin{pmatrix} n_1 + \frac{1}{2} \\ n_2 + \frac{1}{2} \end{pmatrix}, \mathfrak{Im}\, \tau_f \cdot B \right\rangle - \pi \left\langle B, \mathfrak{Im}\, \tau_f \cdot B \right\rangle} \\ &= \sum_{(n_1, n_2) \in \mathbb{Z}^2} e^{-\pi \left\langle \begin{pmatrix} n_1 + \frac{1}{2} \\ n_2 + \frac{1}{2} \end{pmatrix} + B, \mathfrak{Im}\, \tau_f \cdot \left(\begin{pmatrix} n_1 + \frac{1}{2} \\ n_2 + \frac{1}{2} \end{pmatrix} + B \right) \right\rangle}. \end{aligned}$$

Now the quadratic form

$$\begin{aligned}
&\left\langle \begin{pmatrix} n_1+\frac{1}{2} \\ n_2+\frac{1}{2} \end{pmatrix} + B, \mathfrak{Im}\,\tau_f \cdot \left(\begin{pmatrix} n_1+\frac{1}{2} \\ n_2+\frac{1}{2} \end{pmatrix} + B \right) \right\rangle \\
&= (\mathfrak{Im}\,\tau_1 - \mathfrak{Im}\,\tau_{12})(n_1+\frac{1}{2}+B_1)^2 + (\mathfrak{Im}\,\tau_{12})(n_1+n_2+1+B_1+B_2)^2 \\
&\quad + (\mathfrak{Im}\,\tau_2 - \mathfrak{Im}\,\tau_{12})(n_2+\frac{1}{2}+B_2)^2. \\
&\geq \frac{\mathfrak{Im}\,\tau_1}{2}\left((n_1+\frac{1}{2}+B_1)^2 + (n_2+\frac{1}{2}+B_2)^2 \right).
\end{aligned}$$

First, note that this is always $\geq \frac{\mathfrak{Im}\,\tau_1}{4}(1-O(\varepsilon))$. Moreover, once $||\vec{n}|| \gg 1$, we find that it is

$$\geq \frac{\mathfrak{Im}\,\tau_1}{4}\left(1+\Omega(||\vec{n}||_2^2) - O(\varepsilon)\right),$$

from which the claim follows. □

We will next upper bound $\mathfrak{Im}\,\tau_1$ via a study of Igusa invariants.

Lemma 5.2.24.

$$\mathfrak{Im}\,\tau_1 \leq \frac{10}{\pi}h(f) - \frac{1}{3\pi}\log|\Delta_f| + O(1).$$

Proof. Consider the reduced Igusa invariant (note: this notation differs from Igusa's original, but follows Streng [98], at least up to normalizing (absolute) constants):

$$i_3(f) := \frac{I_4^5}{I_{10}^2},$$

where[11], writing $\rho_1, \ldots, \rho_5$ for the roots of f,

$$I_4 := \sum_{\sigma\in S_5} (\rho_{\sigma(1)}-\rho_{\sigma(2)})^2(\rho_{\sigma(2)}-\rho_{\sigma(3)})^2(\rho_{\sigma(3)}-\rho_{\sigma(1)})^2(\rho_{\sigma(4)}-\rho_{\sigma(5)})^2\cdot\rho_{\sigma(4)}^2\cdot\rho_{\sigma(5)}^2$$

[11] Here I_4 is the usual Igusa-Clebsch invariant of the binary sextic $\sum_{i=0}^5 a_i X^{5-i}Y^{i+1}$ — recall that one of the branch points of the curve is at ∞, thus the zero at $Y=0$. We have computed the invariant via $(X,Y)\mapsto(Y,X)$, which replaces the roots with their inverses, and the usual definition for sextics.

and

$$I_{10} := \Delta_f.$$

Note that, by Lemma 5.2.27, $|I_4| \ll H(f)^{12}$. It follows therefore that

$$|i_3(f)| \ll \frac{H(f)^{60}}{|\Delta_f|^2}.$$

Now, by [98] (originally in Igusa's [58] — see page 848 — and apparently already computed in Bolza's [27]), we have also that

$$i_3(f) = \frac{\left(\sum_{\chi \text{ even}} \theta_\chi(0)^8\right)^5}{\left(\prod_{\chi \text{ even}} \theta_\chi(0)^2\right)^2}.$$

Applying Proposition 5.2.17 and assuming that $\mathfrak{Im}\,\tau_1 \gg 1$ (else we are done), we find that:

$$|i_3(f)| \asymp \frac{e^{4\pi(\mathfrak{Im}\,\tau_1 + \mathfrak{Im}\,\tau_2 - \mathfrak{Im}\,\tau_{12})}}{\min(1, |\tau_{12}|)} \gg \max(e^{6\pi\,\mathfrak{Im}\,\tau_1}, e^{4\pi\,\mathfrak{Im}\,\tau_2}).$$

Combining this with

$$|i_3(f)| \ll \frac{H(f)^{60}}{|\Delta_f|^2}$$

gives the claim. □

After all this work, we have finally found that:

Corollary 5.2.25. *Let $\tilde{P} \in \tilde{K}_f(\mathbb{C})$ and ρ a root of f. Then:*

$$\hat{\lambda}_\infty(\tilde{P}) \geq \log|\ell_\rho(\tilde{P})| + \frac{5}{12}\log|\Delta_f| - 9h(f) - O(\delta h(f)).$$

Proof. We simply combine Lemmas 5.2.13, 5.2.15, 5.2.16, and 5.2.24. □

Moreover, as Lemma 5.2.23 shows, with an extra hypothesis the above bound can be significantly improved. Namely, we have the following:

Corollary 5.2.26. *Let Z be a set-theoretic section of $\mathbb{C}^2 \twoheadrightarrow \mathbb{C}^2/(\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2) \simeq J_f(\mathbb{C}) \twoheadrightarrow K_f(\mathbb{C})$. Let $\tilde{P} \in \tilde{K}_f(\mathbb{C})$ be such that $Z(P) =: A + \tau_f \cdot B$ with $||A||, ||B|| \ll \varepsilon$. Let ρ be a root of f. Then:*

$$\hat{\lambda}_\infty(\tilde{P}) \geq \log |\ell_\rho(\tilde{P})| + \frac{1}{4} \log |\Delta_f| - 4h(f) - O(\varepsilon h(f)).$$

Proof. Again combine Lemmas 5.2.13, 5.2.15, and 5.2.24, except now use Lemma 5.2.17. □

We next claim that, in fact, these lower bounds are quite good for our purposes. To show this we will need to know something about the roots of f. Crucially, we will use that the x^4 coefficient — the sum of the roots — vanishes. This will ensure that there are two roots of f that are $\gg H(f)$ away from each other (and both of size $\asymp H(f)$). First we must guarantee at least *one* root of size $H(f)$ — this exists for totally general reasons.

Lemma 5.2.27.

$$\max_{f(\rho)=0} |\rho| \asymp H(f).$$

Proof. The upper bound follows from the fact that if $|z| \geq 100H(f)$, then $|f(z)| \gg |z|^5$. The lower bound follows from the fact that, if, for all ρ such that $f(\rho) = 0$, $|\rho| \leq \frac{H(f)}{100}$, then since

$$a_i = (-1)^{5-i} \sum_{S \in \binom{\mathrm{roots}(f)}{i}} \prod_{\rho \in S} \rho,$$

we would have that

$$|a_i|^{\frac{1}{i}} < \frac{H(f)}{2}$$

for all i, a contradiction. □

Next we find the two large roots that are far away from each other.

Lemma 5.2.28. $\exists \alpha \neq \beta : f(\alpha) = f(\beta) = 0$, *and:*

$$|\alpha|, |\beta|, |\alpha - \beta| \geq \frac{H(f)}{10^{10}}.$$

Proof. Lemma 5.2.27 produces an α with $f(\alpha) = 0$ and $|\alpha| > \frac{H(f)}{100}$. Now, if for all ρ such that $f(\rho) = 0$, either $|\alpha - \rho| < \frac{|\alpha|}{100}$ or $|\rho| < \frac{|\alpha|}{100}$, then, writing $k := \#|\{\rho : f(\rho) = 0, |\alpha - \rho| < \frac{|\alpha|}{100}\}| \geq 1$, we would have that

$$0 = \sum_{f(\rho)=0} \rho = k\alpha + \sum_{f(\rho)=0, |\alpha-\rho|<\frac{|\alpha|}{100}} \rho - \alpha + \sum_{f(\rho)=0, |\rho|<\frac{|\alpha|}{100}} \rho,$$

and the first term dominates in size, a contradiction. Thus there is a root β such that $|\beta| > \frac{|\alpha|}{100}$ and $|\alpha - \beta| > \frac{|\alpha|}{100}$, as desired. □

Now take α_*, β_* as in Lemma 5.2.28.

Now, we will only ever apply our lower bound on $\hat{\lambda}_\infty$ to points of the form $\kappa(P)$ or $\kappa(P - Q)$ for $P \neq \pm Q \in C_f(\mathbb{Q})$. For these points, we note that

$$\ell_\rho(0, 1, x, x^2) = x - \rho,$$

and

$$\ell_\rho\left(1, X + x, Xx, \frac{2a_5 + a_4(X + x) + 2a_3Xx + a_2Xx(X + x) + X^2x^2(X + x) + 2Yy}{(X - x)^2}\right)$$
$$= (X - \rho)(x - \rho).$$

Thus,

$$|\ell_{\alpha_*}(0, 1, x, x^2) - \ell_{\beta_*}(0, 1, x, x^2)| = |\alpha_* - \beta_*| \gg H(f).$$

This implies (using Lemma 5.2.25) that:

Corollary 5.2.29. *Let $P \in C_f(\mathbb{Q})$. Then:*

$$\hat{\lambda}_\infty(\tilde{P}) \geq \max(\frac{1}{2}\lambda_\infty(P), h(f)) + \frac{5}{12}\log|\Delta_f| - 9h(f) - O(\delta h(f)).$$

(We have used the simple bound $|x-\rho| \gg |x|$ if $|x| \gg \delta^{-1}H(f)$ and ρ is a root of f.)

Similarly, for $\hat{\lambda}_\infty(P-Q)$, we get (using Corollary 5.2.26):

Corollary 5.2.30. *Let $P \neq \pm Q \in C_f(\mathbb{Q})$ be such that $x(P)$ and $x(Q)$ are closest to the same element of $\{\alpha_*, \beta_*\}$ and such that the coset $\Psi_f^{-1}(P-Q) \subseteq \mathbb{C}^2$ of $\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2$ contains an element $A + \tau_f \cdot B$ with $||A||, ||B|| \ll \varepsilon$. Then:*

$$\hat{\lambda}_\infty(P-Q) \geq \frac{1}{4}\log|\Delta_f| - 2h(f) - O(\varepsilon h(f)).$$

Proof. Note that, in general, $\max(|x-\alpha_*|, |x-\beta_*|) \gg H(f)$ (since $H(f) \ll |\alpha_* - \beta_*| \leq |x-\alpha_*| + |x-\beta_*|$). Without loss of generality, let us suppose that β_* is the closest of $\{\alpha_*, \beta_*\}$ to both $x(P) =: X$ and $x(Q) =: x$. Since

$$\begin{aligned}
&\ell_{\alpha_*}\left(1, X+x, Xx, \frac{2a_5 + a_4(X+x) + 2a_3Xx + a_2Xx(X+x) + X^2x^2(X+x) - 2Yy}{(X-x)^2}\right)\\
&= (X-\alpha_*)(x-\alpha_*)\\
&\gg H(f)^2,
\end{aligned}$$

the result follows from Corollary 5.2.26. □

Having completed our analysis of $\hat{\lambda}_\infty$, let us return to the postponed analysis of $\hat{h}(P-Q)$ for points without an unusually large x-coordinate.

5.2.2.4 Lower bounds on $\hat{h}(P-Q)$.

Next we will lower bound, for P and Q non-small points, $\hat{h}(P-Q)$. Since our lower bound on the height of a non-small point is so large, we will not need to do any delicate analysis for the lower bound (much like the upper bound). The only difficulty will be in guaranteeing that Corollary 5.2.30 is applicable if the points do not have big x-coordinates. To do this we will, of course, introduce a further partition of our points.

But first, let us finish off the case of points with large x-coordinate.

Lemma 5.2.31. *Let $P \neq \pm Q \in \mathrm{II}_f$ be such that $|x(P)|, |x(Q)| \gg \delta^{-\delta^{-1}} H(f)$ and $y(P), y(Q) \geq 0$. Then:*

$$\hat{h}(P-Q) \geq \frac{1}{2} h_K(P) + \frac{1}{2} h_K(Q) - \frac{17}{3} h(f) - O(\delta h(f)).$$

Proof. Write $x(P) =: X =: \frac{S}{D^2}$ and $x(Q) =: x =: \frac{s}{d^2}$, both in lowest terms, so that we are analyzing the point

$$\left(1, X+x, Xx, \frac{2a_5 + a_4(X+x) + 2a_3 Xx + a_2 Xx(X+x) + X^2x^2(X+x) + 2Yy}{(X-x)^2}\right).$$

Recall that we have already lower bounded $\hat{\lambda}_\infty(P-Q)$ — namely, Lemma 5.2.21 tells us that:

$$\hat{\lambda}_\infty(P-Q) \geq \frac{1}{2}\lambda_\infty(\kappa(P)) + \frac{1}{2}\lambda_\infty(\kappa(Q)) + h(f) - O(\delta h(f)).$$

To lower bound $\hat{\lambda}_p(P-Q)$, we will simply lower bound $\lambda_p(P-Q)$ and apply Corollary 5.2.8. To lower bound

$$\lambda_p(P-Q) := \log p \cdot \max_i(-v_p(\kappa(P-Q)_i)),$$

we first observe that, of course, since the first coordinate is 1, it follows that $\lambda_p(P-Q) \geq 0$ always. Next, if $\max(-v_p(X), -v_p(x)) > 0$, it follows (by breaking into cases based on whether or not both are positive) that

$$\max(-v_p(X+x), -v_p(Xx)) \geq 2v_p(D) + 2v_p(d).$$

Hence we have found that

$$\sum_p \lambda_p(P-Q) \geq 2\log|d| + 2\log|D|,$$

which is to say that

$$\sum_p \lambda_p(P-Q) \geq \sum_p \frac{1}{2}\lambda_p(\kappa(P)) + \frac{1}{2}\lambda_p(\kappa(Q)).$$

Applying Stoll [97] (i.e. Corollary 5.2.8), we find that

$$\sum_p \hat{\lambda}_p(P-Q) \geq \sum_p \frac{1}{2}\lambda_p(\kappa(P)) + \frac{1}{2}\lambda_p(\kappa(Q)) - \frac{1}{3}\log|\Delta_f|.$$

Combining this with the lower bound on $\hat{\lambda}_\infty(P-Q)$ and the fact that $|\Delta_f| \ll H(f)^{20}$ gives the claim. □

Write

$$\mathcal{G} := \left[-\frac{1}{2}, \frac{1}{2}\right]^{\times 2} + \tau_f \cdot \left[-\frac{1}{2}, \frac{1}{2}\right]^{\times 2}.$$

Write

$$\mathcal{G}^{(i_1,\ldots,i_4)} := \left(\left[\frac{i_1}{2N}, \frac{i_1+1}{2N}\right] \times \left[\frac{i_2}{2N}, \frac{i_2+1}{2N}\right]\right) + \tau_f \cdot \left(\left[\frac{i_3}{2N}, \frac{i_3+1}{2N}\right] \times \left[\frac{i_4}{2N}, \frac{i_4+1}{2N}\right]\right),$$

where $N \asymp \delta^{-1}$ (thus this is a partition into $O(1)$ parts, since $\delta \gg 1$). Note that

$$\mathcal{G} = \bigcup_{i_1=-N}^{N} \bigcup_{i_2=-N}^{N} \bigcup_{i_3=-N}^{N} \bigcup_{i_4=-N}^{N} \mathcal{G}^{(i_1,i_2,i_3,i_4)}.$$

Now let

$$Z : J_f(\mathbb{C}) \to \mathcal{G}$$

be a set-theoretic section (observe that the map $\mathcal{G} \to \mathbb{C}^2/(\mathbb{Z}^2 + \tau_f \cdot \mathbb{Z}^2) \simeq J_f(\mathbb{C})$ is surjective). Next let

$$\mathrm{II}_f^{(i_1,i_2,i_3,i_4)} := \mathrm{II}_f \cap Z^{-1}(\mathcal{G}^{(i_1,i_2,i_3,i_4)}).$$

(Similarly with decorations such as $\uparrow, \downarrow, \bullet$ added, and for $\mathrm{III}_f^{(i_1,\ldots,i_4)}$.) Thus if $P, Q \in \mathrm{II}_f^{(i_1,i_2,i_3,i_4)}$, we have that

$$Z(P) - Z(Q) = A + \tau_f \cdot B$$

with

$$||A||, ||B|| \ll \delta.$$

Finally, recall (via Lemma 5.2.28) that we chose two roots α_*, β_* of f such that

$$|\alpha_*|, |\beta_*|, |\alpha_* - \beta_*| \gg H(f).$$

So let

$$\mathrm{II}_f^{\alpha_*} := \{P \in \mathrm{II}_f : |x(P) - \alpha_*| \leq |x(P) - \beta_*|\}$$

and, similarly,

$$\mathrm{II}_f^{\beta_*} := \{P \in \mathrm{II}_f : |x(P) - \beta_*| \leq |x(P) - \alpha_*|\}.$$

That is, $\mathrm{II}_f^{\alpha_*}$ is the set of points of II_f whose x-coordinates are closest to α_*, and similarly for $\mathrm{II}_f^{\beta_*}$. We similarly define $\mathrm{III}_f^{\alpha_*}$ and $\mathrm{III}_f^{\beta_*}$. Finally, for $\rho \in \{\alpha_*, \beta_*\}$, define

$$\mathrm{II}_f^{(i_1,i_2,i_3,i_4),\rho} := \mathrm{II}_f^{(i_1,i_2,i_3,i_4)} \cap \mathrm{II}_f^{\rho},$$

and similarly for III_f, and all other decorations.

Having suitably refined our partition, let us now deal with points whose x-coordinate is not so large.

Lemma 5.2.32. *Let $\rho \in \{\alpha_*, \beta_*\}$. Let $P \neq \pm Q \in \mathrm{II}_f^{(i_1,i_2,i_3,i_4),\rho}$ be such that $|x(P)|, |x(Q)| \ll \delta^{-\delta^{-1}} H(f)$. Then:*

$$\hat{h}(P - Q) \geq \frac{1}{2} h_K(P) + \frac{1}{2} h_K(Q) - \frac{17}{3} h(f) - O(\delta h(f)).$$

Of course the same result holds for $\mathrm{III}_f^{(i_1,\ldots,i_4),\rho}$ as well.

Proof. The argument is precisely the same as for Lemma 5.2.31, except that for the lower bound on $\hat{\lambda}_\infty$ we use Corollary 5.2.30 (— this is where we use the partition), and we note that

$$\sum_p \frac{1}{2}\lambda_p(\kappa(P)) + \frac{1}{2}\lambda_p(\kappa(Q)) \geq \frac{1}{2} h_K(P) + \frac{1}{2} h_K(Q) - 2h(f).$$

Thus we find that:

$$\begin{aligned} \hat{h}(P - Q) &\geq \frac{1}{2} h_K(P) + \frac{1}{2} h_K(Q) + \frac{1}{4}\log|\Delta_f| - 4h(f) - \frac{1}{3}\log|\Delta_f| + O(\delta h(f)) \\ &= \frac{1}{2} h_K(P) + \frac{1}{2} h_K(Q) - \frac{17}{3} h(f) + O(\delta h(f)) \end{aligned}$$

as desired. □

Now that we have adequately (thanks to our strong lower bounds on the heights of medium points) lower bounded the canonical height of a difference, we may finally prove the claimed gap principles. To do so we will need some final preparatory work to ensure the points we consider are indeed suitably close (so that they may be seen to repulse). That is to say, we will have to ensure that their canonical heights are comparable — note that, at the moment, we may only a priori ensure that their *naïve* heights are comparable. But it turns out we have done enough to guarantee the former as well.

5.2.2.5 Partitioning based on the size of $\hat{h}$.

The following shows that we may now guarantee that h_K and $\hat{h}$ are within a multiplicative constant in our range.

Lemma 5.2.33. *Let $P \in C_f(\mathbb{Q}) - \mathrm{I}_f$. Then:*

$$\hat{h}(P) \asymp h_K(P).$$

Proof. Write

$$\hat{h}(P) = h_K(P) + \left[\hat{\lambda}_\infty(\kappa(P)) - \lambda_\infty(\kappa(P))\right] + \sum_p \left[\hat{\lambda}_p(\kappa(P)) - \lambda_p(\kappa(P))\right].$$

By Corollary 5.2.8,

$$\sum_p |\hat{\lambda}_p(\kappa(P)) - \lambda_p(\kappa(P))| \le \frac{1}{3} \log |\Delta_f| + O(1) \le \frac{20}{3} h(f) + O(1).$$

For the upper bound, note that, from the doubling formulas, evidently

$$\lambda_\infty(2R) - 4\lambda_\infty(R) \le 12h(f) + O(1),$$

so that[12]

$$\hat{\lambda}_\infty(\kappa(P)) - \lambda_\infty(\kappa(P)) \leq 4h(f) + O(1),$$

and hence

$$\hat{h}(P) \leq h_K(P) + O(h(f)),$$

which is enough since $h(P) \gg h(f)$ since P is not small.

The lower bound is a bit more difficult, and for it we will break into cases.

If $|x(P)| > \delta^{-\delta^{-1}}H(f)$, then $h(P) \geq (c_\uparrow - \delta)h(f) = \left(\frac{25}{3} - \delta\right)h(f)$, and thus $h_K(P) = 2h(P) > (\frac{50}{3} - 2\delta)h(f)$. Moreover we have seen (Lemma 5.2.20) that

$$\hat{\lambda}_\infty(\kappa(P)) \geq \lambda_\infty(\kappa(P)) - O(\delta h(f)).$$

Therefore we find that, in this case,

$$\hat{h}(P) \geq h_K(P) - \frac{20}{3}h(f) - O(\delta h(f)).$$

Since $h_K(P) > \left(\frac{50}{3} - \delta\right)h(f)$ this is enough.

If $|x(P)| < \delta^{-\delta^{-1}}H(f)$, then $h_K(P) \geq 2(c_\downarrow - \delta)h(f) = (16 - 2\delta)h(f)$. Also since $\lambda_\infty(\kappa(P)) \geq \lambda_\infty(\kappa(P)) + \frac{5}{12}\log|\Delta_f| - 9h(f) - O(\delta h(f))$ by Corollary 5.2.29, we may follow the argument in the above case to find that

$$\begin{aligned}
\hat{h}(P) &\geq h_K(P) + \frac{5}{12}\log|\Delta_f| - 10h(f) - \frac{1}{3}\log|\Delta_f| - O(\delta h(f)) \\
&= h_K(P) + \frac{1}{12}\log|\Delta_f| - 10h(f) - O(\delta h(f)) \\
&\geq h_K(P) - 10h(f) + O(\delta h(f))
\end{aligned}$$

[12] In fact it is rather easy to get that, for $R \in K_f(\mathbb{Q})$, $\hat{\lambda}_\infty(R) - \lambda_\infty(R) \leq 3h(f) + O(\varepsilon h(f))$ by following the same analysis done in Lemma 5.2.4. Indeed, one finds that $|(2^N R)_i| \ll \max_{\alpha_1+\cdots+\alpha_4=4^N} H(f)^{4^{N+1}-4+i-\alpha_1-\cdots-4\alpha_4} H_K(P)^{4^N}$ and concludes in the same way. Note that this argument gives a bound of $\hat{\lambda}_\infty(\kappa(P)) - \lambda_\infty(\kappa(P)) \leq 2h(f) + O(\varepsilon h(f))$, since $\kappa(P)_1 = 0$, and so $\alpha_1 = 0$ is forced (whence the maximum is achieved at $\alpha_2 = 4^N$, rather than $\alpha_1 = 4^N$).

which is again enough. This completes the argument. □

Thus Lemma 5.2.33 furnishes us with constants μ, ν with $\mu \asymp 1, \nu \asymp \delta^{-\delta^{-1}}$ such that, for all $P \in \mathrm{II}_f$, we have that both $\hat{h}(P) \in [\mu^{-1} h_K(P), \mu h_K(P)]$, and $h_K(P) \in [\nu^{-1} h(f), \nu h(f)]$. In order to break into points with very (multiplicatively) close heights (and canonical heights, since a priori they may still be wildly different), we will partition as follows. Note that this is precisely the situation in which we have a chance of seeing a repulsion phenomenon — our points from now on will be close in size in the Mordell-Weil lattice.

Note that

$$[\mu^{-1}, \mu] \subseteq \bigcup_{i=-O(\delta^{-1})}^{O(\delta^{-1})} [(1+\delta)^i, (1+\delta)^{i+1}],$$

and similarly for $[\nu^{-1}, \nu]$ (except with the bounds on the union changed to $O(\delta^{-2} \log \delta^{-1})$). Define the following partition of II_f into $\delta^{-O(1)}$ many pieces:

$$\mathrm{II}_f^{[i,j]} := \left\{ P \in \mathrm{II}_f \;\middle|\; \begin{array}{c} \hat{h}(P) \in [(1+\delta)^i h_K(P), (1+\delta)^{i+1} h_K(P)] \\ \text{and } h_K(P) \in [(1+\delta)^j h(f), (1+\delta)^{j+1} h(f)] \end{array} \right\},$$

and similarly with all other decorations added — e.g.,

$$\mathrm{II}_f^{\uparrow,(i_1,i_2,i_3,i_4),\rho,[i,j]} := \mathrm{II}_f^{\uparrow,(i_1,i_2,i_3,i_4),\rho} \cap \mathrm{II}_f^{[i,j]}.$$

We also define $\mathrm{III}_f^{[[i]]}$, etc. (thus also e.g. $\mathrm{III}_f^{\bullet,(i_1,\ldots,i_4),\rho,[[i]]}$) in a similar way, except without the second condition — that is, we only impose that $\hat{h}(P) \in [(1+\delta)^i h_K(P), (1+\delta)^{i+1} h_K(P)]$.

Note that, by construction, if $P, Q \in \mathrm{II}_f^{[i,j]}$, then

$$\left| \frac{h_K(P)}{h_K(Q)} - 1 \right|, \left| \frac{\hat{h}(P)}{\hat{h}(Q)} - 1 \right| \ll \delta.$$

This will allow us to replace e.g. $h_K(Q)$ by $h_K(P)$ (and the same for $\hat{h}$) in the expression for $\cos\theta_{P,Q}$ without incurring a nontrivial error. Having defined this partition, let us now finally prove the promised gap principles for $\hat{h}$.

5.2.2.6 The explicit gap principles via switching.

Let us now, finally, prove the gap principles. First, we will deal with the case of points with large x-coordinate.

Lemma 5.2.34. *Let $P \neq \pm Q \in \mathrm{II}_f^{\uparrow,(i_1,\ldots,i_4),\rho,[i,j]}$ and such that $y(P), y(Q) \geq 0$. Then:*

$$\cos\theta_{P,Q} \leq \frac{39}{59} + O(\delta) \leq 0.6334.$$

Proof. We break into cases based on whether $\hat{h}(P) \geq h_K(P) - \frac{5}{3}h(f)$ or not. If $\hat{h}(P) \geq h_K(P) - \frac{5}{3}h(f)$, then we use the formula

$$\cos\theta_{P,Q} = \frac{\hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}}.$$

Now since $\hat{h}(Q) = \hat{h}(P) \cdot (1 + O(\delta))$, we find that:

$$\cos\theta_{P,Q} \leq \frac{\hat{h}(P+Q)}{2\hat{h}(P)} - 1 - O(\delta).$$

We now apply our hypothesis for this case, namely that $\hat{h}(P) \geq h_K(P) - \frac{5}{3}h(f)$, to find that

$$\cos\theta_{P,Q} \leq \frac{\hat{h}(P+Q)}{2h_K(P) - \frac{10}{3}h(f)} - 1 - O(\delta).$$

Now we apply Lemma 5.2.5 and the fact that $h_K(Q) = h_K(P) \cdot (1 + O(\delta))$ to find that

$$\cos\theta_{P,Q} \leq \frac{1}{2} + \frac{6h(f)}{3h_K(P) - 5h(f)} + O(\delta).$$

Since $h_K(P) \geq \frac{50}{3}h(f) - O(\delta h(f))$, we find that

$$\cos\theta_{P,Q} \leq \frac{19}{30} + O(\delta),$$

finishing this case.

Thus we are left with the case of $\hat{h}(P) < h_K(P) - \frac{5}{3}h(f)$, for which we use the formula

$$\cos\theta_{P,Q} = \frac{\hat{h}(P) + \hat{h}(Q) - \hat{h}(P-Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}}.$$

Again this is simply

$$\cos\theta_{P,Q} \leq 1 - \frac{\hat{h}(P-Q)}{2\hat{h}(P)} + O(\delta),$$

and now we use our hypothesis to get that

$$\cos\theta_{P,Q} \leq 1 - \frac{\hat{h}(P-Q)}{2h_K(P) - \frac{10}{3}h(f)}.$$

But now Lemma 5.2.31 tells us that therefore

$$\cos\theta_{P,Q} \leq \frac{1}{2} + \frac{6h(f)}{3h_K(P) - 5h(f)} + O(\delta),$$

which is the same expression we got in the previous case, QED. □

Now let us prove the gap principle for points whose x-coordinate is not so large.

Lemma 5.2.35. *Let $P \neq \pm Q \in \mathrm{II}_f^{(i_1,i_2,i_3,i_4),\rho,[i,j]}$ be such that $|x(P)|, |x(Q)| \ll \delta^{-\delta^{-1}}H(f)$. Then:*

$$\cos\theta_{P,Q} \leq \frac{64}{95} + O(\delta) \leq 0.6737.$$

Proof. The proof is the same, except instead we use Lemmas 5.2.4 and 5.2.32, and we split into cases based on whether $\hat{P} \geq h_K(P) - \frac{1}{6}h(f)$ or not. In both cases we get that

$$\cos\theta_{P,Q} \leq \frac{1}{2} + \frac{33h(f)}{12h_K(P) - 2h(f)} + O(\delta).$$

Using $h_K(P) \geq 16h(f) - O(\delta h(f))$ gives the claim. □

Thus we have proved our desired gap principles.

5.2.2.7 Concluding via the sphere-packing argument.

As in Helfgott-Venkatesh [55], we will use the Kabatiansky-Levenshtein bound. We defer its statement (and a corollary) to Chapter 6 — see specifically Theorem 6.2.3 and Corollary 6.2.4.

Let us now bound the number of medium points on our curves.

Proposition 5.2.36.

$$\#|\mathrm{II}_f| \ll 1.645^{\mathrm{rank}(J_f(\mathbb{Q}))}.$$

Proof. We may assume without loss of generality that, for all $P \in \mathrm{II}_f$, $y(P) \geq 0$.[13] Since

$$\mathrm{II}_f = \bigcup_{\rho\in\{\alpha_*,\beta_*\}} \bigcup_{?\in\{\uparrow,\bullet,\downarrow\}} \bigcup_{i_1=0}^{O(\delta^{-1})} \cdots \bigcup_{i_4=0}^{O(\delta^{-1})} \bigcup_{i=-O(\delta^{-1})}^{O(\delta^{-1})} \bigcup_{j=-\delta^{-O(1)}}^{\delta^{-O(1)}} \mathrm{II}_f^{?,(i_1,\ldots,i_4),\rho,[i,j]}$$

is a partition into $\delta^{-O(1)} = O(1)$ parts, it suffices to prove this bound for each of the parts of the partition. But for each $P \neq Q \in \mathrm{II}_f^{?,(i_1,\ldots,i_4),\rho,[i,j]}$ (the bound is trivial when $Q = -P$), we have proven that $\cos\theta_{P,Q} \leq 0.6737$ (and in fact the bound is even a bit better when $? =\uparrow$). It follows then from Corollary 6.2.4 that

$$\#|\mathrm{II}_f^{?,(i_1,\ldots,i_4),\rho,[i,j]}| \ll 1.645^{\mathrm{rank}(J_f(\mathbb{Q}))},$$

as desired. □

[13] Of course this is not literally true, but the resulting bound will only be worsened by a factor of 2 to make up for this assumption.

This completes the medium point analysis.

5.2.3 Conclusion of the proof.

Thus we have completed the proof of Theorem 5.1.1. Let us combine the ingredients to conclude.

Proof of Theorem 5.1.1. By Lemmas 5.2.1 and 5.2.2, we have seen that

$$\operatorname*{Avg}_{f\in\mathcal{F}_{\mathrm{univ.}}:H(f)\leq T} \#|\mathrm{I}_f| \leq o_{T\to\infty}(1).$$

But, by Proposition 5.2.36 above and Theorem 6.1.1 of Chapter 6, we find that

$$\#|\mathrm{II}_f \cup \mathrm{III}_f| \ll 1.888^{\mathrm{rank}(J_f(\mathbb{Q}))} \leq 2^{\mathrm{rank}(J_f(\mathbb{Q}))} \leq \#|\mathrm{Sel}_2(J_f)|.$$

Finally, by Theorem 5.1.2,

$$\operatorname*{Avg}_{f\in\mathcal{F}_{\mathrm{univ.}}:H(f)\leq T} \#|\mathrm{Sel}_2(J_f)| \ll 1.$$

□

Chapter 6

$C(K)^{\mathrm{large}}$: large rational points on hyperbolic curves.

This chapter is based on Section 5.3 of my [5].

Abstract.

Let K be a number field. Let C/K be a smooth projective curve over K of genus $g > 1$. Write $J := \mathrm{Jac}\, C$. We prove that there is an explicit constant $\kappa_g \in \mathbb{R}^+$ depending only on g such that:

$$\#|\{P \in C(K) : h(P) \geq \kappa_g \cdot h(C)\}| \ll 1.872^{\mathrm{rank}\, J(K)},$$

where $h(C)$ is the height of C under its tricanonical embedding into projective space. We note that one can reduce the 1.872 to 1.311 once the genus of C is larger than an explicit absolute constant.

6.1 Introduction.

6.1.1 Main theorem.

In this chapter we prove the following theorem. The proof is a straightforward combination of the gap principles of Mumford [73] and Vojta [103] and the Kabatiansky-Levenshtein bound [60] on the size of a spherical code.

Theorem 6.1.1. *There is an explicit effectively computable function* $\kappa_\bullet : \mathbb{Z}^+_{\geq 2} \to \mathbb{R}^+$ *with the following property.*

Let K *be a number field. Let* C/K *be a smooth projective curve over* K *of genus* $g > 1$*. Let* $h(C)$ *be the height of the image of* C *under its tricanonical embedding* $C \hookrightarrow \mathbb{P}^{5g-6}$*. Write* $J := \operatorname{Jac} C$*. Then:*

$$\#|\{P \in C(K) : h(P) \geq \kappa_g \cdot h(C)\}| \ll 1.872^{\operatorname{rank} J(K)},$$

where $h(C)$ *is the height of* C *under its tricanonical embedding into projective space.*

Remark 6.1.2. *One can reduce the* 1.872 *to* 1.311 *once the genus of* C *is larger than an explicit effectively computable absolute constant.*

6.2 Inputs.

Let us state the three inputs that we will use in the proof of Theorem 6.1.1.

First, Mumford's gap principle.

Theorem 6.2.1 (Mumford's gap principle, cf. the "basic estimate" on page 1014 of Mumford's [73] and Proposition $9.4.5$ of Bombieri-Gubler's [29].)**.** *There is an explicit effectively computable function* $\kappa_{\bullet,\bullet} : \mathbb{Z}^+_{\geq 2} \times (0,1) \to \mathbb{R}^+$ *with the following property.*

Let K *be a number field. Let* C/K *be a smooth projective curve over* K *of genus* $g > 1$*. Let* $h(C)$ *be the height of* C *under its tricanonical embedding. Write* $J := \operatorname{Jac} C$*.*

Let $P_0 \in (\mathrm{Div}(C) \otimes_{\mathbb{Z}} \mathbb{Q})^1$ be a degree 1 divisor on C with rational coefficients. Embed $C \hookrightarrow J = \mathrm{Pic}^0(C)$ via $P \mapsto [P - P_0]$. Let $\frac{1}{g} < \alpha < 1$. Let $P \neq Q \in C(K)$ be such that

$$\left(1 + \kappa_{g,\alpha}^{-1}\right) \cdot h(Q) \geq h(P) \geq h(Q) \geq \kappa_{g,\alpha} \cdot h(C).$$

Then:

$$\cos \theta_{P,Q} \leq \alpha.$$

Here we have written

$$\cos \theta_{P,Q} := \frac{\hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}},$$

where $\hat{h}$ is the canonical height on the Jacobian J/K and we have suppressed the embedding $C \hookrightarrow J$ via P_0 in the notation.

Note that Proposition $9.4.5$ in Bombieri-Gubler's [29] (and certainly in Mumford's original "basic estimate" in [73]) is insufficient to prove Theorem 6.2.1, because the implicit constants (here $\kappa_{\bullet,\bullet}$) there are not made sufficiently precise. Thankfully, as noted in $11.9.3$ of Bombieri-Gubler's [29], this issue arose in Bombieri-Granville-Pintz's [28] — see their proof of Lemma 5 in [28] for a determination of the dependence on g and α of the implicit constants in Proposition $9.4.5$ of Bombieri-Gubler's [29].

Next we state Vojta's gap principle.

Theorem 6.2.2 (Vojta's gap principle, cf. Proposition 3.1 of Vojta's [103] and Theorem $11.9.1$/Remark $11.9.2$ of Bombieri-Gubler's [29].)**.** *There is an explicit effectively computable function $\kappa_{\bullet,\bullet} : \mathbb{Z}^+_{\geq 2} \times (0,1) \to \mathbb{R}^+$ with the following property.*

Let K be a number field. Let C/K be a smooth projective curve over K of genus $g > 1$. Let $h(C)$ be the height of C under its tricanonical embedding. Write $J := \mathrm{Jac}\, C$. Let $P_0 \in (\mathrm{Div}(C) \otimes_{\mathbb{Z}} \mathbb{Q})^1$ be a degree 1 divisor on C with rational coefficients. Embed

$C \hookrightarrow J = \mathrm{Pic}^0(C)$ via $P \mapsto [P - P_0]$. Let $\frac{1}{\sqrt{g}} < \alpha < 1$. Let $P \neq Q \in C(K)$ be such that

$$h(P) \geq \kappa_{g,\alpha} \cdot h(Q) \geq \kappa_{g,\alpha}^2 \cdot h(C).$$

Then:

$$\cos\theta_{P,Q} \leq \alpha.$$

Again we note that the determination of the dependence on g and α of the implicit constants in Theorem $11.9.1$ of Bombieri-Gubler's [29] (and certainly in Proposition 3.1 of Vojta's original [103]) is insufficient to prove Theorem 6.2.2, and we again use the proof of Lemma 5 Bombieri-Granville-Pintz's [28] (again, see $11.9.3$ of Bombieri-Gubler's [29]).

Finally, let us state the Kabatiansky-Levenshtein bound on the sizes of spherical codes.

Theorem 6.2.3 (Theorem 4 in Kabatiansky-Levenshtein's [60]). *Let $n \in \mathbb{Z}^+$. Let $\alpha \in (0,1)$. Let $A \subseteq S^{n-1} \subseteq \mathbb{R}^n$ be such that for all $v \neq w \in A$,*

$$\cos\theta_{v,w} \leq \alpha.$$

Then:

$$\begin{aligned}\#|A| \ll \exp\Bigg(n \cdot \Bigg[\left(\frac{1+\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\log\left(\frac{1+\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\\ -\left(\frac{1-\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\log\left(\frac{1-\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\Bigg]\Bigg).\end{aligned}$$

From this we easily derive the following corollary.

Corollary 6.2.4. *Let $n \in \mathbb{Z}^+$. Let $\alpha \in (0,1)$. Let $A \subseteq \mathbb{R}^n$ be such that for all $v \neq w \in A$,*

$$\cos\theta_{v,w} \leq \alpha.$$

Then:

$$\#|A| \ll \exp\left(n \cdot \left[\left(\frac{1+\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\log\left(\frac{1+\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\right.\right.$$
$$\left.\left. - \left(\frac{1-\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\log\left(\frac{1-\sin(\arccos(\alpha))}{2\sin(\arccos(\alpha))}\right)\right]\right).$$

Proof. Without loss of generality, $0 \notin A$ (since removing one element does not affect the bound). The map $\varphi : A \to S^{n-1}$ via $v \mapsto \frac{v}{|v|}$ is then injective, since otherwise if $v \neq w \in A$ were to map to the same point, then it would be the case that $\cos\theta_{v,w} = \left\langle \frac{v}{|v|}, \frac{w}{|w|} \right\rangle = 1 > \alpha$, a contradiction. Now just apply Theorem 6.2.3 to $\varphi(A)$. □

This completes the discussion of the inputs into the proof of Theorem 6.1.1.

6.3 Proof of Theorem 6.1.1.

We will instead prove a slightly worse bound of $\ll 1.888^{\operatorname{rank}\operatorname{Jac}(C)(K)}$ and indicate how to optimize further in a remark after the end of the argument.

Proof of Theorem 6.1.1. Let $\delta \in \mathbb{R}^+$ with $\delta \ll_g 1$ be a(n explicit effectively computable) parameter that we will optimize at the end of the argument. Let $\alpha := \frac{3}{4} + \delta^{\frac{1}{2}}$. Note that $\alpha > \frac{1}{\sqrt{2}}$.

Write

$$C(K)^{\text{large}} := \{P \in C(K) : h(P) \geq \kappa_{g,\alpha}^2 \cdot h(C)\},$$

where $\kappa_{\bullet,\bullet}$ is such that both of Theorems 6.2.1 and 6.2.2 hold.

Let $S \subseteq C(K)^{\text{large}}$ be maximal such that: for all $P \neq Q \in S$, we have that $\cos\theta_{P,Q} \leq \alpha$. Of course, by Kabatiansky-Levenshtein (Corollary 6.2.4), once $\delta \ll 1$, it follows that $\#|S| \ll 1.888^{\operatorname{rank} J(K)}$. Thus it suffices to show that $\#|C(K)^{\text{large}}| \ll \#|S|$.

Now observe that, by maximality, for all $P \in C(K)^{\mathrm{large}}$, there is a $Q \in S$ such that $\cos\theta_{P,Q} > \alpha$. Of course by Theorem 6.2.2 it follows that $\hat{h}(P) \asymp_g \hat{h}(Q)$. Thus, it follows that:

$$C(K)^{\mathrm{large}} = \bigcup_{Q\in S} \bigcup_{|k|\ll_g 1} C(K)^{\mathrm{large},(Q,k)},$$

where

$$C(K)^{\mathrm{large},(Q,k)} := \{P \in C(K)^{\mathrm{large}} : \cos\theta_{P,Q} > \alpha, \frac{\hat{h}(P)}{\hat{h}(Q)} \in [(1+\delta)^k, (1+\delta)^{k+1}]\}.$$

Since this partition is into $\ll_g \#|S|$ parts, it suffices to show that each $\#|C(K)^{\mathrm{large},(Q,k)}| \ll_g 1$.

Let, for each $P \in C(K)^{\mathrm{large},(Q,k)}$,

$$v_P := \frac{P}{|P|} - \frac{Q}{|Q|} \in J(K) \otimes_{\mathbb{Z}} \mathbb{R},$$

where $|\bullet| := \sqrt{\hat{h}(\bullet)}$. Write

$$\langle \bullet, \bullet' \rangle := \frac{\hat{h}(\bullet + \bullet') - \hat{h}(\bullet) - \hat{h}(\bullet')}{2},$$

so that

$$\cos\theta_{\bullet,\bullet'} = \frac{\langle \bullet, \bullet' \rangle}{|\bullet||\bullet'|}.$$

Now, if $C(K)^{\mathrm{large},(Q,k)} = \emptyset$, then we are done. Else let R be such that $|v_R|$ is minimal.

The claim is that, for all $R \neq P \in C(K)^{\mathrm{large},(Q,k)}$, we have that $|v_P| \gg 1$. Either $|v_R| \geq \frac{1}{2} - \delta^{\frac{1}{2}}$, in which case we are done, or not. If both $|v_R|, |v_P| < \frac{1}{2} - \delta^{\frac{1}{2}}$, then

$$|v_P - v_R| \leq |v_P| + |v_R| < 1 - 2\delta^{\frac{1}{2}}.$$

But we also have that

$$|v_P - v_R|^2 = \left|\frac{P}{|P|} - \frac{R}{|R|}\right|^2 = 2 - 2\cos\theta_{P,R}.$$

However, since both P and R are in $C(K)^{\text{large},(Q,k)}$, by Theorem 6.2.1 we find that

$$\cos\theta_{P,R} \leq \frac{1}{2} + O_g(\delta),$$

whence

$$|v_P - v_R|^2 \geq 1 - O_g(\delta),$$

a contradiction. So, since $\delta \ll_g 1$, on removing R from $C(K)^{\text{large},(Q,k)}$ we find that all remaining $|v_P| \gg 1$.

Now observe that, for $P \neq P' \in C(K)^{\text{large},(Q,k)} - \{R\}$,

$$\begin{aligned} |v_P||v_{P'}|\cos\theta_{v_P,v_{P'}} &= \langle v_P, v_{P'}\rangle \\ &= 1 - \cos\theta_{P,Q} - \cos\theta_{P',Q} + \cos\theta_{P,P'} \\ &< \cos\theta_{P,P'} - \frac{1}{2} - 2\delta^{\frac{1}{2}}. \end{aligned}$$

But since $P \neq -P'$ (else the following claim is trivial anyway) and both are in $C(K)^{\text{large},(Q,k)}$, again by Theorem 6.2.1, we find that

$$\cos\theta_{P,P'} \leq \frac{1}{2} + O_g(\delta).$$

Thus

$$|v_P||v_{P'}|\cos\theta_{v_P,v_{P'}} < -2\delta^{\frac{1}{2}} + O_g(\delta).$$

Since we have already established that $|v_P|, |v_{P'}| \gg 1$ (and of course $\delta \ll_g 1$), this gives the claim that $\cos\theta_{P,P'} \leq -\Omega_g(\delta^{\frac{1}{2}})$.

But it then follows that:

$$\begin{aligned}
0 &\leq \left| \sum_{P \in C(K)^{\mathrm{large},(Q,k)} - \{R\}} \frac{P}{|P|} \right|^2 \\
&= \left(\#|C(K)^{\mathrm{large},(Q,k)} - \{R\}|\right) + \sum_{P \neq P' \in C(K)^{\mathrm{large},(Q,k)} - \{R\}} \cos\theta_{P,P'} \\
&\leq \left(\#|C(K)^{\mathrm{large},(Q,k)} - \{R\}|\right) - \Omega_g(\delta^{\frac{1}{2}}) \cdot \left(\#|C(K)^{\mathrm{large},(Q,k)} - \{R\}|\right)^2.
\end{aligned}$$

Rearranging now gives the result. □

Remark 6.3.1. *Let us now comment on optimizing the* 1.888 *to* 1.872 *and eventually* 1.311 *for* g *(explicitly and effectively computably) sufficiently large. First, instead of choosing* $\alpha := \frac{3}{4} + \delta^{\frac{1}{2}}$*, we will choose* $\alpha \sim 0.7406$*. Moreover, instead using the actual lower bound* $|v_P| \geq \frac{1}{2} - \delta^{\frac{1}{2}}$ *that we established for* $P \in C(K)^{\mathrm{large},(Q,k)} - \{R\}$ *rather than the crude* $|v_P| \gg 1$*, we instead find that, for all* $P \neq P' \in C(K)^{\mathrm{large},(Q,k)} - \{R\}$*,*

$$\cos\theta_{v_P,v_{P'}} \leq \frac{\frac{3}{2} - 2\alpha}{|v_P||v_{P'}|} + O_g(\delta^{\frac{1}{2}}) \leq 6 - 8\alpha + O_g(\delta^{\frac{1}{2}}).$$

Now one uses Kabatiansky-Levenshtein to bound both S *(where the repulsion bound is* $\cos\theta \leq \alpha + O_g(\delta^{\frac{1}{2}})$*, which results in a bound of* $\#|S| \leq 1.85149^{\mathrm{rank}\, J(K)}$*) and* $C(K)^{\mathrm{large},(Q,k)}$ *(where the repulsion bound is* $\cos\theta \leq 6 - 8\alpha + O_g(\delta^{\frac{1}{2}})$*, which results in a bound of* $\#|C(K)^{\mathrm{large},(Q,k)}| \ll 1.01077^{\mathrm{rank}\, J(K)}$*), and multiplies these bounds together to conclude.*

If we are yet more precise and use $\frac{1}{g}$ *instead of* $\frac{1}{2}$ *in the application of Mumford's gap principle (see the applications of Theorem 6.2.1 in the proof of Theorem 6.1.1 above) we immediately obtain the following: for*

$$\max\left(\frac{1}{\sqrt{g}}, \frac{1}{4} + \frac{3}{4g}\right) < \alpha < \frac{1}{2} + \frac{1}{2g}$$

and $\delta \in \mathbb{R}^+$ *with* $\delta \ll_g 1$*, one gets*

$$\#|C(K)^{\text{large}}| \ll M\left(\operatorname{rank}(J_f(\mathbb{Q})), \alpha+\varepsilon\right) \cdot M\left(\operatorname{rank}(J_f(\mathbb{Q})), \frac{1+\frac{1}{g}-2\alpha}{\frac{1}{2}-\frac{1}{2g}}+\varepsilon\right),$$

where

$$M(n,\eta) := \max\{\#|S| : S \subseteq S^{n-1}, \forall v \neq w \in S, \cos\theta_{v,w} \leq \eta\}.$$

Taking $\alpha \sim 0.4818$ *(which is only admissible when* $g \gg 1$*), we improve the constant to* 1.311*, as claimed.*

Part II

Algorithms for rational points on hyperbolic curves.

Chapter 7

$C(K)$: Fontaine-Mazur.

This chapter is based on joint work with Brian Lawrence.

Abstract.

We give an algorithm that, on input (g, K, S), with $K/\mathbb{Q}$ a number field and S a finite set of places of K, outputs the finite set $\mathcal{A}_g(\mathfrak{o}_{K,S})$ of principally polarized g-dimensional abelian varieties defined over K and with good reduction outside S, along with an unconditional certificate of correctness of the output. Assuming the Fontaine-Mazur, Grothendieck-Serre, absolute Hodge, and Tate conjectures, we prove this algorithm always terminates in finite time.

Consequently, we give an algorithm that, on input (C, K) with C/K a smooth projective hyperbolic curve over a number field K, outputs $C(K)$, along with an unconditional certificate of correctness of the output. Again, assuming the Fontaine-Mazur, Grothendieck-Serre, absolute Hodge, and Tate conjectures, we prove this algorithm always terminates in finite time.

In other words, we give a positive solution, conditional on these standard conjectures, to an effective form of the Shafarevich Conjecture, and thus to Hilbert's tenth problem for rational points on curves over number fields[1].

[1] This is often called an effective form of the Mordell Conjecture/Faltings' Theorem, but because there are differing meanings of the phrase "effective Mordell Conjecture" we have used the terminology of Hilbert's tenth problem.

The key idea is to use a day/night procedure: given a tuple $(a_\mathfrak{p})_{\mathfrak{p}\in T}$ of putative Dirichlet coefficients at $\mathfrak{p} \in T$ of $L(s, A)$, by day one examines mod-ℓ^n Galois representations that are unramified outside S to try to rule out the tuple, and by night one searches abelian varieties with good reduction outside S to try to certify the tuple. (It is evident by purity that the set of tuples one has to consider is in an explicit finite set.) The key point is that, assuming these standard conjectures, if the tuple arises from an ℓ-adic Galois representation satisfying finitely-falsifiable hypotheses (about Hodge-Tate weights, purity, etc.), then a positive integral multiple of it actually arises from an abelian variety, via a standard construction introduced by Deligne in his proof of the Weil conjectures for K3 surfaces via Kuga-Satake.

7.1 Introduction.

7.1.1 Problem.

Let C/K be a smooth projective hyperbolic curve over a number field. It is a famous theorem of Faltings that $C(K)$ is finite. We are interested in finding $C(K)$ given C/K.

7.1.2 Main theorem.

In this chapter we prove the following theorem.

Theorem 7.1.1. *There is an algorithm that, on input (g, K, S) with $g \in \mathbb{N}$, $K/\mathbb{Q}$ a number field, and S a finite set of places of K, outputs $\mathcal{A}_g(\mathfrak{o}_{K,S})$ (along with unconditional proof of correctness of the output), the finite set*[2] *of g-dimensional abelian varieties over K with*

[2] From now on we drop the phrase "principally polarized" by using Zarhin's trick. Let us repeat that, as in the rest of this thesis, we completely ignore stack-theoretic issues, since they are irrelevant for these effectivity issues. Nonetheless let us mention that, if one prefers, one can also run the

good reduction outside S. Moreover, assuming the Fontaine-Mazur, Grothendieck-Serre, absolute Hodge, and Tate conjectures, this algorithm always terminates in finite time.

In particular, there is an algorithm that, on input (C, K) with $K/\mathbb{Q}$ a number field and C/K a smooth projective hyperbolic curve over K, outputs $C(K)$ (along with unconditional proof of correctness of the output), and, assuming the Fontaine-Mazur, Grothendieck-Serre, absolute Hodge, and Tate conjectures, always terminates in finite time.

The algorithm and its subroutines are detailed in Section 7.3 (the algorithm itself is detailed in Section 7.3.1).

We emphasize that we will prove that, if the algorithm terminates, its output is unconditionally correct — however, the proof that the algorithm always terminates will be conditional on the aforementioned conjectures.

We emphasize this point because one has much the same situation with the higher-and-higher-descents algorithm[3] for determining the rank of an elliptic curve — one runs the algorithm, and, if it terminates, one has unconditional proof that the rank is the given output. However we also emphasize that there is one completely essential difference between the two, which is that the below algorithms are not written to terminate in an at all reasonable amount of time (and it is arguable whether they can ever be optimized to do so, since e.g. they depend on enumerating finite-image Galois representations with bounded ramification).

algorithm instead using the (explicit) extension K'/K (with S' the finite set of places of K' above S) produced by taking the compositum of all extensions of K with suitably bounded degree and ramification (for example, so that "$K(A[210])$" $\subseteq K'$ for all abelian varieties in question) and then search only for points in $\mathcal{A}_{g,n}(\mathfrak{o}_{K',S'})$ (thus in a fine moduli space), i.e. abelian varieties with full level-n structure defined over K', since the field of moduli agrees with the field of definition in this case.

[3] — specifically, the day/night procedure alluded to in the paper [100] of Tate first conjecturing the finiteness of $Ш_{E/K}$ where, "by day", one searches for points of larger and larger height, and, "by night", one computes $r_{2^n}(E) := \log_2 \#|2^{n-1}\mathrm{Sel}_{2^n}(E/K)|$, say. The algorithm terminates if the "by day" points span a rank-$r_{2^n}(E)$ lattice in $E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$. To prove that the algorithm always terminates requires assuming finiteness of $Ш_{E/K}[2^\infty]$ (in this setup).

7.1.3 Difficulty.

Upon examining e.g. Vojta's proof of Faltings' Theorem it is clear where the difficulty lies. Specifically, in Vojta's argument one passes to cones in $J(K) \otimes_{\mathbb{Z}} \mathbb{R}$, where $J := \operatorname{Jac} C$ is the Jacobian of C. Inside each cone, Vojta proves that, if there is a (large) K-point of C in this cone, then all other K-points of C in the cone have height bounded explicitly in terms of this point. However one has no a priori way of knowing if there is a K-point of C in the cone, and thus no way of knowing when to terminate a search for K-points of C. The same issue arises in Roth's theorem in Diophantine approximation.

On the surface one has the same obstacle in Faltings' first proof. Specifically, in Faltings' argument the finitely many cones are replaced[4] by the finitely many tuples of integers that could arise as traces, at an explicit finite set of primes, of Galois representations of type relevant to the problem. The bound on the heights of all points in the cone in terms of the height of one point in the cone is replaced by a bound on the Faltings height of all abelian K-varieties in a K-isogeny class in terms of the Faltings height of one abelian K-variety in the K-isogeny class — a bound made explicit by Raynaud (Masser-Wüstholz later gave another explicit bound using transcendence techniques).

Thus it seems that one has the same obstacle in all known proofs of Faltings' Theorem.[5]

However this is too quick an evaluation: one has no information a priori about whether or not a given cone contains a K-point of C in Vojta's argument, but in Faltings' argument the analogous question is whether or not a given Galois repre-

[4] This is, of course, a historically backwards way of presenting the proofs.

[5] Faltings' second proof (of a much stronger theorem, to be clear) has the first obstacle, and the recent proof by Lawrence-Venkatesh has the second obstacle, along with the issue that one also has to be able to conclude in finite time that a given zero of a certain p-adic analytic function is not algebraic.

sentation arises as the ℓ-adic representation associated to an abelian variety over K. The latter is, of course, a quite well-studied sort of question.

7.1.4 Approach.

The approach taken in this chapter should then be clear.[6] One does a day/night search given a tuple, indexed by a finite set of primes, of integers as above: by night, one searches for an abelian variety over K with the given integers as Frobenius traces of its ℓ-adic representation[7], and then, by day, one searches for a counterexample to a short list of obvious necessary conditions to being a tuple of traces associated to an abelian variety. For example, the tuple must be the tuple of traces of a mod-ℓ^N representation for all N, and the latter must "seem" pure of weight 1, at least at primes much smaller than ℓ^N, etc. (See Section 7.3.1.1 for a more precise discussion.) Then the assumed conjectures combine to imply that such a process must terminate, because one proves that said conjectures imply that tuples surviving the "by day" checks forever do indeed arise[8] from abelian varieties over K.

[6] We note that we only realized such an approach might possibly work upon learning of Patrikis-Voloch-Zarhin's masterful [77]. The massive influence of that paper on this chapter should be evident.

[7] This is not quite correct, but it is accurate enough for this motivating discussion. The point is that the method only gives an abelian variety A/K whose traces are

$$\mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A)) = d \cdot a_{\mathfrak{p}}$$

for a $d \in \mathbb{Z}^+$ and all $\mathfrak{p} \in T$ (standard examples involving abelian varieties with maximal quaternionic multiplication show that one cannot take $d = 1$ in general). This, and indeed even the weaker statement that the $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-representation $V_\ell(A)$ just admit a factor ρ with $\mathrm{tr}(\rho(\mathrm{Frob}_{\mathfrak{p}})) = a_{\mathfrak{p}}$ for all $\mathfrak{p} \in T$ if such a ρ exists, is of course also fine for us, since we can determine in finite time if A has a K-isogeny factor with given tuple of Frobenius traces $(a_{\mathfrak{p}})_{\mathfrak{p} \in T}$ using Masser-Wüstholz. (One can check the latter, weaker, condition by using endomorphism estimates of Masser-Wüstholz [67] to explicitly compute $\mathrm{End}^0_K(A)$, thus $\mathrm{End}_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}(V_\ell(A))$, and thus the decomposition of $V_\ell(A)$ into irreducible $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-representations.)

[8] See the previous footnote.

7.2 Preliminary lemmas.

Before specifying the algorithms precisely, let us collect a few standard statements that will be used repeatedly in this chapter (as well as in the next chapters).

7.2.1 Faltings' Lemma.

The first lemma is the usual form of Faltings' Lemma (save perhaps the observation that it also works for the rings $\mathfrak{o}/\lambda^N$).

Lemma 7.2.1 (Faltings, see Satz 5 of his [47]). *Let $d \in \mathbb{Z}^+$. Let $K/\mathbb{Q}$ be a number field and S a finite set of places of K. Let $\mathfrak{o}$ be an order in the ring of integers of a number field and λ be a prime of $\mathfrak{o}$. Let T_λ be a finite set of primes of K that is disjoint from S and prime to $\operatorname{Nm}\lambda$ such that, for all Galois extensions L/K that are unramified outside S and of degree $[L:K] \leq \#|\mathfrak{o}/\lambda|^{2d^2}$, the map $T \to \operatorname{Gal}(L/K)/$conj. via $\mathfrak{p} \mapsto \operatorname{Frob}_\mathfrak{p}$ is surjective. Let $R := \mathfrak{o}_\lambda$ or $\mathfrak{o}/\lambda^N$ for some $N \in \mathbb{N}$. Let $\rho, \rho' : \operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \operatorname{GL}_d(R)$ be unramified outside S and such that $\operatorname{tr}(\rho(\operatorname{Frob}_\mathfrak{p})) = \operatorname{tr}(\rho'(\operatorname{Frob}_\mathfrak{p}))$ for all $\mathfrak{p} \in T_\lambda$.*

Then: $\operatorname{tr} \circ \rho = \operatorname{tr} \circ \rho'$ *on* $\operatorname{Gal}(\overline{\mathbb{Q}}/K)$.

Proof. The standard proof extends upon noting that, for $M \subseteq R^{\oplus n}$, $\#|M/\lambda| \leq \#|R/\lambda|^n$, as can be seen[9] by e.g. computing the Tor_1.

Namely, it evidently suffices to show that the R-span of $\operatorname{im}(\rho \oplus \rho' : \operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \operatorname{GL}_d(R)^{\times 2})$ inside $M_d(R)^{\times 2}$ is in fact spanned by $\bigcup_{\mathfrak{p} \in T_\lambda} (\rho(\operatorname{Frob}_\mathfrak{p}), \rho'(\operatorname{Frob}_\mathfrak{p}))$, where $\operatorname{Frob}_\mathfrak{p} \subseteq \operatorname{Gal}(\overline{\mathbb{Q}}/K)$ is the Frobenius conjugacy class of $\mathfrak{p}$. To do this one uses Nakayama to reduce mod λ, after which it follows from the hypothesis on T_λ. □

We record the following algorithmic form of the above statement, since we will cite it repeatedly in the next chapters. We emphasize that it is simple to give an

[9] Brian Lawrence points out a far more elegant proof: take the preimage $\tilde{M}$ of M under the evident surjection $\mathfrak{o}_\lambda^{\oplus n} \twoheadrightarrow R^{\oplus n}$ and then run the usual argument: $\tilde{M}$ is a free $\mathfrak{o}_\lambda$-module of rank $\leq n$, now reduce mod λ, QED.

explicit output directly (in the form of "all primes $\mathfrak{p} \notin S$ with $\mathrm{Nm}\,\mathfrak{p} \ll_{d,K,S,N} 1$", with the implicit constant computed explicitly in terms of d, K, S, N), by combining an explicit form of the Chebotarev density theorem with Minkowski's theorem, but we have not bothered.

7.2.1.1 FaltingsPrimeList(d, K, S, N):

Input: $d \in \mathbb{Z}^+$, $K/\mathbb{Q}$ a number field, S a finite set of places of K, and $N \in \mathbb{Z}^+$.
Output: T, a finite set of primes of K with the following properties:

- T is disjoint from S and prime to N.
- Let $E/\mathbb{Q}$ be a number field and $\mathfrak{q} \subseteq \mathfrak{o}_E$ with $\mathfrak{q}|(N)$ a prime of E such that $\mathrm{Nm}\,\mathfrak{q} \leq N$. Let $R := \mathfrak{o}_{E,\mathfrak{q}}$ or $\mathfrak{o}_E/\mathfrak{q}^n$ for $n \in \mathbb{Z}^+$. Let $\rho, \rho' : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_d(R)$ be unramified outside S and such that $\mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})) = \mathrm{tr}(\rho'(\mathrm{Frob}_\mathfrak{p}))$ for all $\mathfrak{p} \in T$. Then: $\mathrm{tr}(\rho(g)) = \mathrm{tr}(\rho'(g))$ for all $g \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$.

Algorithm:

1. Let T be the minimal (with respect to the absolute norm and lexicographic ordering) finite set prime to $S \cup \{(N)\}$ and such that, for all Galois extensions L/K of degree $\leq N^{10^{10} \cdot d^{10^{10}}}!$ that are unramified outside S, the map $T \to \mathrm{Gal}(L/K)/\mathrm{conj.}$ via $\mathfrak{p} \mapsto \mathrm{Frob}_\mathfrak{p}$ is surjective.
2. Output T.

7.2.1.1.1 Explanation in words.

This is simply the statement of Lemma 7.2.1 translated into algorithmic form (though we have replaced irrelevant constants by 10^{10}).

7.2.2 Masser-Wüstholz.

For reference we next state the Masser-Wüstholz bound (though we note that work of Raynaud [81], making explicit Faltings' argument, would certainly also suffice). Note that the original work of Masser-Wüstholz left one constant ineffective (because of a compactness argument using the Baily-Borel/Satake compactification of A_g to compare the Faltings and naïve heights), but work of Bost-David [31] rectified this. We will state the completely explicit bound of Gaudron-Rémond [50] instead for clarity.

Theorem 7.2.2 (Gaudron-Rémond, see Theorem 1.4 of their [50])**.** *Let $A, A'/K$ be K-isogenous abelian varieties over a number field K. Write $g := \dim A$. Then: there is an isogeny $\varphi : A \to A'$ of degree*

$$\deg \varphi \leq \left((14g)^{64g^2} \cdot [K:\mathbb{Q}] \cdot \max(h(A), \log [K:\mathbb{Q}], 1)^2\right)^{2^{10}g^3} =: \kappa(A),$$

where $h(A)$ is the Faltings height of A using Faltings' original normalization.

Consequently,

$$|h(A') - h(A)| \leq \frac{1}{2} \log \kappa(A).$$

The last line is a consequence of the usual change-in-height formula for the Faltings height under an isogeny. We will frequently use, in this and future chapters, the notation

$$\text{MasserWüstholz}(A, K) := h(A) + \frac{1}{2} \log \kappa(A)$$

for an upper bound on the height of an abelian K-variety K-isogenous to A/K.

7.2.3 Bost.

Finally, though we will not use it in this chapter (but rather in future chapters), for ease of reference let us state Bost's lower bound on the Faltings height of an abelian variety of given dimension as well.

Theorem 7.2.3 (Bost, Gaudron-Rémond (see Corollary 8.4 and then paragraph 2.3 of their [51]))**.** *Let $A/\overline{\mathbb{Q}}$ be an abelian variety. Then:*

$$h(A) \geq -\frac{\log(2\pi^2)}{2} \cdot \dim A.$$

Again, in future chapters we will use the notation

$$\text{Bost}(g) := \frac{\log(2\pi^2)}{2} \cdot g,$$

so that an abelian variety $A/\overline{\mathbb{Q}}$ satisfies $h(A) \geq -\text{Bost}(\dim A)$.

7.3 The algorithm and its subroutines.

Let us now precisely specify the algorithms alluded to in the introduction.

7.3.1 Shafarevich(g, K, S):

Input: $g, K, S : g \in \mathbb{N}, K/\mathbb{Q}$ a number field, and S a finite set of primes of K.

Output: $\mathcal{A}_g(\mathfrak{o}_{K,S})$ (and unconditional proof of correctness of the output).

Algorithm:

1. If $g \leq 1$, output $\mathcal{A}_g(\mathfrak{o}_{K,S})$ trivially (using Baker's bounds on heights of S-integral points on elliptic curves when $g = 1$).

2. Let $N := 10^{10}$. Let $\ell \geq 10^{10}$ be a prime of $\mathbb{Z}$ unramified in K and not lying under any of the primes in S.

3. Let $T :=$ output of FaltingsPrimeList$(2g, K, S, \ell)$.

4. Let $(M, \tilde{T}, Z) :=$ output of TraceTuples(g, K, S, T, ℓ, N).

5. Let, for $d \in \mathbb{Z}^+$ with $d \leq N$, $B_d :=$ output of AbelianVarieties$(d \cdot g, K, S, N)$.

6. For each $\vec{a} \in Z$: let $(d_{\vec{a}}, \tilde{B}_{\vec{a}}/K)$ be such that $d_{\vec{a}} \in \mathbb{Z}^+$, $d_{\vec{a}} \leq N$, $\tilde{B}_{\vec{a}} \in B_{d_{\vec{a}}}$ is of dimension $d_{\vec{a}} \cdot g$, and, for all $\mathfrak{p} \in \tilde{T}$, one has

   $$\mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright H^1_{\text{ét.}}((\tilde{B}_{\vec{a}})_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)) = d_{\vec{a}} \cdot a_{\mathfrak{p}}.$$

   If no such pair exists, increment $N \mapsto M + 1$ and return to Step 2. If the pair found has $d_{\vec{a}} > 1$ and NotAPower$(\tilde{B}_{\vec{a}}, K, S, g, d_{\vec{a}})$ holds, then remove $\vec{a}$ from Z. Otherwise let $\tilde{\tilde{B}}_{\vec{a}}/K$ be such that $\tilde{B}_{\vec{a}} \sim_K \tilde{\tilde{B}}_{\vec{a}}^{\times d_{\vec{a}}}$.

7. Let $C_N :=$ output of FillIsogenyClasses$(\{\tilde{\tilde{B}}_{\vec{a}} : \vec{a} \in Z\}, g, K)$.

8. Output[10] C_N.

7.3.1.1 Explanation in words.

Step 1 deals with the easy (given Baker's bounds) cases of $g \leq 1$.

Step 2 chooses an auxiliary prime ℓ and initializes a loop variable (informally, in the "day/night" language, N measures the "number of days and nights").

Step 3 chooses, via (the explicit form of) Faltings' Lemma, a finite set of primes T for which the Frobenius traces at primes in T determine the trace function of a Galois representation $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(R)$ with $R = \mathbb{Z}_\ell$ or $\mathbb{Z}/\ell^n$ for some $n \in \mathbb{Z}^+$.

Step 4 runs the "by day" procedure. Let us act like $T = \tilde{T}$ for the sake of this description (see Section 7.3.2.1).[11] Specifically, it produces the finite set of

[10] We ignore the question of finding the (potentially empty) set of principal polarizations on each of the abelian varieties in C_N, since it is irrelevant. If one prefers one can simply conclude by enumerating points $A \in \mathcal{A}_g(\mathfrak{o}_{K,S})$ of bounded height $h(A) \leq \max_{B \in C_N} h(B)$ instead.

[11] The point here is that traces at just primes in T determine abelian K-varieties of dimension g up to K-isogeny, but not a priori abelian K-varieties of dimension $N \cdot g$ up to K-isogeny. Suppose

tuples $(a_{\mathfrak{p}})_{\mathfrak{p}\in T}$ for which there is[12] a $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^N)$ satisfying $\mathrm{tr}(\rho(\mathrm{Frob}_{\mathfrak{p}})) \equiv a_{\mathfrak{p}} \pmod{\ell^N}$ for which:

- (Integrality) $a_{\mathfrak{p}} \in \mathbb{Z}$ for all $\mathfrak{p} \in T$,
- (Purity at $\mathfrak{p} \in T$)
$$|a_{\mathfrak{p}}| \leq 2g \cdot \sqrt{\mathrm{Nm}\,\mathfrak{p}}$$
for all $\mathfrak{p} \in T$,
- (ρ is unramified outside S) ρ is unramified outside S,
- (ρ has cyclotomic similitude character) the character given by postcomposing ρ with the projection $0 \to \mathrm{Sp} \to \mathrm{GSp} \to \mathbb{G}_m \to 0$ is the mod-ℓ^N reduction of the ℓ-adic cyclotomic character,
- (ρ "seems" pure at small $\mathfrak{p} \notin T$) for all primes $\mathfrak{p} \subseteq \mathfrak{o}_K$ of K with $\mathfrak{p} \notin S$ with $\mathfrak{p} \nmid (\ell)$ and $\mathrm{Nm}\,\mathfrak{p} \leq \ell^{10^{-10}\cdot N}$, there is a unique integer $a_{\mathfrak{p}} \in \mathbb{Z}$ satisfying both $|a_{\mathfrak{p}}| \leq 2g \cdot \sqrt{\mathrm{Nm}\,\mathfrak{p}}$ and $a_{\mathfrak{p}} \equiv \mathrm{tr}(\rho(\mathrm{Frob}_{\mathfrak{p}})) \pmod{\ell^N}$,
- (ρ "seems to arise" from the N-th layer of an ℓ-divisible group) and such that, for all primes $\lambda|(\ell)$ of K, the $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)$-module structure on $(\mathbb{Z}/\ell^N)^{\oplus 2g}$ given by $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}$ extends to a finite flat group scheme over $\mathfrak{o}_{K,\lambda}$.

It is crucial to us that we can get by by only being able to falsify in finite time that e.g. $(a_{\mathfrak{p}})_{\mathfrak{p}\in T}$ arises as the traces of a representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ with integral Frobenius traces that is pure of weight 1 (and similarly for the claim

the tuple $(a_{\mathfrak{p}})_{\mathfrak{p}\in T}$ arises as the traces of a representation $\rho^0 : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Q}_\ell)$. In Step 6 we are really trying to check that B has associated ℓ-adic representation $(\rho^0)^{\oplus d}$, and to do this in finite time we must check an equality of traces at primes $\mathfrak{p} \in \tilde{T}$ — just checking at $\mathfrak{p} \in T$ is inadequate. So the TraceTuples subroutine must also output the $a_{\mathfrak{p}} := \mathrm{tr}(\rho^0(\mathrm{Frob}_{\mathfrak{p}}))$ for $\mathfrak{p} \in \tilde{T}$, and it does precisely this. However this is a minor technical point, so in this description we will act like $T = \tilde{T}$.

[12] We note that, because of Faltings' Lemma with coefficient ring $\mathbb{Z}/\ell^N$, even though there may be many mod-ℓ^N representations ρ satisfying these properties, the trace function $\mathrm{tr} \circ \rho$ is still uniquely determined.

that ρ arises from an ℓ-divisible group). It is not clear how to certify that such a tuple *does* come from such a representation in finite time (after all, it is difficult to specify an ℓ-adic representation in finite form), but to show that it does *not* come from such a representation one need only examine mod-ℓ^N representations for larger and larger N. If there is no such ℓ-adic representation, then for some N there is no mod-ℓ^N representation with the fifth property above, since if for all N there is one with the fifth property above, then by an application of Kőnig's Lemma one sees that there is a compatible system of such, and thus such an ℓ-adic representation.

We note also that it is obvious that the ℓ-adic representation associated to an abelian variety A/K with good reduction outside S has all of the above properties. So at no point do we rule out tuples actually arising from abelian varieties of the desired type.

Step 5 runs the "by night" procedure. One simply enumerates abelian varieties A/K of dimension $d \cdot g$ with $d \in \mathbb{Z}^+$, $d \leq N$, and $h(A) \leq N$. It is clear that this is easily algorithmically possible (use Zarhin's trick, Masser-Wüstholz, and a comparison of the Faltings height and the naïve height under the Baily-Borel/Satake projective embedding of $A_{10^{10} \cdot d \cdot g}$). We note that the factor $d \in \mathbb{Z}^+$ arises because we will prove tuples passing the above "by day" checks for all N have a positive integral multiple that arises from an abelian variety A/K with good reduction outside S.

Let us comment on this result. The argument very much follows arguments in Section 3 of Patrikis-Voloch-Zarhin's [77]. If a tuple $(a_\mathfrak{p})_{\mathfrak{p} \in T}$ passes the "by day" checks for all N, then, again by Kőnig's Lemma, it follows that there is an ℓ-adic representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ that is unramified outside S, is pure of weight 1, satisfies $\mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})) = a_\mathfrak{p}$ for all $\mathfrak{p} \in T$, has similitude character the ℓ-adic cyclotomic character, and for which ρ arises from an ℓ-divisible group over $\mathfrak{o}_{K,\lambda}$

for all $\lambda|(\ell)$. The last two properties imply that ρ is crystalline at primes above ℓ with all Hodge-Tate weights (under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$) $\underbrace{0, \ldots, 0}_{g}, \underbrace{-1, \ldots, -1}_{g}$. Thus by the Fontaine-Mazur conjecture (the semisimplification of) $\rho^0 := \rho \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ occurs (after a Tate twist) in the ℓ-adic étale cohomology of a smooth projective variety X/K. By the Grothendieck-Serre conjecture in fact it is a summand of said cohomology group. By the Tate conjecture the projector onto ρ^0 is a $\mathbb{Q}_\ell$-linear combination of $\dim X$-dimensional correspondences in $X \times X$. However by considering the (semisimple) endomorphism algebra of the relevant motive we see that there must be a $\overline{\mathbb{Q}}$-linear combination of correspondences giving this projector, since the endomorphism algebra is already split upon tensoring up to $\overline{\mathbb{Q}}$, let alone $\mathbb{Q}_\ell$ or $\overline{\mathbb{Q}}_\ell$. So there is a number field $E/\mathbb{Q}$ for which there is an E-linear combination of correspondences projecting onto ρ^0. Our d is simply $d := [E : \mathbb{Q}]$. We find a rank $d \cdot (2g)$ motive M with $\mathbb{Q}$-coefficients (arising from the rank $2g$ motive with E-coefficients that is given by the produced projector) with ℓ-adic realization $(\rho^0)^{\oplus d}$. By taking Betti realizations we find a polarizable $\mathbb{Q}$-Hodge structure of type $\underbrace{(1,0), \ldots, (1,0)}_{d \cdot g}, \underbrace{(0,1), \ldots, (0,1)}_{d \cdot g}$, corresponding to the aforementioned Hodge-Tate weights. Riemann's classification produces an isogeny class of abelian varieties $A/\mathbb{C}$ with the corresponding $\mathbb{Q}$-Hodge structure, and the absolute Hodge conjecture allows us to descend to the isogeny class of an abelian variety $A/\overline{\mathbb{Q}}$ with this $\mathbb{Q}$-Hodge structure. By a restriction of scalars down to K and a use of the Tate conjecture for abelian varieties we eventually realize $(\rho^0)^{\oplus d}$ as the ℓ-adic representation associated to an abelian variety over K.

Step 6 enforces the aforementioned result. If the "by night" abelian varieties do not account for all the tuples that have survived all the "by day" checks thus far, then we have not run the checks with large enough N. So we increment N and return to the beginning of the loop (namely Step 4). The aforementioned result implies this check will eventually pass and we will pass to Step 7. Note that, if

$(\rho^0)^{\oplus d}$ is the ℓ-adic representation associated to a $d \cdot g$-dimensional abelian variety A/K, then ρ^0 is the ℓ-adic representation associated to a g-dimensional abelian variety B/K if and only if $A \sim_K B^{\times d}$. So, with A/K in hand, we may simply check if A/K is a d-th power up to K-isogeny using Masser-Wüstholz. This is what the NotAPower call does for us, and so we end up with only those tuples $(a_\mathfrak{p})_{\mathfrak{p} \in T}$ that genuinely arise (i.e. no multiple necessary) from abelian varieties over K.

Now we are essentially done. Step 7 is necessary because we have thus far only worked up to K-isogeny, and thus only produced at least one abelian K-variety with good reduction outside S in each K-isogeny class of such. But of course given A/K, we have a bound on the heights of all B/K with $B \sim_K A$, by Masser-Wüstholz. The FillIsogenyClasses call serves to implement this.

Thus we conclude.

7.3.2 TraceTuples(g, K, S, T, ℓ, N):

Input: g, K, S, T, ℓ, N.

Output: $(M, \tilde{T}, Z)$:

$$\tilde{T} := \bigcup_{d=1}^{N} \text{output of FaltingsPrimeList}(d \cdot (2g), K, S, \ell),$$

Z a finite set of elements of $\mathbb{Z}^{\tilde{T}}$ such that, for all $B \in \mathcal{A}_g(\mathfrak{o}_{K,S})$,

$$\mathrm{tr}(\mathrm{Frob}_\mathfrak{p} \curvearrowright H^1_{\text{ét.}}(B_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell))_{\mathfrak{p} \in \tilde{T}} \in Z$$

and $M \in \mathbb{Z}^+$ with $M \geq N$.

Algorithm:

1. Let $T_N := \bigcup_{d=1}^{N}$ output of FaltingsPrimeList$(d \cdot (2g), K, S, \ell)$.

2. Let $M := g^{10^{10}} \cdot N^{10^{10}} \cdot \max_{\mathfrak{p} \in T_N} (\mathrm{Nm}\,\mathfrak{p})^{10^{10}}$.

3. Let

$$R_M := \left\{ \begin{array}{l} \rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^M) : \\ \quad \rho \text{ unramified outside } S, \\ \quad \text{SeemspDivisible}(\rho, g, K, \ell, \lambda, M) \text{ returns True for all primes } \lambda|(\ell) \text{ of } K, \\ \quad \text{and, } \forall \mathfrak{p} \subseteq \mathfrak{o}_K \text{ with } \mathfrak{p} \notin S,\ \mathfrak{p} \nmid (\ell), \text{ and } \mathrm{Nm}\,\mathfrak{p} \leq \ell^N, \\ \quad \exists! a_\mathfrak{p} \in \mathbb{Z} \text{ with } |a_\mathfrak{p}| \leq 2g \cdot \sqrt{\mathrm{Nm}\,\mathfrak{p}} \text{ and } a_\mathfrak{p} \equiv \mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})) \pmod{\ell^M} \end{array} \right\}.$$

4. Let

$$Z := \{(a_\mathfrak{p})_{\mathfrak{p} \in T_N} \in \mathbb{Z}^{T_N} \mid \exists \rho \in R_M : \forall \mathfrak{p} \in T_N, |a_\mathfrak{p}| \leq 2g \cdot \sqrt{\mathrm{Nm}\,\mathfrak{p}}, a_\mathfrak{p} \equiv \mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})) \pmod{\ell^M}\}.$$

5. Output (M, T_N, Z).

7.3.2.1 Explanation in words.

We first essentially repeat the summary in Section 7.3.1.1: one should think of this algorithm as producing the finite set of tuples $(a_\mathfrak{p})_{\mathfrak{p} \in T}$ for which there is a $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^N)$ satisfying $\mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})) \equiv a_\mathfrak{p} \pmod{\ell^N}$ for which:

- (Integrality) $a_\mathfrak{p} \in \mathbb{Z}$ for all $\mathfrak{p} \in T$,
- (Purity at $\mathfrak{p} \in T$)

  $$|a_\mathfrak{p}| \leq 2g \cdot \sqrt{\mathrm{Nm}\,\mathfrak{p}}$$

  for all $\mathfrak{p} \in T$,
- (ρ is unramified outside S) ρ is unramified outside S,
- (ρ has cyclotomic similitude character) the character given by postcomposing ρ with the projection $0 \to \mathrm{Sp} \to \mathrm{GSp} \to \mathbb{G}_m \to 0$ is the mod-ℓ^N reduction of the ℓ-adic cyclotomic character,

- (ρ "seems" pure at small $\mathfrak{p} \notin T$) for all primes $\mathfrak{p} \subseteq \mathfrak{o}_K$ of K with $\mathfrak{p} \notin S$ with $\mathfrak{p} \nmid (\ell)$ and $\operatorname{Nm} \mathfrak{p} \leq \ell^{10^{-10} \cdot N}$, there is a unique integer $a_{\mathfrak{p}} \in \mathbb{Z}$ satisfying both $|a_{\mathfrak{p}}| \leq 2g \cdot \sqrt{\operatorname{Nm} \mathfrak{p}}$ and $a_{\mathfrak{p}} \equiv \operatorname{tr}(\rho(\operatorname{Frob}_{\mathfrak{p}})) \pmod{\ell^N}$,
- (ρ "seems to arise" from the N-th layer of an ℓ-divisible group) and such that, for all primes $\lambda | (\ell)$ of K, the $\operatorname{Gal}(\overline{\mathbb{Q}}_\ell / K_\lambda)$-module structure on $(\mathbb{Z}/\ell^N)^{\oplus 2g}$ given by $\rho|_{\operatorname{Gal}(\overline{\mathbb{Q}}_\ell / K_\lambda)}$ extends to a finite flat group scheme over $\mathfrak{o}_{K,\lambda}$.

However for the purposes of the algorithm in Section 7.3.1, we must also know the traces at $\mathfrak{p} \in T_N$, where $T \subseteq T_N$ and traces at T_N determine a G-dimensional abelian K-variety up to K-isogeny for all $G \leq N \cdot g$. The point is that, while the tuple $(a_{\mathfrak{p}})_{\mathfrak{p} \in T}$ determines a g-dimensional semisimple ℓ-adic representation that is unramified outside S uniquely (if it exists), the tuple $(d \cdot a_{\mathfrak{p}})_{\mathfrak{p} \in T}$ may a priori not determine such a $d \cdot g$-dimensional Galois representation uniquely. So we simply keep track of traces at more primes than those just in T throughout the argument by including as well the $a_{\mathfrak{p}}$ for $\mathfrak{p} \in T_N$ in the output.

7.3.3 AbelianVarieties(g, K, S, N):

Input: g, K, S, N.
Output: B, a finite subset of $\mathcal{A}_g(\mathfrak{o}_{K,S})$.
Algorithm:

1. Output the g-dimensional abelian varieties A/K with good reduction outside S and height $h(A) \leq N$.

We note that one can also, if one prefers, search through isogeny factors of Jacobians of curves of bounded genus (using the Castelnuovo bound), via Masser-Wüstholz.

7.3.3.1 Explanation in words.

This algorithm expresses the fact that the set $A_g(K)$ is recursively enumerable — in other words, it allows us to search through abelian varieties (here of larger and larger height).

7.3.4 FillIsogenyClasses(B, g, K):

Input: B, g, K : B a finite set of g-dimensional abelian varieties over K, $K/\mathbb{Q}$ a number field.

Output: C, the finite subset of $A_g(K)$ containing exactly the abelian K-varieties K-isogenous to an abelian variety in B.

Algorithm:

1. Let[13] $W := \max_{A \in B}$ MasserWüstholz(A, K).
2. Let $G := \{H \mid \exists A \in B, n \in \mathbb{Z}^+$ with $n \leq W : H \subseteq A[n]$ a K-subgroup$\}$.
3. Let $C := \{A/H : A \in B, H \in G\}$.
4. Output C.

7.3.4.1 Explanation in words.

This algorithm produces the union of all K-isogeny classes intersecting the set B of g-dimensional abelian varieties over K.

In future chapters we will simply apply Masser-Wüstholz to deduce a height bound on a K-isogenous abelian variety and then enumerate bounded-height g-dimensional abelian varieties over K. The method given here is a bit less impractical, and we have included it in this chapter just to indicate a slightly better way to proceed. (The same goes for the NotAPower subroutine below.)

[13] That is to say, W is the Masser-Wüstholz bound on the minimal degree of a K-isogeny between two g-dimensional, K-isogenous abelian varieties defined over K, say A and A', where $A \in B$.

7.3.5 SeemspDivisible$(\rho, g, K, \ell, \lambda, M)$:

Input: $\rho, g, K, \ell, \lambda, M$, with $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^M)$.

Output: True or False.

Algorithm:

1. Let $A := \mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}((\mathbb{Z}/\ell^M)^{\oplus 2g}, \overline{\mathbb{Q}}_\ell)$ and $A^\vee := \mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}(A, K_\lambda)$. Write $\mathfrak{o}_A \subseteq A, \mathfrak{o}_{A^\vee} \subseteq A^\vee$ for the rings of integers of these étale K_λ-algebras.

2. For each $R \subseteq \mathfrak{o}_A$ with $\mathrm{Hom}_{\mathfrak{o}_K}(R, \mathfrak{o}_K) \subseteq \mathfrak{o}_{A^\vee}$, check if both R and $\mathrm{Hom}_{\mathfrak{o}_K}(R, \mathfrak{o}_K)$ are closed under the ring multiplications of A and $A^\vee$, respectively. If so, return True.

3. Return False.

7.3.5.1 Explanation in words.

A is the Hopf algebra associated to the finite commutative étale group scheme G_ρ/K_λ, where $G_\rho := (\mathbb{Z}/\ell^M)^{\oplus 2g}$ as abelian groups, with Galois module structure given by ρ. We simply check that the finite commutative étale group scheme over K_λ has an integral model over $\mathfrak{o}_{K,\lambda}$. Such an integral model must be of the form $\mathrm{Spec}\, R$ with $R \subseteq \mathfrak{o}_A$ a subring, and such that (by Cartier duality) $\mathrm{Hom}_{\mathfrak{o}_K}(R, \mathfrak{o}_K) \subseteq \mathfrak{o}_{A^\vee}$ is also a subring. And indeed the latter property of a subring $R \subseteq \mathfrak{o}_A$ implies in reverse that $\mathrm{Spec}\, R$ is such an integral model. It remains to note that there are only finitely many abelian groups (let alone subrings) $\mathrm{Hom}_{\mathfrak{o}_K}(\mathfrak{o}_{A^\vee}, \mathfrak{o}_K) \subseteq R \subseteq \mathfrak{o}_A$.

We note that, for $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ and $\rho_n := \rho \bmod \ell^n$, if SeemspDivisible$(\rho_n, g, K, \ell, \lambda, n)$ returns True for all n, then the ℓ-divisible group G_ρ/K_λ whose Tate module is given by ρ (and is explicitly constructed to have n-th layer given by G_{ρ_n}/K_λ, with notation as in the previous paragraph) is therefore the generic fibre of an ℓ-divisible group $\mathcal{G}_\rho/\mathfrak{o}_{K,\lambda}$ (whose n-th layers are the subrings R_n found in the course of the SeemspDivisible$(\rho_n, g, K, \ell, \lambda, n)$ algorithm — note that we implicitly use Kőnig's Lemma here to make sure we choose a compatible system of R_n). By Tate's Hodge-Tate decomposition or else a consideration of the Dieudonné functor, it follows that ρ is crystalline at λ with Hodge-Tate weights all 0 or -1.

If moreover ρ has cyclotomic similitude character, then we find that, by autoduality, there must be exactly as many 0's as there are -1's — that is to say, in such a case, the $2g$-dimensional representation ρ is crystalline (thus de Rham) at λ with Hodge-Tate weights $\underbrace{0, \ldots, 0}_{g}, \underbrace{-1, \ldots, -1}_{g}$.

In any case we find that, if $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)} : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda) \to \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ does *not* arise from an ℓ-divisible group over $\mathfrak{o}_{K,\lambda}$ (in particular this implies that $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}$ does not arise from the ℓ-adic Tate module of an abelian variety over K_λ with good reduction at λ) for each prime $\lambda \subseteq \mathfrak{o}_K$ with $\lambda|(\ell)$, then there is a prime $\lambda \subseteq \mathfrak{o}_K$ with $\lambda|(\ell)$ and an n for which SeemspDivisible($\rho \bmod \ell^n, g, K, \ell, \lambda, n$) returns False. This is precisely what we need for the algorithm in Section 7.3.1.

7.3.6 NotAPower(B, K, S, g, d):

Input: B, K, S, g, d : B/K an abelian variety of dimension $d \cdot g$ with good reduction outside S.

Output: True or False.

Algorithm:

1. Let[14] $W :=$ MasserWüstholz(B, K).

2. Let $C := \{B/H : n \in \mathbb{Z}^+, n \leq W, H \subseteq B[n]$ a K-subgroup$\}$.

3. Check[15] if there is a $\tilde{B}/K$ with $\tilde{B}^{\times d} \in C$.

[14] Again, this is to say that, for all C/K K-isogenous to B/K, there is a K-isogeny $B \sim_K C$ of degree $\leq W$.

[15] For example by computing the maximal height of an element of C and then enumerating bounded height abelian varieties over K.

4. If so, return False. Else, return True.

7.3.6.1 Explanation in words.

This algorithm simply checks if B/K is a d-th power up to K-isogeny.

7.4 Proof of Theorem 7.1.1.

In this section we prove the following theorem.

Theorem 7.4.1.

- *(Proof of correctness assuming termination.) If Shafarevich(g, K, S) terminates, the output is unconditionally correct.*
- *(Proof of termination.) Assume the Fontaine-Mazur, Grothendieck-Serre, Tate, and absolute Hodge conjectures. Then: Shafarevich(g, K, S) always terminates.*

Evidently this implies Theorem 7.1.1.

7.4.1 Proof of correctness.

Proof of correctness assuming termination. Let us first prove the first claim in Theorem 7.4.1. Let Y be the output of Shafarevich(g, K, S). The claim is that $Y = \mathcal{A}_g(\mathfrak{o}_{K,S})$. That $Y \subseteq \mathcal{A}_g(\mathfrak{o}_{K,S})$ is evident: for all N, all $B \in B_d$ of the form $B \sim_K \tilde{B}^{\times d}$ have B/K of good reduction outside S, and thus $\tilde{B}/K$ of good reduction outside S, so that $C_N \subseteq \mathcal{A}_g(\mathfrak{o}_{K,S})$ for all N. So let us show the reverse inclusion.

Let $A \in \mathcal{A}_g(\mathfrak{o}_{K,S})$. Let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ be its ℓ-adic representation. Then, using the notation of the TraceTuples subroutine, evidently $\rho \bmod \ell^M \in R_M$ for all M. Hence, in Step 4, we find that $(\mathrm{tr}(\mathrm{Frob}_\mathfrak{p} \curvearrowright H^1_{\text{ét.}}(A_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)))_{\mathfrak{p} \in \tilde{T}} \in Z$.

Then, because of the check in Step 6, there is a $d \in \mathbb{Z}^+$ and an $\tilde{A}/K \in B_d$ with $\tilde{A} \sim_K A^{\times d}$. Hence $A \in C_N$, whence $A \in Y$, as desired.

Thus the first claim follows. □

7.4.2 Proof of termination.

Proof of termination. Now let us turn to the second claim of Theorem 7.4.1. Assume the Fontaine-Mazur, Grothendieck-Serre, Tate, and absolute Hodge conjectures. We need to show that, for N sufficiently large, the check performed in Step 6 passes. Note: for $N' \geq N$, writing $\tilde{T}_n$ and Z_n for the values of $\tilde{T}$ and Z computed in the loop with parameter $N = n$ and writing $\pi_n : \mathbb{Z}^{\tilde{T}_n} \to \mathbb{Z}^T$ for the projection $(a_\mathfrak{p})_{\mathfrak{p}\in\tilde{T}_n} \mapsto (a_\mathfrak{p})_{\mathfrak{p}\in T}$, we have that:

$$\pi_{N'}(Z_{N'}) \subseteq \pi_N(Z_N) \subseteq \pi_{10^{10}}(Z_{10^{10}}),$$

which is finite. Thus we need to show: for all

$$(a_\mathfrak{p})_{\mathfrak{p}\in T} \in \bigcap_{N\geq 10^{10}} \pi_N(Z_N),$$

there is a $d \in \mathbb{Z}^+$, an $\vec{a} \in Z_{10^{10}\cdot d}$, and a $d \cdot g$-dimensional abelian variety B/K with good reduction outside S such that

$$d \cdot a_\mathfrak{p} = \operatorname{tr}(\operatorname{Frob}_\mathfrak{p} \curvearrowright H^1_{\text{ét.}}(B_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell))$$

for all $\mathfrak{p} \in \tilde{T}_{10^{10}\cdot d}$. Now, $(a_\mathfrak{p})_{\mathfrak{p}\in T} \in \bigcap_{N\geq 10^{10}} \pi_N(Z_N)$ implies that, for N sufficiently large (in terms of $(a_\mathfrak{p})_{\mathfrak{p}\in T}$), there is an $\vec{a} \in Z_{10^{10}\cdot N}$ and a

$$\rho_N : \operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \operatorname{GSp}_{2g}(\mathbb{Z}/\ell^N)$$

that is unramified outside S, has similitude character the mod-ℓ^N reduction of the ℓ-adic cyclotomic character, is such that SeemspDivisible$(\rho_N, g, K, \ell, \lambda, N)$ returns

True for all $\lambda | (\ell)$, and is such that $\mathrm{tr}(\rho(\mathrm{Frob}_{\mathfrak{p}})) \equiv a_{\mathfrak{p}} \pmod{\ell^N}$ for all $\mathfrak{p} \in \tilde{T}_{10^{10} \cdot N}$. Applying Kőnig's Lemma (and the fact that there are only finitely many representations $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^N)$ that are unramified outside S in the first place), we find that we may assume without loss of generality that the ρ_N form a compatible system. Let

$$\begin{aligned} \tilde{\rho} &:= \varprojlim \rho_N, \\ \tilde{\rho}^0 &:= \tilde{\rho} \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell, \\ \rho^0 &:= (\tilde{\rho}^0)^{s.s.}, \end{aligned}$$

where the superscript "$s.s.$" indicates that we have semisimplified.

Evidently $\tilde{\rho}$, $\tilde{\rho}^0$, and ρ^0 each have similitude character the ℓ-adic cyclotomic character.

We claim that ρ^0 is unramified outside S and de Rham at primes above ℓ with Hodge-Tate weights (under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$) $\underbrace{0, \ldots, 0}_{g}, \underbrace{-1, \ldots, -1}_{g}$. The first part is evident since, for $\mathfrak{p} \notin S$ with $\mathfrak{p} \nmid (\ell)$, and for all N, ρ_N is trivial on the inertia group $I_{\mathfrak{p}}$, whence ρ^0 is trivial on $I_{\mathfrak{p}}$ as well (indeed, for $g \in I_{\mathfrak{p}}$, evidently $\tilde{\rho}(g) = \mathrm{id}$, so that g acts as the identity on each Jordan-Hölder constituent of $\tilde{\rho}^0$, whence $\rho^0(g) = \mathrm{id}$).

As for the second part, we follow the discussion in Section 7.3.5.1. Let $\lambda \subseteq \mathfrak{o}_K$ be a prime of K with $\lambda | (\ell)$. To say that SeemspDivisible$(\rho_N, g, K, \ell, \lambda, N)$ returns True for all N is to say that, for all N, the finite commutative étale group scheme G_{ρ_N}/K_λ, given by the the $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)$-module $(\mathbb{Z}/\ell^N)^{\oplus 2g}$ with action given by $\rho_N|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}$, has an integral model $\mathcal{G}_{\rho_N}/\mathfrak{o}_{K,\lambda}$. By compatibility of the ρ_N we find that the ℓ-divisible group G_ρ/K_λ (which we note has n-th layer G_{ρ_n}/K_λ) whose Tate module is given by $\tilde{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}$ has an integral model $\mathcal{G}_\rho/\mathfrak{o}_{K,\lambda}$ an ℓ-divisible group over $\mathfrak{o}_{K,\lambda}$ (implicitly we use Kőnig's Lemma to choose the $\mathcal{G}_{\rho_N}/\mathfrak{o}_{K,\lambda}$ compatibly). It follows that $\tilde{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/K_\lambda)}$ is crystalline at λ with all Hodge-Tate weights either 0 or -1.

Now because $\tilde{\rho}$ has similitude character the ℓ-adic cyclotomic character it follows by autoduality that its Hodge-Tate weights under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$ are $\underbrace{0, \ldots, 0}_{g}, \underbrace{-1, \ldots, -1}_{g}$.

It finally remains to deal with the Hodge-Tate weights of ρ^0. First, since the property of being crystalline is preserved under passing to subquotients (and ρ^0 is a direct sum of Jordan-Hölder constituents of $\tilde{\rho}^0$), it follows that ρ^0 is crystalline under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$ as well. Moreover, because the D_{HT} functor is exact on the category of Hodge-Tate representations (and thus on the subcategory of crystalline representations), it follows that the Hodge-Tate weights of ρ^0 and $\tilde{\rho}^0$ match (since they are both multiset unions of the Hodge-Tate weights of the local representations of the various Jordan-Hölder constituents of $\tilde{\rho}^0$). Thus we have that ρ^0 is crystalline (hence de Rham) under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$ with Hodge-Tate weights $\underbrace{0, \ldots, 0}_{g}, \underbrace{-1, \ldots, -1}_{g}$.

Therefore, by virtue of being finitely ramified, semisimple, and de Rham at primes above ℓ, it follows from combining the Fontaine-Mazur conjecture and the Grothendieck-Serre conjecture that there is a smooth projective variety X/K, an $i \in \mathbb{N}$, and a $k \in \mathbb{Z}$ for which ρ^0 is a summand of $H^i_{\text{ét.}}(X_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)(k)$.

By the Tate conjecture, the projector $H^i_{\text{ét.}}(X_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)(k) \to H^i_{\text{ét.}}(X_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)(k)$ corresponding to projection onto the summand ρ^0 is given by a $\mathbb{Q}_\ell$-linear combination of (maps induced by) correspondences in $X \times X$. Proceeding exactly as in the first paragraph of the proof of Lemma 3.3 in Patrikis-Voloch-Zarhin's [77], we find that, because of the Tate conjecture, we may in fact take the coefficients of the linear combination to lie in $\overline{\mathbb{Q}}$. In other words, there is a finite set $\mathscr{C}$ of $\dim X$-dimensional correspondences $C \subseteq X \times X$ and $\alpha_C \in \overline{\mathbb{Q}}$ for which

$$\pi := \sum_{C \in \mathscr{C}} \alpha_C \cdot C_* \in \mathrm{End}(H^i_{\text{ét.}}(X_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell))$$

is the projector onto ρ^0.

Let $E := \mathbb{Q}(\{\alpha_C\}_{C \in \mathscr{C}})$, a number field. Thus the pure motive over K with E-coefficients given by

$$\tilde{M} := (X, \pi, k)$$

has ℓ-adic realization ρ^0 — i.e., $\tilde{M}_\ell \cong \rho^0$.

Let M be the pure motive over K with $\mathbb{Q}$-coefficients induced by $\tilde{M}$ (so that, e.g., $M_\ell \cong \tilde{M}_\ell^{\oplus[E:\mathbb{Q}]} \cong (\rho^0)^{\oplus[E:\mathbb{Q}]}$).

Let M_{B} be the Betti realization of M. Since the Hodge-Tate weights (under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$) of M_ℓ are $\underbrace{0, \ldots, 0}_{g \cdot [E:\mathbb{Q}]}, \underbrace{-1, \ldots, -1}_{g \cdot [E:\mathbb{Q}]}$, it follows that M_{B} is a polarizable $\mathbb{Q}$-Hodge structure of type $\underbrace{(1,0), \ldots, (1,0)}_{g \cdot [E:\mathbb{Q}]}, \underbrace{(0,1), \ldots, (0,1)}_{g \cdot [E:\mathbb{Q}]}$.

We now argue as in Blasius's [25] and Patrikis-Voloch-Zarhin's [77] (note that, just as in Blasius, the latter argument only needs[16] the absolute Hodge conjecture). By the absolute Hodge conjecture, there is an absolute Hodge class in $H^{2\dim X}_{\mathrm{B}}(X \times X, \mathbb{Q})$ inducing $H^i_{\mathrm{B}}(X, \mathbb{Q}) \twoheadrightarrow M_{\mathrm{B}} \subseteq H^i_{\mathrm{B}}(X, \mathbb{Q})$. Let $L/\mathbb{Q}$ be the fixed field of this absolute Hodge class — because the class is absolute Hodge, $L/\mathbb{Q}$ is a number field. By enlarging L if necessary, we may assume $K \subseteq L$.

Now, by the Riemann classification, there is a unique isogeny class of abelian varieties $A/\mathbb{C}$ with $H^1_{\mathrm{B}}(A(\mathbb{C}), \mathbb{Q}) \cong M_{\mathrm{B}}$. Because this isogeny class is countable, it follows that we may take $A/\overline{\mathbb{Q}}$ — i.e., without loss of generality A is defined over $\overline{\mathbb{Q}}$. By again enlarging L if necessary, without loss of generality A is defined over L.

[16] Admittedly in the context of this chapter this is an irrelevant point, since the absolute Hodge conjecture and the Tate conjecture combine to imply the Hodge conjecture.

Therefore, using the same absolute Hodge class and the comparison theorem between ℓ-adic and Betti cohomology, we find that

$$H^1_{\text{ét.}}(A_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \cong \mathrm{Res}^L_K(M_\ell).$$

Thus we find a surjection

$$H^1_{\text{ét.}}(\mathrm{Res}^L_K(A)_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \twoheadrightarrow M_\ell,$$

where as usual $\mathrm{Res}^L_K(A)$ is the Weil restriction of scalars of A.

We now repeat the argument we gave to construct M, except with X replaced by $\mathrm{Res}^L_K(A)$. By Faltings' proof of the Tate conjecture for abelian varieties and semisimplicity of $\mathrm{End}_K(\mathrm{Res}^L_K(A)) \otimes_{\mathbb{Z}} \mathbb{Q}$ we see that this surjection is induced by a projector in $\mathrm{End}_K(\mathrm{Res}^L_K(A)) \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}$. Replacing E by an extension (thus A by a power) if necessary, we therefore find a projector in $\mathrm{End}_K(\mathrm{Res}^L_K(A)) \otimes_{\mathbb{Z}} \mathbb{Q}$.

It follows that there is an isogeny factor $\tilde{A}/K$ of $\mathrm{Res}^L_K(A)$ such that:

$$H^1_{\text{ét.}}(\tilde{A}_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \cong M_\ell.$$

Thus by the Néron-Ogg-Shafarevich criterion it follows that the $[E : \mathbb{Q}] \cdot g$-dimensional abelian variety $\tilde{A}/K$ has good reduction outside S. Moreover it follows that

$$\mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright H^1_{\text{ét.}}(\tilde{A}_{/\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)) = [E : \mathbb{Q}] \cdot a_{\mathfrak{p}}$$

for all $\mathfrak{p} \in \tilde{T}_{10^{10} \cdot [E:\mathbb{Q}]}$. Thus we have produced the desired abelian variety $\tilde{A}$ (and the desired $d \in \mathbb{Z}^+$, namely $d := [E : \mathbb{Q}]$), and the proof of the theorem is complete. □

Chapter 8

$C_f(K), \mathcal{V}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$: Bogomolov-Tschinkel.

Abstract.

We give an algorithm that, on input (C, K) with C/K a smooth projective hyperbolic hyperelliptic curve over a number field K, outputs $C(K)$, along with a certificate of correctness of the output. Assuming a modularity conjecture for fake elliptic curves over all number fields, we prove this algorithm always terminates in finite time with the correct output.

The key trick is to use a theorem of Bogomolov-Tschinkel to reduce the problem to that of finding the rational points on the single curve $C_6 : y^2 = x^6 + 1$ over an explicit extension L/K, and then to notice that C_6 is (a twist of) a Shimura curve with level structure to produce a family of abelian surfaces with quaternionic multiplication (i.e. "fake elliptic curves") over C_6.

8.1 Introduction.

8.1.1 Main theorem.

In this chapter we prove the following theorem.

Theorem 8.1.1. *1. The effective Shafarevich conjecture for genus 2 Jacobians over number fields implies the effective Mordell conjecture for all smooth projective hyperbolic solvable covers of $\mathbb{P}^1$ (in particular, for hyperbolic hyperelliptic curves) over number fields.*

2. Moreover, Conjecture 8.1.2 implies that there is a finite time algorithm that, on input (C, K) with C/K a smooth projective hyperbolic curve over a number field K that can be realized as a solvable cover of $\mathbb{P}^1$ over $\overline{\mathbb{Q}}$, outputs $C(K)$. In particular, Conjecture 8.1.2 implies that there is a finite time algorithm that, on input (C, K) with C/K hyperbolic hyperelliptic, outputs $C(K)$.

It is worth noting that Baker's effective solution of S-unit equations provides a solution of an effective Shafarevich conjecture for genus 2 (indeed, any hyperelliptic) curves, but this is not really relevant to the effective Shafarevich conjecture for their Jacobians.

We also point out that Levin [66] proved that the effective Shafarevich conjecture for Jacobians of hyperelliptic curves of genus g would imply an effective Siegel theorem for integral points on hyperelliptic curves of genus g. Theorem 8.1.1 of course improves this statement.

8.1.2 Main conjecture.

Let us now state Conjecture 8.1.2.

Conjecture 8.1.2 (Cf. e.g. Taylor's 1994 ICM address [101], after "Further, it is now standardly conjectured that…")**.** *Let $F/\mathbb{Q}$ be a number field. Let $K/\mathbb{Q}$ be a number* *field, with r_1 real places and r_2 complex places (so that $r_1 + 2r_2 = d := [K : \mathbb{Q}]$). Let $\mathfrak{Z}_K := (\mathfrak{h}_2^{\pm})^{r_1} \times \mathfrak{h}_3^{r_2}$, where $\mathfrak{h}_2^{\pm} := \mathbb{C} - \mathbb{R}$ and $\mathfrak{h}_3$ is hyperbolic 3-space. Let*

$$Y_1(1) := \mathrm{GL}_2(K)\backslash \left(\left(\mathrm{GL}_2(\mathbb{A}_{K,f})/\prod_{\mathfrak{p}} \mathrm{GL}_2(\mathfrak{o}_{K,\mathfrak{p}})\right) \times \mathfrak{Z}_K\right),$$

where $\mathbb{A}_{K,f} := K \otimes_{\mathfrak{o}_K} \prod_{\mathfrak{p}} \mathfrak{o}_{K,\mathfrak{p}}$ is the ring of finite adeles of K. Let $\mathbb{T}_1(1)$ denote the image of the Hecke algebra in the endomorphisms of $H^d(Y_1(1), \mathbb{C})$ and $S_{\mathfrak{p}}, T_{\mathfrak{p}} \in \mathbb{T}_1(1)$ the standard Hecke operators.

Then: there is a bijection between nontrivial characters $\theta : \mathbb{T}_1(1) \to F$ and isogeny classes of $2 \cdot [F : \mathbb{Q}]$-dimensional abelian varieties A/K without sufficiently many complex multiplications over K, with good reduction everywhere, and with K-endomorphism algebra $\mathrm{End}^0_K(A)$ containing a quaternion algebra over F, such that, for all coprime primes ℓ of $\mathbb{Z}$ and $\mathfrak{p}$ of K,

$$2 \cdot [F : \mathbb{Q}] \cdot \theta(T_{\mathfrak{p}}) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A)).$$

Let us immediately emphasize that we do not mean to suggest that this is a conjecture of Taylor's, and indeed we will essentially use the statement in the same way he does in [101], namely to provide a context in which to discuss a new technique.[1]

In any case, here

$$V_\ell(A) := T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$$

with $T_\ell(A) := \varprojlim A[\ell^n]$ the usual ℓ-adic Tate module of A. Recall that a generalized fake elliptic curve over K is an abelian variety A/K for which its algebra of K-endomorphisms contains a quaternion algebra with centre a totally real number field of degree $\frac{\dim A}{2}$ over $\mathbb{Q}$. We note that we will only really need the above conjecture in the particular case of abelian surfaces (i.e. when the centre $F = \mathbb{Q}$), in

[1] In fact the existence of non-PEL-type quaternionic Shimura varieties seems to complicate things quite a bit over a general number field, because of Deligne's absolute Hodge theorem for abelian varieties (consider e.g. the ℓ-adic monodromy groups of fibres in the hypergeometric family of abelian varieties associated to a nonarithmetic triangle group). In any case the techniques we give will of course be robust to any such small modifications.

which case the terminology "fake elliptic curve" is used. A more general conjecture can be made by adding level structures (thus the notation $\mathbb{T}_1(1), Y_1(1)$, etc.), but for us this is irrelevant because we use an explicit form of Grothendieck's semistable reduction theorem to pass to an explicit extension over which our abelian varieties have good reduction everywhere (recall that division algebras have no nontrivial unipotents).

8.1.3 A theorem of Bogomolov-Tschinkel.

We, of course, use a theorem of Bogomolov-Tschinkel [26], though in an improved form due to Poonen [80]. Let

$$C_6 : y^2 = x^6 + 1.$$

Theorem 8.1.3 (Poonen, Bogomolov-Tschinkel). *Let (C, K, f) be such that C/K is a smooth projective hyperbolic curve over a number field K and $f : C \to \mathbb{P}^1$ is a solvable cover of $\mathbb{P}^1$ defined over K. Then: there is an effectively computable (in terms of (C, K, f)) tuple $(\tilde{C}, L, \varphi, g)$ with $\tilde{C}/L$ a smooth projective hyperbolic curve over a number field L, $\varphi : \tilde{C} \to C$ an étale map defined over L, and $g : \tilde{C} \to C_6$ a nonconstant map defined over L.*

The reason this is relevant to us is the following particular case of a standard theorem of Chevalley-Weil [38] (see also pages 65-67 of Weil's collected works [106]).

Theorem 8.1.4 (Chevalley-Weil). *Let K be a number field. Let $C, C'/K$ be hyperbolic curves and $\varphi : C \to C'$ an étale map defined over K. Then: there is an explicitly computable finite extension L/K such that*

$$C'(K) \subseteq \varphi(C(L)).$$

(In other words, all K-rational points of C lift to L-rational points of C' over an explicitly computable finite extension L/K — indeed, L is computed as the compositum of all extensions of K which have explicitly bounded degree and which are unramified outside an explicit finite set of places.)

8.1.4 Main idea.

The content of this chapter amounts to the following very simple observation.

Lemma 8.1.5. *The Belyi map $C_6 \to \mathbb{P}^1$ via $(x, y) \mapsto x^6$ realizes C_6 as a twist of a Shimura curve. Consequently, there is a nonconstant map*

$$C_6 \to A_2$$

defined over $\overline{\mathbb{Q}}$, with image the locus of isomorphism classes of principally polarized abelian surfaces with quaternionic multiplication by an explicit order in the discriminant 6 (thus indefinite) quaternion algebra over $\mathbb{Q}$.

The point is that, via Chevalley-Weil (Theorem 8.1.4) and Bogomolov-Tschinkel (Theorem 8.1.3) as above, we are reduced to finding, for an explicit number field L/K and finite set S of places of L, the abelian surfaces A/L with good reduction outside S and admitting $B_6 \hookrightarrow \mathrm{End}^0_L(A)$, where $B_6/\mathbb{Q}$ is the quaternion algebra over $\mathbb{Q}$ ramified exactly at 2 and 3.

We then proceed to, "by day", diagonalize the actions on $H^d(Y_1(1), \mathbb{C})$ of more and more Hecke operators $T_\mathfrak{p}$, with $\mathfrak{p}$ of larger and larger norm — $\mathrm{Nm}\,\mathfrak{p} \leq X$, say — and discard characters not taking rational values at all such $T_\mathfrak{p}$. "By night", we enumerate abelian surfaces A/L of larger and larger height and compute their $a_\mathfrak{p}$. If an A/L has $(a_\mathfrak{p})_{\mathrm{Nm}\,\mathfrak{p} \leq X}$ in the "by day" set of tuples, and X is sufficiently large (so that Faltings' Lemma, i.e. Lemma 7.2.1 of Chapter 7, applies), then by Conjecture 8.1.2 A/L must be L-isogenous to an abelian surface of the desired type.

By invoking Masser-Wustholz we bound the height of an abelian surface over L in the corresponding isogeny class, and then move to the next tuple.

And again by Conjecture 8.1.2, all the abelian surfaces we search for must 'always' (i.e. for all X) match a tuple in the "by day" set, and a tuple in the "by day" set that 'always survives' (i.e. arises from a character $\mathbb{T}_1(1) \to \mathbb{Q}$) must come from such an abelian surface. So eventually in the "by night" step we will find an abelian surface matching an 'always surviving' tuple, so that the algorithm will terminate.

8.2 The algorithm.

Let us now precisely specify the algorithm alluded to above. In fact we will give an algorithm solving a slight generalization of the problem at hand: we will give an algorithm that finds the S-integral K-rational points on the canonical integral model of a PEL-type quaternionic Shimura variety in finite time — i.e. we will allow our abelian varieties A/K to have bad reduction at some places, and endomorphisms by an order in a totally indefinite quaternion algebra over a larger number field of degree $\frac{\dim A}{2}$ over $\mathbb{Q}$ (rather than just by an order in the indefinite quaternion algebra of discriminant 6 over $\mathbb{Q}$). Again, the case relevant to Theorem 8.1.1 is the case of $\mathfrak{o}$ the maximal order in the indefinite quaternion algebra $B_6/\mathbb{Q}$ of discriminant 6.

We write $V_{\mathfrak{o}}$ for the Shimura variety corresponding to $\mathfrak{o}$ (without loss of generality $V_{\mathfrak{o}}$ will be defined over K), and $\mathcal{V}_{\mathfrak{o}}$ for its canonical integral model over $\mathfrak{o}_{K,S}$ (S will always be sufficiently large so that there are no subtle issues about integral models of our varieties).

8.2.1 GeneralizedFakeEllipticCurves($\mathfrak{o}, K, S$):

Input: $\mathfrak{o}, K, S$, with $\mathfrak{o}$ an order in a quaternion algebra $B := \mathfrak{o} \otimes_{\mathbb{Z}} \mathbb{Q}$ with centre a totally real number field F, $K/\mathbb{Q}$ an explicitly sufficiently large number field (so that K splits B and $V_{\mathfrak{o}}$ is defined over K), and S an explicitly sufficiently large finite set of places of K (so that S contains all infinite primes of K, all ramified primes of K, all primes of K over $2, 3, 5$, and 7, and all primes not coprime to the ramified primes of B).

Output: $\mathcal{V}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$, in other words[2] the $2 \cdot [F : \mathbb{Q}]$-dimensional abelian varieties A/K with good reduction outside S and admitting a map $\mathfrak{o} \hookrightarrow \mathrm{End}_K(A)$.

Algorithm:

1. Let ℓ be a prime of $\mathbb{Z}$ not lying below any prime of S. Let E/K be the compositum of all extensions of K unramified outside S and of degree $\leq 10^{10 \cdot [F:\mathbb{Q}]}$. Let T be the finite set of places of E lying over a place in S. Let r_1 and r_2 be the respective numbers of real and complex places of E, and let $d := [E : \mathbb{Q}]$.

2. Let $\mathcal{V}_{\mathfrak{o}}^{\mathrm{CM}}(\mathfrak{o}_{E,T})$ be the set of abelian varieties of the form A^{Φ}/E, where Φ is a CM type of the maximal imaginary CM subfield E/E^{CM} of E (if one does not exist then $\mathcal{V}_{\mathfrak{o}}^{\mathrm{CM}}(\mathfrak{o}_{E,T}) := \emptyset$), and A^{Φ} is the CM abelian variety corresponding to the reflex CM type of (E^{CM}, Φ).

3. Let $\Sigma :=$ output of FaltingsPrimeList$(2 \cdot [F : \mathbb{Q}], E, T, \ell)$.

4. Let $N := 10^{10}$.

5. Let
$$\Sigma_N := \Sigma \cup \{\mathfrak{p} \subseteq \mathfrak{o}_E : \mathrm{Nm}\,\mathfrak{p} \leq N, (\mathrm{Nm}\,\mathfrak{p}, \ell \cdot \prod_{\mathfrak{q} \in T} \mathrm{Nm}\,\mathfrak{q}) = 1\}.$$

6. Let
$$\Lambda_N := \{\theta : \mathbb{T}_1(1) \to \mathbb{C} : \forall \mathfrak{p} \in \Sigma_N, \theta(T_{\mathfrak{p}}) \in \mathfrak{o}_F\}.$$

[2] Recall that in this thesis we completely ignore stack-theoretic issues (since they are irrelevant).

7. Let

$$\mathscr{C}_N := \left\{ A/E \;\middle|\; \begin{array}{c} \dim A = 2\cdot[F:\mathbb{Q}], h(A) \leq N, \\ \exists \theta \in \Lambda_N : \forall \mathfrak{p} \in \Sigma_N, 2\cdot[F:\mathbb{Q}]\cdot\theta(T_\mathfrak{p}) = \operatorname{tr}(\mathrm{Frob}_\mathfrak{p} \curvearrowright V_\ell(A)) \end{array} \right\}.$$

8. If, for some $\theta \in \Lambda_N$, there is *no* $A \in \mathscr{C}_N$ for which

$$2\cdot[F:\mathbb{Q}]\cdot\theta(T_\mathfrak{p}) = \operatorname{tr}(\mathrm{Frob}_\mathfrak{p} \curvearrowright V_\ell(A))$$

for all $\mathfrak{p} \in \Sigma_N$, then increment $N \mapsto N+1$ and return to Step 5.

9. Let[3]

$$H := \max_{A \in \mathscr{C}_N \cup \mathcal{V}_\mathfrak{o}^{\mathrm{CM}}(\mathfrak{o}_{E,T})} \text{MasserWüstholz}(A, E).$$

10. Output the points in $\mathcal{V}_\mathfrak{o}(\mathfrak{o}_{K,S})$ of height $\leq H$.

8.2.1.1 Explanation in words.

The extension E/K is chosen so that, for an abelian variety A/K with quaternionic multiplication by B/F of dimension $\dim A = 2\cdot[F:\mathbb{Q}]$, the base change A/E has good reduction everywhere (see Lemma 8.3.2).

So we search instead for abelian varieties over E with good reduction everywhere. (This is why we have formulated Conjecture 8.1.2 sans level structure.)

Because Conjecture 8.1.2 concerns non-CM abelian varieties, we must treat CM abelian varieties by hand. Of course this is simple: if an abelian variety A/E has sufficiently many complex multiplications over E, its λ-adic (with λ a prime of F) Galois representations are abelian, and the characteristic polynomials of the $\mathrm{Frob}_\mathfrak{p}$

[3] Here $\text{MasserWüstholz}(A,K) \in \mathbb{R}^+$ is an explicitly computed constant such that, for all B/K abelian K-varieties that are K-isogenous to A/K,

$$h(B) \leq \text{MasserWüstholz}(A,K).$$

See Theorem 7.2.2 of Chapter 7 for precise constants.

have coefficients in F. Thus by the main result of Henniart's [56], it follows that they are direct sums of the λ-adic realizations of algebraic Hecke characters of E.

By the classification of algebraic Hecke characters (and the fact that the λ-adic representation of A/E is pure of weight 1), it follows that if A has sufficiently many complex multiplications over E then it must be of the form given in Step 2. So Step 2 indeed deals with the CM abelian varieties.

Now we discuss the non-CM abelian varieties.

The purpose of Step 3 is to ensure our sets of primes in Step 5 are sufficiently large to 'separate L-functions', so to speak. And indeed by definition the $(a_{\mathfrak{p}})_{\mathfrak{p}\in\Sigma}$ determines such an abelian variety A/E up to E-isogeny.

Step 4 gives the initial conditions of a loop. Step 5 defines the set of primes $\mathfrak{p}$ for which we will diagonalize the actions of the Hecke operators $T_{\mathfrak{p}}$ on $H^d(Y_1(1), \mathbb{C})$, in order to check rationality of the various characters of the Hecke algebra prime-by-prime. Step 6 defines the characters that 'look' rational — this set is computed by exact linear algebra and knowledge of characteristic polynomials of Hecke operators.[4] In our "day/night" description of the algorithm, Step 6 is the "by day" step.

Step 7, then, is the "by night" step. In it we simply enumerate abelian varieties of the correct dimension and larger and larger height, and compute their tuples $(a_{\mathfrak{p}})_{\mathfrak{p}\in\Sigma_N}$ to see if they match the characters found in Step 6. If there is at least one that does not match, we repeat the loop after incrementing N — this is Step 8.

Once we reach Step 9 we know that all characters in Λ_N must have matched some abelian variety in $\mathscr{C}_N$ (see Conjecture 8.1.2 to see why we will always exit the loop and reach Step 9). We are essentially done, except that we have only determined a set of abelian varieties with the property that the abelian varieties

[4] Specifically, one takes more and more Hecke operators so that all common eigenspaces are one-dimensional (a rank calculation), and then after that one can express the statement that $T_{\mathfrak{q}}$ acts by $\theta(T_{\mathfrak{q}})$ on the common eigenspace of the $T_{\mathfrak{p}}$ as the vanishing of a determinant.

we are interested in are all *E-isogenous* to an element in this set. So we use Masser-Wüstholz (we could also use a result of Raynaud, making precise Faltings' isogeny estimate) to produce a height bound on the abelian varieties we are interested in, and then enumerate in Step 10.[5]

8.3 Proof of termination and correctness.

In this section we prove the following theorem.

Theorem 8.3.1. *Let $K/\mathbb{Q}$ be a number field. Let S be a finite set of places of K. Let $F/\mathbb{Q}$ be a totally real number field. Let $\mathfrak{o} \subseteq B$ be an order in a totally indefinite quaternion algebra B/F over F. Then: on input $(\mathfrak{o}, K, S)$, the algorithm specified in Section 8.2.1 terminates with output $\mathcal{V}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$.*

We will see in Section 8.4 that this implies Theorem 8.1.1.

8.3.1 Preliminary lemmas.

We first prove Lemma 8.1.5. We note that the Lemma is implicit in a table of Wolfart (see pages 17 and 18 of Wolfart's [107], particularly the fourth line of the table on page 18).

Proof of Lemma 8.1.5. The degree 12 map $C_6 \to \mathbb{P}^1$ via $(x, y) \mapsto x^6$ is unramified outside $\{0, 1, \infty\}$, and, over these points, has ramification degrees $6, 2, 6$, respectively. In fact it also exhibits C_6 as a Galois ramified cover of $\mathbb{P}^1$, with Galois group $\mathbb{Z}/2 \times \mathbb{Z}/6$, with action

$$(a, b) \cdot (x, y) := (\zeta_6^b \cdot x, (-1)^a \cdot y).$$

[5] We note that this is an especially lazy way of completing the endgame, since we already have explicit representatives of each isogeny class we are interested in in hand. We also note that implicit here is a comparison between the Faltings height and the naïve height arising from a Baily-Borel/Satake compactification — it is much easier to enumerate bounded-height points in projective space, after all.

Thus we have produced an unramified $\mathbb{Z}/2 \times \mathbb{Z}/6$-Galois cover $C_6 \to \mathbb{P}^1(2,6,6) := \mathfrak{h}/\Delta(2,6,6)$. On the other hand, $\mathfrak{h}/[\Delta(2,6,6), \Delta(2,6,6)] \to \mathfrak{h}/\Delta(2,6,6) = \mathbb{P}^1(2,6,6)$ is an unramified cover with Galois group $\Delta(2,6,6)^{\mathrm{ab}} \simeq \mathbb{Z}/2\times\mathbb{Z}/6$. It is fairly simple (see e.g. Lemma $2.1.2$ of [76]) to classify such covers, and there is only one. Because $\Delta(2,6,6)$ is arithmetic, the claim follows. □

We will also use the following standard observation:

Lemma 8.3.2. *Let $F/\mathbb{Q}$ be a totally real number field. Let B/F be a quaternion algebra over F. Let $K/\mathbb{Q}$ be a number field, and S a finite set of places of K containing all places of K above $2,3,5,7$, and ∞. Let L/K be the compositum of all extensions of degree $\leq 10^{10\cdot g}$ that are unramified outside S. Let A/K be an abelian variety of dimension $2g$ admitting $B \hookrightarrow \mathrm{End}^0_K(A)$. Then: A/L has good reduction everywhere.*

Proof. This is because of the explicit form of Grothendieck's semistable reduction theorem: A/E has semistable reduction because E contains the field $K(A[210])$, for example, since said field has suitably bounded degree and ramification.

Thus at a finite prime the special fibre of the Néron model is semiabelian. Writing T for its torus part, we find (since endomorphisms act on the special fibre by naturality of the Néron model) $B \hookrightarrow \mathrm{End}^0(T) \simeq M_{\dim T}(\mathbb{Q})$, an injection since B is a division algebra. In other words we find a B-module structure on $\mathbb{Q}^{\dim T}$. Hence

$$4 \cdot [F:\mathbb{Q}] = \dim_{\mathbb{Q}} B \mid \dim T \leq 2 \cdot [F:\mathbb{Q}].$$

Therefore $\dim T = 0$ and so A has good reduction at this prime.[6] □

[6] This is in fact the same argument as is given in Boutot's proof of the Proposition in Section 5 of his Expose III in [32], as well as essentially the same argument as is given in Ribet's proof of Theorem 3 in his [82].

In the particular case that is relevant to us — namely $B = B_6$, the indefinite quaternion algebra of discriminant 6 over $F = \mathbb{Q}$ — it is simpler to just note that the 2-adic and 3-adic Galois representations land in the units of the division algebras $B \otimes_{\mathbb{Q}} \mathbb{Q}_2$ and $B \otimes_{\mathbb{Q}} \mathbb{Q}_3$, which have no nontrivial unipotents.

8.3.2 Proof of Theorem 8.3.1.

We break the proof into two parts for clarity. First we prove that, if the algorithm terminates, its output is correct. After that we will prove that the algorithm always terminates.

8.3.2.1 Proof of correctness.

Proof of correctness assuming termination. We first deal with the proof of correctness. Because of the last step of the algorithm in Section 8.2.1, it suffices to show that any abelian variety A/K with $\mathfrak{o}$-multiplication over K and good reduction outside S has height bounded by the H computed in Step 9 of the algorithm. By Lemma 8.3.2, A/E has good reduction everywhere. Moreover $h(A/E) = h(A/K)$ since the Faltings height is unchanged under base extension.

We now split into two cases.

If A/E has sufficiently many complex multiplications over E, then we claim that A is E-isogenous to an element of the set $\mathcal{V}_{\mathfrak{o}}^{\mathrm{CM}}(\mathfrak{o}_{E,T})$ defined in Step 2. Indeed, we repeat the argument given in Section 8.2.1.1: its λ-adic (with λ a prime of F) Galois representations are abelian, and the characteristic polynomials of the $\mathrm{Frob}_{\mathfrak{p}}$ have coefficients in F. Thus by the main result of Henniart's [56], it follows that these λ-adic representations are direct sums of λ-adic realizations of algebraic Hecke characters χ of E.

By the classification of algebraic Hecke characters (and the fact that the λ-adic representation of A/E is pure of weight 1 — see the proof of Lemma 9.3.4 in Chapter 9 for a lengthier explanation), the claim follows.

Therefore in case A/E has CM, it follows that $h(A) \leq H$, by Masser-Wüstholz (see Theorem 7.2.2 in Chapter 7 for a precise statement).

Now we deal with the case that A/E does not have CM. By Conjecture 8.1.2 applied to E, there is a $\theta : \mathbb{T}_1(1) \to F$ with

$$2 \cdot [F : \mathbb{Q}] \cdot \theta(T_{\mathfrak{p}}) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A))$$

for all $\mathfrak{p}$ prime to ℓ. Thus $\theta \in \Lambda_N$ for all N.

Because we have assumed the algorithm terminates, eventually we reach Step 9. Let M be the value of the parameter N when the algorithm passes to Step 9. Since $\theta \in \Lambda_M$, it follows that there is an $A' \in \mathscr{C}_M$ for which

$$2 \cdot [F : \mathbb{Q}] \cdot \theta(T_{\mathfrak{p}}) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A'))$$

for all $\mathfrak{p} \in \Sigma_M$. Hence $\mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A)) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A'))$ for all $\mathfrak{p} \in \Sigma_M$, hence there is an E-isogeny $A \sim_E A'$, since $\Sigma \subseteq \Sigma_M$.

Hence by Masser-Wüstholz $h(A) \leq H$, as desired. □

8.3.2.2 Proof of termination.

Proof of termination. Now let us prove that the algorithm terminates. Evidently Steps $1, 2, 3, 4, 5, 6$, and 7 terminate. If we reach Step 9 (i.e. Step 8 terminates without returning to Step 5), then, since it is evident that Steps 9 and 10 terminate, we are done. So it suffices to show that we reach Step 9. In other words, we must show that, for N sufficiently large, and for all $\theta \in \Lambda_N$, there is an $A \in \mathscr{C}_N$ for which

$$2 \cdot [F : \mathbb{Q}] \cdot \theta(T_{\mathfrak{p}}) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A))$$

for all $\mathfrak{p} \in \Sigma_N$. But $\Lambda_N \subseteq \Lambda_{10^{10}}$ for all $N \geq 10^{10}$, and the latter set is finite. Thus for N sufficiently large, all $\theta \in \Lambda_N$ are such that $\theta : \mathbb{T}_1(1) \to F$. Thus, for such N, and for $\theta \in \Lambda_N$, by Conjecture 8.1.2 there is an abelian variety A/E with $\dim A = 2 \cdot [F : \mathbb{Q}]$, with good reduction everywhere, and for which

$$2 \cdot [F : \mathbb{Q}] \cdot \theta(T_{\mathfrak{p}}) = \mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}} \curvearrowright V_\ell(A))$$

for all $\mathfrak{p}$ prime to ℓ. Increasing N if necessary so that $h(A) \leq N$, we find that this $A \in \mathscr{C}_N$. So we are done. □

8.4 Proof of Theorem 8.1.1.

Let us first set notation.

Because $C_6/\mathbb{Q}$ is a twist of a Shimura curve, there is a number field F such that C_6/F is (the base change of) a Shimura curve over F. Thus there is a map $\psi : C_6 \to A_2$ defined over F.

Let T be a finite set of primes of F that is sufficiently large so that the map $\psi : C_6 \to A_2$ extends to

$$\Psi : \mathcal{C}_6 \to \mathcal{A}_2,$$

a map of integral models over $\mathfrak{o}_{F,T}$, where $\mathcal{C}_6$ is the minimal proper regular model of C_6 over $\mathfrak{o}_{F,T}$ and $\mathcal{A}_2$ is the canonical integral model of A_2 over $\mathfrak{o}_{F,T}$ — note that we implicitly take T to be very large, so that we avoid any subtleties whatsoever about these integral models (e.g. $\mathcal{C}_6$ is smooth and has the Néron property over $\mathfrak{o}_{F,T}$).

We first prove part 1 of Theorem 8.3.1.

Proof of Theorem 8.1.1, part 1. Assume an effective form of the Shafarevich conjecture for Jacobians of genus two curves over number fields — in other words, that $A_2(\mathfrak{o}_{K,S})$ is effectively computable in terms of (K, S) with K a number field and S a finite set of places of S.[7] Let C/K be a smooth projective hyperbolic curve

[7] This is equivalent because, by Baker, the effective Shafarevich conjecture holds for moduli of elliptic curves, so that we may ignore the reducible locus.

over a number field K that is equipped with a map $C \to \mathbb{P}^1$ that is defined over K and that realizes C as a solvable cover of $\mathbb{P}^1$. By Poonen/Bogomolov-Tschinkel (Theorem 8.1.3), there is an effectively computable (in terms of this data) tuple $(\tilde{C}, L, \varphi, f)$ with $\varphi : \tilde{C} \to C$ étale, $f : \tilde{C} \to C_6$ nonconstant, and everything defined over the number field L/K. By Chevalley-Weil (Theorem 8.1.4), there is an effectively computable $\tilde{L}/L$ for which

$$C(K) \subseteq C(L) \subseteq \varphi(\tilde{C}(\tilde{L})).$$

By replacing $\tilde{L}$ by $F\tilde{L}$ if necessary, without loss of generality $F \subseteq \tilde{L}$. Note that

$$f(\tilde{C}(\tilde{L})) \subseteq C_6(\tilde{L}),$$

and that it is easy to determine if an element of $C_6(\tilde{L})$ lies in $f(\tilde{C}(\tilde{L}))$.

It follows that we need only determine $C_6(\tilde{L})$. Let S be a finite set of primes of $\tilde{L}$ containing all primes over T. Then, (the isomorphism by the Néron property, though we do not need it)

$$C_6(\tilde{L}) \simeq \mathcal{C}_6(\mathfrak{o}_{\tilde{L},S}) \xrightarrow{\Psi} \mathcal{A}_2(\mathfrak{o}_{\tilde{L},S}).$$

By hypothesis, we may effectively determine $\mathcal{A}_2(\mathfrak{o}_{\tilde{L},S})$. We conclude by testing each element of $\mathcal{A}_2(\mathfrak{o}_{\tilde{L},S})$ for membership in the image of ψ, which is easy. Thus the first part follows. □

Finally we prove that Theorem 8.3.1 implies part 2 of Theorem 8.1.1.

Proof that Theorem 8.3.1 implies part 2 of Theorem 8.1.1. We modify the above proof of part 1 of Theorem 8.1.1 by computing $\Psi(\mathcal{C}_6(\mathfrak{o}_{\tilde{L},S}))$ assuming Conjecture 8.1.2 instead of an effective form of the Shafarevich conjecture for genus two Jacobians.

We note that the image of the map $\psi : C_6 \to A_2$ lies in the locus of principally polarized abelian surfaces with endomorphisms by an order in the indefinite quaternion algebra $B_6/\mathbb{Q}$ of discriminant 6. Therefore each point of $\Psi(\mathcal{C}_6(\mathfrak{o}_{\tilde{L},S}))$ is a principally polarized abelian surface over $\tilde{L}$ with good reduction outside S and $\tilde{L}$-endomorphisms by an explicit order $\mathfrak{o}$ in B_6. Because it is easy to check if an $\tilde{L}$-rational point is in the image of ψ, and also if it lifts to an $\tilde{L}$-rational point of C_6 via $\psi : C_6 \to A_2$, it suffices to compute the principally polarized abelian surfaces over $\tilde{L}$ with good reduction outside S and $\tilde{L}$-endomorphisms by $\mathfrak{o}$. By Theorem 8.3.1, this is done by the algorithm specified in Section 8.2.1 on input $(\mathfrak{o}, L, S)$. $\square$

This completes the proof of Theorem 8.1.1.

Chapter 9

$\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$: potential modularity and unconditional algorithms.

Abstract.

We give a finite-time algorithm that takes as input $(\mathfrak{o}, K, S)$ with $K/\mathbb{Q}$ totally real of odd degree, S a finite set of places of K, and $\mathfrak{o}$ an order in a totally real field, and outputs the abelian K-varieties with $\mathfrak{o}$-multiplication over K with good reduction outside S — in other words, the $\mathfrak{o}_{K,S}$-points of the (canonical integral model of the) Hilbert modular variety $\mathcal{H}_{\mathfrak{o}}$ associated to $\mathfrak{o}$.

It follows that there is a finite-time algorithm that takes as input (C, K), where C/K is a smooth projective (necessarily hyperbolic) curve admitting a nonconstant map defined over an odd-degree totally real field to a Hilbert modular variety, and K is an odd-degree totally real field, and outputs $C(K)$.

Because such curves C abound, this gives an abundance of curves with rational points for which there is a *completely unconditional* algorithm determining their rational points over all odd-degree totally real extensions.

The key point of this chapter is that one can apply known potential modularity theorems in this setup, because one can find a point guaranteed by the Moret-Bailly theorem in finite time.

We note also that one can remove the phrase "odd-degree" above if one also allows the algorithm to either output $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$ (or similarly $C(K)$) or an unconditional disproof of the absolute Hodge conjecture — this arises because it is not yet known unconditionally that *all* parallel weight two Hilbert modular eigencuspforms over even-degree totally real fields are associated to GL_2-type abelian varieties, but this statement does follow (in quantitative form, basically because one can always associate an explicit motive to the symmetric square lift) from the absolute Hodge conjecture.

9.1 Introduction.

In this chapter we apply the key underlying idea of Chapter 8, namely (and roughly) that one can replace the conjectures in the realm of motives used in Chapter 7 by instead appealing to conjectures in the realm of modularity, in order to produce an *unconditional* algorithm determining the rational points on a large class of curves. The key point is that the hypotheses one needs — (potential) modularity of abelian varieties in one direction, and the existence of abelian varieties associated to automorphic representations in the other direction — are available in considerable generality unconditionally.[1]

However, there is a tradeoff. Recall that, in Faltings' original argument [47], one reduces (using Parshin's observation that an old construction of Kodaira of complete curves in M_g can be applied in this context) the problem of rational points on curves to the problem of S-integral points on $\mathcal{A}_g$. Corresponding to this, one could replace the motivic hypotheses of Chapter 7 with suitable modular hypotheses on Galois representations valued in GSp_{2g}, and the ideas of Chapter 7 would

[1] It is not such a surprise a posteriori that the argument uses potential modularity results to establish a consequence of motivic conjectures, since after all it is through modularity (and the knowledge of the cohomology of Shimura varieties) that one can currently prove cases of the Fontaine-Mazur conjecture in the first place.

still work. However of course when $g > 2$ these modular hypotheses are also out of reach, so this is not obviously progress.

So it seems we do not improve our position by replacing motives with automorphic representations. However nothing forces us to use Kodaira's construction in this context — indeed we have already seen in Chapter 8 that finding maps to more restricted Shimura varieties can considerably improve our situation. And the point here is the following: since Galois representations valued in GL_2 (which, in our context, correspond to GL_2-type abelian varieties) are much better understood, why not work inside a Hilbert modular variety $H_{\mathfrak{o}}$ instead of the whole of A_g? So we see the tradeoff: Kodaira's construction produces, for any hyperbolic curve, a nonconstant map to some A_g, whereas we are not sure which curves admit such maps to some Hilbert modular variety $H_{\mathfrak{o}}$. Moreover, since we only understand modularity questions over CM fields (and can only always construct abelian varieties associated to the relevant type of automorphic representations when the field is totally real of odd degree), we need hypotheses about the field of definition of the nonconstant map from the curve to this $H_{\mathfrak{o}}$.

Nonetheless, since Hilbert modular varieties are quasiprojective with very small (indeed, zero-dimensional) boundary, complete curves on them abound. In any case, this explains the source of the hypotheses on our curves in this section.

9.1.1 Main theorem.

We prove the following theorem.

Theorem 9.1.1. *There is a finite-time algorithm that computes*

$$(\mathfrak{o}, K, S) \mapsto \mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S}),$$

where $\mathfrak{o}$ is an order in a totally real field, $K/\mathbb{Q}$ is a totally real field with $[K:\mathbb{Q}]$ odd, S is an explicitly sufficiently large (in terms of K and $\mathfrak{o}$) finite set of places of K, and $\mathcal{H}_{\mathfrak{o}}$ is the canonical $\mathbb{Z}[S^{-1}]$-model of the Hilbert modular variety associated to $\mathfrak{o}$.

Unsurprisingly, essentially the same finite-time algorithm also computes, on input $(\mathfrak{o}, K, S)$ with $\mathfrak{o}$ an order in a CM (instead of just totally real) field $L/\mathbb{Q}$, the abelian varieties A/K with good reduction outside S, with $\dim A = [L:\mathbb{Q}]$, and K-endomorphisms by $\mathfrak{o}$, but we have stated the above theorem for totally real endomorphism fields for convenience.

The case $K=\mathbb{Q}$ of the above was done by von Känel [104] (Murty-Pasten [75] similarly, but independently, dealt with the case $K=\mathbb{Q}$ and $\mathfrak{o}=\mathbb{Z}$) using the resolution of Serre's conjecture over $\mathbb{Q}$ by Khare-Wintenberger [61, 62]. Moreover these works give quite explicit height bounds on the relevant abelian varieties, whereas we have phrased the above theorem in terms of finite-time algorithms for simplicity (and to match the other main theorems of Part II of this thesis). However one could in principle rephrase our argument as giving an explicit height bound on the relevant A/K — see the discussion in Section 9.2.3.1, and note that every other step of our argument is already made explicit below.

We will mostly focus on the odd-degree case below, but we note that in Section 9.5 we explain how to modify the arguments given to prove Theorem 9.1.1 in order to prove the following.

Theorem 9.1.2. *There is a finite-time algorithm that, on input $(\mathfrak{o}, K, S)$, outputs either $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$ or else an unconditional disproof of the absolute Hodge conjecture. Here $\mathfrak{o}$ is an order in a totally real field, $K/\mathbb{Q}$ is a totally real field, S is an explicitly sufficiently large (in terms of K and $\mathfrak{o}$) finite set of places of S, and $\mathcal{H}_{\mathfrak{o}}$ is the canonical $\mathbb{Z}[S^{-1}]$-model of the Hilbert modular variety associated to $\mathfrak{o}$.*

We note that the absolute Hodge conjecture arises because of Blasius's construction [25] of abelian varieties associated to parallel weight 2 Hilbert modular eigencuspforms in cases where the usual construction (Jacquet-Langlands transfer to a quaternion algebra split at only one place at infinity and take the corresponding quotient of the Jacobian of the corresponding Shimura curve) fails.

9.1.2 Main corollary and motivation.

Let us first detail the main corollary and motivation for Theorem 9.1.1, and then discuss the key ideas that go into its proof.

Definition 9.1.3. *Let $K/\mathbb{Q}$ be a totally real field with $[K : \mathbb{Q}]$ odd. A smooth projective hyperbolic curve C/K is called* oddly totally really nice *if and only if there is an order $\mathfrak{o}$ in a totally real field and a finite-to-one map $C \to H_{\mathfrak{o}}$ defined over an odd-degree totally really field, where $H_{\mathfrak{o}}$ is the Hilbert modular variety associated to $\mathfrak{o}$.*

Corollary 9.1.4. *There is a finite-time algorithm that computes*

$$(C, K) \mapsto C(K),$$

where C/K is an oddly totally really nice curve and $K/\mathbb{Q}$ is totally real with $[K : \mathbb{Q}]$ odd.

Assuming Theorem 9.1.1, the proof is immediate: one can find a map by searching; once one has found such a map then the map extends to S-integral models for S explicit and large — e.g., to $\mathcal{C} \to \mathcal{H}_{\mathfrak{o}}$, where $\mathcal{C}$ is the minimal proper regular model of C over $\mathfrak{o}_{K,S}$ (which is smooth for S sufficiently large and thus satisfies the Néron property, though this is not necessary here) and $\mathcal{H}_{\mathfrak{o}}$ is the canonical model of $H_{\mathfrak{o}}$ over $\mathbb{Z}[(\operatorname{Nm} S)^{-1}]$ — and then one has reduced to computing the S-integral points on the canonical integral model of a Hilbert modular variety.

Note that such curves abound because Hilbert modular varieties can be compactified into projective varieties by points, so that general curve sections and their (e.g. ramified) covers give examples of such curves. One has so much freedom that one can prescribe lots of points to lie on these curves as well.[2]

We note here that not only is there no algorithm to find rational points on smooth projective hyperbolic curves over number fields that provably always terminates (though recall that Chapter 7 gives an algorithm with unconditionally correct output and which terminates conditional on standard conjectures) — even over $\mathbb{Q}$, even when the Chabauty hypothesis holds, etc. — there is *also* no unconditional algorithm to even determine the ranks of the Jacobians of such curves.[3]

Counterintuitively, we produce here smooth projective hyperbolic curves C/K over odd-degree totally real fields K for which, for every odd-degree totally real extension L/K, one can unconditionally determine $C(L)$ in finite time, but for which no unconditional method is known to determine $\operatorname{rank}(\operatorname{Jac} C)(L)$. Indeed, for very large L/K, one expects the algebraic rank of $\operatorname{Jac} C$ over such L to also be large (or even $\operatorname{Jac} C$ might not be GL_2-type) since e.g. we may specify many points to lie in $C(K)$, so that the usual unconditional method of determining ranks of abelian varieties — namely, constructing suitable Heegner points and using a Gross-Zagier-type formula/Kolyvagin-type Euler systems argument — must fail.

9.1.3 Main ideas.

Let $\mathfrak{o}$ be an order in a totally real field F and λ a prime of $\mathfrak{o}$ with residue characteristic ℓ, which we will assume to be prime to all data in sight. Let $K/\mathbb{Q}$ be a totally real field with $[K : \mathbb{Q}]$ odd. Let S be a finite set of places of K.

[2] This comment is necessary because Shimura curves associated to indefinite quaternion algebras over $\mathbb{Q}$ also lie inside these Hilbert modular varieties, but they lack real points. In other words there is already an extremely fast algorithm for determining their points over totally real fields.

[3] Assuming finiteness of the Tate-Shafarevich group the evident day/night descent/search procedure works, as first pointed out by Tate (see his comments after Conjecture 4.1 in [100]) when conjecturing the finiteness of the Tate-Shafarevich group.

As usual, we will completely ignore stack-theoretic issues and just discuss the question of determining the $[F : \mathbb{Q}]$-dimensional abelian K-varieties with $\mathfrak{o}$-multiplication defined over K and good reduction outside S, since this is evidently sufficient.

If we knew that all the abelian varieties in question were modular (in the sense that their L-functions match the L-functions of parallel weight 2 Hilbert modular eigencuspforms over K), then we would immediately be done via an essentially one-line argument: all K-simple[4] such abelian varieties would be quotients of the Jacobian of a single Shimura curve[5], and thus Masser-Wüstholz would bound all their heights[6]. This is essentially the argument used by von Känel [104] and Murty-Pasten [75] to find $\mathbb{Z}[S^{-1}]$-points on these Hilbert modular varieties (i.e. to treat the case $K = \mathbb{Q}$), using the resolution of Serre's conjecture by Khare-Wintenberger [61, 62]: there, the corresponding Shimura curves are modular curves, and one is dealing with classical weight 2 modular forms.

We will instead use potential modularity theorems. After all, Serre's conjecture is not available in this generality, and moreover the available modularity lifting theorems are weaker. Naturally it is then the inclusion of the word 'potential' that presents the major obstacle to be overcome. Let us now describe the technique. We will first describe it in significantly less streamlined form, in order to explain a trick ("moving in a compatible family") which we are able to avoid using here by using our Lemma 9.3.3 (an explicit form of a large-residual-image result of Dimitrov [44])

[4] Non-simple such abelian varieties are all of the form $B^{\times n}$ for a B/K with $\mathfrak{o}'$-multiplication, where $\mathfrak{o}'$ is an order in a subfield $F' \subseteq F$ — see Lemma 9.3.1.

[5] — namely, corresponding to a quaternion algebra over K unramified at exactly one infinite place and at all finite places, and with level structure explicitly depending on S.

[6] Let A/K be an abelian variety with K-quotient B/K. By Poincaré complete reducibility, there is a C/K with $A \sim_K B \times C$. By Masser-Wüstholz and Bost's lower bound for the Faltings height, it follows that:

$$h(B) \leq h(B \times C) + O_{\dim A}(1) \ll_{h(A), \dim A} 1.$$

rather than Lemma 7.1.3 of [3] (which we will henceforth call the "ten-author paper"), but which may prove useful in other adaptations.

By Faltings' Lemma, there is an explicitly computable finite set T_λ of places of K that is disjoint from S and such that trace functions[7] of representations $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_\lambda)$ that are unramified outside S are determined by their values at $\mathrm{Frob}_\mathfrak{p}$ for $\mathfrak{p} \in T_\lambda$. In fact for ease of notation we may replace the order $\mathfrak{o}$ by the maximal order $\mathfrak{o}_F$ since this simply enlarges the set of Galois representations in question. Because of Lemma 9.3.1 we may act like all of our abelian varieties are K-simple for the sake of this discussion.

The Galois representations of the abelian varieties we are interested in are all pure of weight 1 and have F-rational traces. Thus their traces at $\mathrm{Frob}_\mathfrak{p}$ for $\mathfrak{p} \in T_\lambda$, say $a_\mathfrak{p} \in \mathfrak{o}$, satisfy

$$|a_\mathfrak{p}|_v \leq 2\sqrt{\mathrm{Nm}\,\mathfrak{p}}$$

for all places $v|\infty$ of F. Hence the tuples $(a_\mathfrak{p})_{\mathfrak{p}\in T_\lambda}$ lie in an explicit finite set, say Λ.

Our goal is to compute a finite set $\mathcal{F}$ of odd-degree totally real extensions L/K that has the property that any $A \in \mathcal{H}_\mathfrak{o}(\mathfrak{o}_{K,S})$ is modular over some $L \in \mathcal{F}$. Then we may repeat the above argument (transfer to a quaternion algebra, look at the Jacobian of the corresponding Shimura curve with level structure, and use Masser-Wüstholz to deduce a height bound) for each $L \in \mathcal{F}$ to deduce a bound on the heights of points in $\mathcal{H}_\mathfrak{o}(\mathfrak{o}_{K,S})$, after which enumeration of bounded-height points suffices. To do this we first compute which of the $\vec{a} \in \Lambda$ are congruent to $(t(\mathrm{Frob}_\mathfrak{p}))_{\mathfrak{p}\in T_\lambda}$, with t running over the trace functions of the finitely many possi-

[7] By the *trace function* of $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_n(R)$ we simply mean the class function $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to R$ given by $\mathrm{tr} \circ \rho$. We note that Faltings' Lemma holds not only for $R = \mathbb{Z}_\ell$ or $\mathfrak{o}_\lambda$ or similar rings, but also for $R = \mathbb{Z}/\ell^n$ or $\mathfrak{o}/\lambda^n$, in the sense that the values of a trace function of a representation into $\mathrm{GL}_n(R)$ at primes in T_λ determines it on all of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ — see Lemma 7.2.1 of Chapter 7. Of course when $R = \mathbb{Z}/\ell^n$ or $\mathfrak{o}/\lambda^n$ Brauer-Nesbitt is not applicable and so for consistency reasons we avoid talking of semisimple representations etc. even when $R = \mathfrak{o}_\lambda$.

ble mod-λ representations (computed using Hermite-Minkowski). We throw out those that do not "lift" to a mod-λ representation.

Then we work with the mod-λ representations $\bar{\rho}$.

- If $\bar{\rho}$ satisfies the Taylor-Wiles condition (namely that $\bar{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K(\zeta_\ell))} \otimes_{\mathfrak{o}/\lambda} \overline{\mathbb{F}}_\ell$ is irreducible), then we follow Snowden (thus ultimately Taylor) and compute a totally real point on a twist of a Hilbert modular variety.[8] The computation is guaranteed to terminate by a theorem of Moret-Bailly.

  Using that point, we follow Snowden[9] to produce a totally real Galois extension L/K over which any of our abelian varieties in question with mod-λ representation $\bar{\rho}$ become modular over L. Let H be a 2-Sylow subgroup of $\mathrm{Gal}(L/K)$. Then L/L^H is a solvable extension, and $L^H/\mathbb{Q}$ is a totally real field of odd degree. Therefore by solvable descent for GL_2, all such abelian varieties with mod-λ representation $\bar{\rho}$ are also modular over L^H. Thus we add L^H to $\mathcal{F}$ and move to the next mod-λ representation.

- If $\bar{\rho}$ does not satisfy the Taylor-Wiles condition, we "move" in a compatible family until it does. Ultimately by a theorem of Larsen (we follow the argument in the ten-author paper [3]), it essentially follows that either $\vec{a}$ is not the tuple of traces at T_λ of a compatible family of representations $(\rho_{\mathfrak{q}} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{\mathfrak{q}}))_{\mathfrak{q}}$ that is unramified outside S, or there is a large prime $\mathfrak{q} \subseteq \mathfrak{o}$ for which there is a mod-$\mathfrak{q}$ representation $\bar{\rho}'$ with $\mathrm{tr}(\bar{\rho}'(\mathrm{Frob}_{\mathfrak{p}})) \equiv a_{\mathfrak{p}} \pmod{\mathfrak{q}}$ for all $\mathfrak{p} \in T_\lambda$ and satisfying the Taylor-Wiles condition.

[8] In fact we have to do this for a finite set of such varieties, one for each definite type function (in the sense of Snowden's [95]) on primes above ℓ, but this is immaterial.

[9] Note that if we had instead directly used Corollary 7.1.11 in the ten-author paper [3], along with the note that the extension produced may be taken to be unramified above infinite primes (thus in the case of a totally real field one gets a totally real extension), we would be able to avoid the discussion of the Taylor-Wiles condition by just producing compatible families. We use Snowden's [95] because Lemma 9.3.3 will allow us to work at one prime only, rather than with a compatible family. Moreover, we learned the idea of using solvable descent to produce an *odd-degree* extension from his [95].

However there is a serious obstruction to be overcome: informally, we "know" the what the traces must be of such a compatible family at T_λ, namely we know $(a_\mathfrak{p})_{\mathfrak{p}\in T_\lambda}$. But to run the above arguments with a different prime $\mathfrak{q}$ we must know the traces at a set $T_\mathfrak{q}$, which may well be quite different from T_λ — in other words, knowing only the traces at T_λ a priori will *not* determine a $\mathfrak{q}$-adic representation for other primes $\mathfrak{q}$.

But of course there is a trick: all of these representations must be pure of weight 1 if they are to arise from an abelian variety, and we already "know" all Frobenius traces modulo λ^N for huge N. So, if we are trying to compute $a_\mathfrak{p}$ for $\mathfrak{p} \in T_\mathfrak{q}$, for example, we may first take N very large in terms of $\mathfrak{p}$ and compute $a_\mathfrak{p} \bmod \lambda^N$ by using a mod-λ^N representation ρ with correct Frobenius traces at T_λ — if one does not exist, then we throw out the tuple $\vec{a}$. Then, having computed what $a_\mathfrak{p}$ must be modulo λ^N, we observe that we also have the inequalities $|a_\mathfrak{p}|_v \leq 2\sqrt{\mathrm{Nm}\,\mathfrak{p}}$ for all $v|\infty$. For N large enough there is at most one element of $\mathfrak{o}$ satisfying these constraints. If there is none then we again throw out $\vec{a}$ — otherwise, the unique element is our $a_\mathfrak{p}$. In this way we compute a tuple of Frobenius traces at $T_\mathfrak{q}$ for other primes $\mathfrak{q}$.

Then, for larger and larger $\mathfrak{q}$ and N, we use this tuple of traces at each $T_\mathfrak{q}$ to check if there is a mod-$\mathfrak{q}^N$ representation of the desired type with said traces. If there is none, we again throw out $\vec{a}$. If there is, we check the Taylor-Wiles condition for the mod-$\mathfrak{q}$ representation — if it is satisfied we return to the previous case. Otherwise we continue the checks for larger primes.

This step must eventually terminate because a tuple $\vec{a}$ that survives these checks forever (and which does not arise from a CM abelian variety, which is easily checked) must fit into a compatible family to which Larsen's theorem applies.

Thus we compute such a finite set of extensions $\mathcal{F}$, and then take the maximum of the height bounds given by the Masser-Wüstholz argument applied to the Jacobians of the corresponding Shimura curves. This bounds the Faltings heights of the points in $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$, and enumeration of bounded height points in the Baily-Borel/Satake compactification (or else decomposition of the Jacobians of the produced Shimura curves) finishes the algorithm.

Now we comment on the simplification. In fact one can do much better than use the theorem of Larsen (i.e. ultimately Lemma 7.1.3 in the ten-author paper [3]) — specifically, one knows something far stronger than the fact that the mod-$\mathfrak{q}$ Galois representation associated to a relevant (non-CM) abelian variety has large image at a density 1 set of primes $\mathfrak{q} \subseteq \mathfrak{o}_F$: one knows that this is the case for all but finitely many $\mathfrak{q}$, by a theorem[10] of Dimitrov. It is not difficult to make said theorem quantitative (see Lemma 9.3.3 — we simply change one step in his endgame), after which point we may simply run the argument outlined above, except with $\operatorname{Nm}\lambda$ (explicitly) sufficiently large in terms of $\mathfrak{o}, K$, and S, and without a search through primes $\mathfrak{q}$, since we may insist that the Taylor-Wiles condition holds (and indeed that the residual image contains $\mathrm{SL}_2(\mathbb{F}_\ell)$) at our originally chosen prime $\lambda|(\ell)$.

9.2 The algorithm and its subroutines.

Let us now precisely specify the algorithm alluded to above. In fact a slight modification of the algorithm we give solves the slight generalization of the problem where we allow endomorphisms by an order in a CM field (rather than just a totally real field), but for simplicity we will just discuss the case of totally real endomorphism field.

[10] This in fact easily generalizes to compatible families of two-dimensional Galois representations with given Hodge-Tate weights and explicitly bounded Frobenius traces — see Lemma 9.3.3.

9.2.1 IntegralPointsOnHilbertModularVarieties($\mathfrak{o}, K, S$):

Input: $\mathfrak{o}, K, S$, with $\mathfrak{o}$ an order[11] in a totally real number field $\operatorname{Frac}\mathfrak{o} =: F/\mathbb{Q}$, $K/\mathbb{Q}$ a totally real number field with $[K:\mathbb{Q}]$ odd, and S a finite set of places of K.

Output: $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$.

Algorithm:

1. Let $F := \operatorname{Frac}\mathfrak{o}$. Let $g := [F:\mathbb{Q}]$. Let $\mathfrak{m}_K := \prod_{\mathfrak{p}\in S}\mathfrak{p}^{10^{10g}\cdot[K:\mathbb{Q}]}$.

2. $\mathcal{F} :=$ output of RelevantExtensions($F, K, \operatorname{Nm}\mathfrak{m}_K$).

3. Let, for $L \in \mathcal{F}$, $v_L|\infty$ be an infinite place of L. Let $\mathfrak{m}_L := \prod_{\mathfrak{p}\subseteq\mathfrak{o}_L:(\mathfrak{p},\operatorname{Nm}S)\neq 1}\mathfrak{p}^{10^{10g}\cdot[L:\mathbb{Q}]}$, a product over primes of L dividing $\operatorname{Nm}\mathfrak{q}$ for some $\mathfrak{q}\in S$. Let L'/L be[12] the compositum of all Galois extensions of L of degree $\leq 10^{10\cdot g^{10^{10}}}$ which are unramified outside primes $\mathfrak{p}\subseteq\mathfrak{o}_L$ of L with $\mathfrak{p}|\operatorname{Nm}\mathfrak{q}$ and $\mathfrak{q}\in S$.

4. Let, for $L\in\mathcal{F}$, $C_{v_L}(\mathfrak{m}_L)$ be the Shimura curve corresponding to the quaternion algebra over L split at all finite places and at v_L, and ramified at all other infinite places of L, along with full level-$\mathfrak{m}_L$ structure.

[11] In fact it suffices to take $\mathfrak{o}$ to be a maximal order, because all abelian varieties A/K with $\mathfrak{o}$-multiplication over K are K-isogenous to the Serre tensor product $A\otimes_{\mathfrak{o}}\mathfrak{o}_F$, and so determining the abelian varieties with $\mathfrak{o}_F$-multiplication suffices to determine those with $\mathfrak{o}$-multiplication by Masser-Wüstholz. Nonetheless we will deal with all orders $\mathfrak{o}$ without using this reduction.

[12] We remark that this L'/L enters only to indicate how to get around a stack-theoretic issue: if one uses coarse moduli spaces instead, the field of moduli of an $A\in\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$ (namely K) may a priori not be the field of definition (which will be a subfield of "$K(A[210])$" $\subseteq L'$). Thus in Step 6 we use the Masser-Wüstholz bound over L', since such an A/L' will be an L'-isogeny factor of $\operatorname{Jac}C_{v_L}(\mathfrak{m}_L)^{\times[F:\mathbb{Q}]}/L'$. Having said this, we will completely ignore the distinction between the field of moduli and the field of definition (i.e. act like $L'=L$) for the rest of the chapter, since after all we are not using coarse moduli spaces.

5. Let[13]

$$H := \max_{L \in \mathcal{F}} \Big[\text{MasserWüstholz}(\operatorname{Jac} C_{v_L}(\mathfrak{m}_L)^{\times [F:\mathbb{Q}]}, L') + \text{Bost}([F:\mathbb{Q}] \cdot \dim \operatorname{Jac} C_{v_L}(\mathfrak{m}_L)) \Big].$$

6. Output the points in $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$ of height $\leq H$.

9.2.1.1 Explanation in words.

The true work is passed to the RelevantExtensions routine — once one has produced a finite set of odd-degree totally real extensions over which any relevant abelian variety A/K is modular, one knows that A is the n-th power (see Lemma 9.3.1) of a quotient of the Jacobian of a Shimura curve with level structure, via Jacquet-Langlands transfer. (Evidently $n \leq \dim A$.) The level is crudely bounded using Brumer-Kramer [34] and the fact (see e.g. Carayol [35], specifically the equality of ε-factors in his Section 0.5) that the conductors of the λ-adic representations associated to a Hilbert modular eigencuspform agree with the form's level. Thus the height bound H is computed as the maximal height of an isogeny factor of the $(\dim A)$-th power of such a Jacobian, which is bounded via Masser-Wüstholz and Bost's lower bound on the Faltings height.

It is worth noting that implicitly we are using the Baily-Borel (here Satake) compactification of $H_{\mathfrak{o}}$, and a comparison of heights between the Faltings height and

[13] Here MasserWüstholz$(A, K) \in \mathbb{R}^+$ is an explicitly computed constant such that, for all B/K abelian K-varieties that are K-isogenous to A/K,

$$h(B) \leq \text{MasserWüstholz}(A, K).$$

A formula can basically be read off from Masser-Wüstholz [68], though we use Gaudron-Rémond [50] instead — see Theorem 7.2.2 of Chapter 7. Rather than using said formula here, we preferred to be clear about the provenance of the constants. Similarly, Bost$(g) \in \mathbb{R}^+$ is an explicitly computed constant such that, for all abelian varieties $A/\overline{\mathbb{Q}}$ of dimension $\dim A \leq g$, one has

$$h(A) \geq -\text{Bost}(g).$$

A formula can be read off from Pazuki [78] or Gaudron-Rémond [51], for example. See Theorem 7.2.3 of Chapter 7 for Gaudron-Rémond's.

the naïve height on the ambient projective space, to enumerate points of bounded height in $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$.

9.2.2 RelevantExtensions(E, K, s):

Input: E, K, s, with $E/\mathbb{Q}$ totally real, $K/\mathbb{Q}$ totally real of odd degree, and $s \in \mathbb{Z}^+$.
Output: $\mathcal{F}$, a finite set of odd-degree totally real extensions L/K such that, for all primes λ of E with $\operatorname{Nm}\lambda$ prime to s and representations $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\lambda})$ arising from abelian varieties A/K with good reduction at primes not dividing s and admitting an embedding $E \hookrightarrow \mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Q}$, there is an $L \in \mathcal{F}$ such that $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}$ is the λ-adic representation of a parallel weight 2 cuspidal Hilbert modular eigencuspform over L (with coefficient ring a subring of $\mathfrak{o}_E$).
Algorithm:

1. Let $\lambda_0 \subseteq \mathfrak{o}_E$ be a prime of E with $\operatorname{Nm}\lambda_0$ prime to s. Let $\lambda \subseteq \mathfrak{o}_E$ be a prime of E with $\operatorname{Nm}\lambda$ prime to s and satisfying $\operatorname{Nm}\lambda \geq C_{E,K,s,\lambda_0}$, in the notation of Lemma 9.3.3. Let ℓ be the prime of $\mathbb{Z}$ with $\lambda|(\ell)$. Let $S := \{\mathfrak{p}|(s\cdot\ell) : \mathfrak{p} \subseteq \mathfrak{o}_K\}$.

2. Let Ψ_λ be the set of finite-order characters

$$\psi : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathfrak{o}_{E,\lambda}^\times$$

of K that are unramified outside S and primes above ℓ.

3. Let

$$R_\lambda := \left\{ (\rho,\psi) \;\middle|\; \begin{array}{c} \rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_E/\lambda) \text{ odd and unramified outside } S \cup \{\mathfrak{q}|(\ell)\}, \\ \psi \in \Psi_\lambda, \mathrm{SL}_2(\mathbb{F}_\ell) \subseteq \rho(\mathrm{Gal}(\overline{\mathbb{Q}}/K)), \det\rho \equiv \psi\cdot\chi_\ell \pmod{\lambda} \end{array} \right\},$$

where χ_ℓ is the ℓ-adic cyclotomic character.

4. Let, for each $(\rho, \psi) \in R_\lambda$,

$$\tilde{\mathcal{F}}_{(\rho,\psi)} := \text{output of TaylorMoretBaillyExtensions}(\rho, \psi, K, E, \lambda).$$

   Let

$$\mathcal{F}_{(\rho,\psi)} := \left\{ \tilde{L}_{(\rho,\psi)}^{H_{(\rho,\psi)}} : \tilde{L}_{(\rho,\psi)} \in \tilde{\mathcal{F}}_{(\rho,\psi)}, H_{(\rho,\psi)} \subseteq \mathrm{Gal}(\tilde{L}_{(\rho,\psi)}/K) \text{ a 2-Sylow subgroup} \right\}.$$

5. Let $\mathcal{F} := \{K\} \cup \bigcup_{(\rho,\psi)\in R_\lambda} \mathcal{F}_{(\rho,\psi)}$.

6. Output $\mathcal{F}$.

9.2.2.1 Explanation in words.

We repeat that this algorithm is considerably streamlined thanks to Lemma 9.3.3 — had we used a weaker result guaranteeing the same conclusion at a density 1 set of primes (e.g. Lemma $7.1.3$ of the ten-author paper [3]), we would have had to set up a somewhat delicate search through larger and larger primes λ. See Section 9.1.3 for a more precise discussion.

Our goal is to find a finite set of extensions $\mathcal{F}$ such that, for all abelian varieties A/K with good reduction at primes not dividing s and admitting an embedding $E \hookrightarrow \mathrm{End}_K^0(A)$, there is an $L \in \mathcal{F}$ and a parallel weight 2 Hilbert modular eigencuspform f over L with coefficient ring a subring of $\mathfrak{o}_E$ and for which $(\rho_{A,\lambda} \otimes_{\mathfrak{o}_{E,\lambda}} E_\lambda)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)} \cong \rho_{f,\lambda} \otimes_{\mathfrak{o}_{E,\lambda}} E_\lambda$.

So let A/K be such an abelian variety. By always having $K \in \mathcal{F}$, we are done if A admits sufficiently many complex multiplications over a quadratic extension of K. So without loss of generality A/K is not potentially CM.

Now A admits a compatible family of representations $\rho_{A,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\lambda})$. These representations have the following properties. Let $\lambda \subseteq \mathfrak{o}_E$ be a prime of E with $\mathrm{Nm}\,\lambda$ prime to s. Let $\ell \in \mathbb{Z}^+$ be the prime of $\mathbb{Z}$ such that $\lambda|(\ell)$. Then $\rho_{A,\lambda} \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell$ is irreducible, odd, unramified at primes not dividing $s \cdot \ell$, has determinant of the form $\det \rho_{A,\lambda} = (\text{finite order}) \cdot \chi_\ell$, and is not the induction of a character from a quadratic extension of K. Moreover it has Hodge-Tate weights $\{0, -1\}$ under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_\ell$, and is pure of weight 1.

So we see that, taking the $\mathfrak{p}$ of Lemma 9.3.3 to be our λ, and the λ of Lemma 9.3.3 to be our λ_0, it follows that $\mathrm{SL}_2(\mathbb{F}_\ell) \subseteq \overline{\rho}_{A,\lambda}(\mathrm{Gal}(\overline{\mathbb{Q}}/K))$. In other words, the Taylor-Wiles hypothesis is satisfied for $\rho_{A,\lambda}$. This is the key point.

Thus writing $\psi_{A,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathfrak{o}_{E,\lambda}^\times$ for the finite-order character for which $\psi_{A,\lambda} := (\det \rho_{A,\lambda}) \cdot \chi_\ell^{-1}$, it follows that $\psi_{A,\lambda} \in \Psi_\lambda$ and indeed $(\overline{\rho}_{A,\lambda}, \psi_{A,\lambda}) \in R_\lambda$.

Then we simply pass to the TaylorMoretBaillyExtensions algorithm, which produces a Galois extension L/K linearly disjoint from the fixed field of $\ker \overline{\rho}_{A,\lambda}$ (thus the Taylor-Wiles condition still holds, so that one can use available modularity lifting theorems) for which $(\rho_{A,\lambda} \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}$ arises from a parallel weight 2 Hilbert modular eigencuspform over L.

We then use solvable descent for GL_2 to produce a Hilbert modular eigencuspform over L^H, with $H \subseteq \mathrm{Gal}(L/K)$ a 2-Sylow subgroup. Note that $[L^H : K]$ is odd, of course. This completes the discussion.

9.2.3 TaylorMoretBaillyExtensions($\bar{\rho}, \psi, K, E, \mathfrak{q}$):

We closely follow Snowden's [95].

Input: $\bar{\rho}, \psi, K, E, \mathfrak{q}$, with $\mathfrak{q}|(p)$.

Output: $\mathcal{F}$, a finite set of totally real Galois extensions L/K for which, for all odd, weight two[14], finitely ramified $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\mathfrak{q}})$ such that $\rho \bmod \mathfrak{q} \cong \bar{\rho}$ and $\det \rho = \psi \cdot \chi_p$ with χ_p the p-adic cyclotomic character, one has that there is an

[14] Here we follow Snowden's terminology in [95].

$L \in \mathcal{F}$ for which $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}$ is the $\mathfrak{q}$-adic representation of a parallel weight 2 Hilbert modular eigencuspform.

Algorithm: We follow the proof of Theorem 5.1.2 in Snowden's [95] — in particular we use his notation and terminology.

1. Let $\ell \neq p$ with $\ell > 5$ and construct[15] M_ℓ/K and $\rho_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ as in Proposition 5.4.1 of Snowden's [95]. Let $\psi_\ell := \det \rho_\ell \cdot \chi_\ell^{-1}$ be the corresponding finite-order character, where χ_ℓ is the ℓ-adic cyclotomic character.

2. Let M/K be the Galois closure of M_ℓ/K.

3. Let $F/\mathbb{Q}$ be CM and $\mathfrak{p}|(p), \lambda|(\ell)$ be primes of F such that $\operatorname{im}\bar{\rho} \subseteq \mathrm{GL}_2(\mathfrak{o}_F/\mathfrak{p})$ and $\operatorname{im}\bar{\rho}_\ell \subseteq \mathrm{GL}_2(\mathfrak{o}_F/\lambda)$.

4. Construct[16], as in the proof of Corollary 4.2.2 of Snowden's [95], K_1/K, a totally real finite extension disjoint from M, and $\tilde{\psi}$ and $\tilde{\psi}_\ell$ finite-order characters of $\mathrm{Gal}(\overline{\mathbb{Q}}/K_1)$ for which $\tilde{\psi}^2 = \psi|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K_1)}$ and $\tilde{\psi}_\ell^2 = \psi_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K_1)}$.

5. Let $\bar{\rho}_\mathfrak{p} := \tilde{\psi}^{-1} \cdot \bar{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K_1)}, \bar{\rho}_\lambda := \tilde{\psi}_\ell^{-1} \cdot \bar{\rho}_\ell|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K_1)}$. Let t_λ be the type[17] function of ρ_ℓ.

6. Write $\Sigma_p := \{\mathfrak{m} \subseteq \mathfrak{o}_K : \mathfrak{m}|(p)\}$. For each definite[18] type function t on Σ_p, define the Skolem datum $(X_t, \Sigma, (L_v)_{v\in\Sigma}, (\Omega_v)_{v\in\Sigma})$ exactly as in the proof of Proposition 5.3.1 in Snowden's [95]. Find[19] a Galois extension $K_{2,t}/K_1$ that is linearly disjoint from M and a point $x \in X_t(K_{2,t})$. Let K_t'/K be the Galois closure of $K_{2,t}/K$ over K.

[15] Snowden's proof of Proposition 5.4.1 in [95] gives an explicit construction through class field theory.

[16] Again, Snowden's proof already gives an explicit construction.

[17] See Section 2 of Snowden's [95] for precise definitions.

[18] Again, see Section 2 of Snowden's [95] for a precise definition.

[19] This can be done by simply searching points of larger and larger height (note that X_t is a (twist of a) Hilbert modular variety, and is thus projective via e.g. the Baily-Borel/Satake compactification), or else, if one prefers, by making explicit the proof of Moret-Bailly's theorem, e.g. by following Rumely and ultimately reducing to a Fekete-Szegő-like statement.

7. Let

$$\mathcal{F} := \{K'_t/K : t \text{ a definite type function on } \Sigma_p\}.$$

8. Output $\mathcal{F}$.

9.2.3.1 Explanation in words.

This algorithm amounts to copying Snowden's proofs of Theorems 5.1.1 and 5.1.2 in [95] over into algorithmic form. The point is simply that the extensions implicit in the theorems are computable in finite time — the only difficulty lies in finding a suitable point on the constructed variety X_t, and the theorem of Moret-Bailly that is invoked can be rephrased as saying that a simple search for points will eventually terminate (because there is one).

In fact it is worth commenting here that Moret-Bailly's proof of Rumely's theorem (see Section 4 of [71]) is constructive, though compactness arguments are used in later treatments. Moreover, in Rumely's original formulation and proof using Cantor capacities [85, 86], the proof is constructive if complicated, since e.g. the proof of Fekete-Szegő is algorithmic: one first approximates a probability distribution sufficiently well by a sum of delta masses, and then Fekete-Szegő give a very explicit procedure to complete the proof. Thus a blind search is not the only way to proceed.

9.3 Preliminary lemmas.

Before the proof of correctness and termination of the algorithm specified in Section 9.2.1, let us collect some preliminary lemmas.

So as not to repeat ourselves, we first note that we will again make use of the standard results stated in Section 7.2 of Chapter 7.

9.3.1 Decomposition of GL_2-type abelian varieties over the reals.

The following lemma arises because the abelian varieties A/K we search for may not necessarily be K-simple, and modularity only implies that a K-simple such abelian variety is a quotient of the Jacobian of a suitable Shimura curve. The lemma shows that this is not an issue because we produce a d-th root as such a quotient for some $d \leq \dim A$, so we can (and do) just replace the Jacobian with its $(\dim A)$-th power in the Masser-Wüstholz argument.

Lemma 9.3.1. *Let $K/\mathbb{Q}$ be a number field with a real place $K \hookrightarrow \mathbb{R}$. Let $E/\mathbb{Q}$ be a totally real field. Let A/K be an abelian variety of dimension $\dim A = [E : \mathbb{Q}]$ admitting a map $E \hookrightarrow \mathrm{End}_K^0(A) := \mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Q}$. Then: there is a subfield $F \subseteq E$ and a K-simple abelian variety B/K of dimension $\dim B = [F : \mathbb{Q}]$ admitting a map $F \hookrightarrow \mathrm{End}_K^0(B)$ such that $A \sim_K B^{\times \frac{\dim A}{\dim B}}$. In fact one also has that $F \simeq \mathrm{End}_K^0(B)$.*

We note that, if we instead took $E/\mathbb{Q}$ to be imaginary CM in the above statement, then the same conclusion would hold, except we would also have to allow[20] the possibility that $\mathrm{End}_K^0(B)$ is a quaternion algebra over its centre $F \subseteq E$, a CM field of degree $[F : \mathbb{Q}] = \frac{\dim B}{2}$. In this case B/K is again GL_2-type over K, by Lemma 1.3.7.1 (attributed to Tate) of Chai-Conrad-Oort [37], but a priori not by a subfield of E, but rather possibly by a quadratic extension of a subfield of E. This is in fact not an issue, but in order not to complicate the algorithm further we have restricted ourselves to the case of totally real endomorphism fields.

The proof is also routine (see e.g. Ribet [83] — unfortunately he uses that his abelian varieties are defined over $\mathbb{Q}$), but let us isolate one part of the argument where we use that $K/\mathbb{Q}$ has a real place for clarity. (We also use that K has a real place in citing Shimura's [94] to exclude real points on his canonical models of his varieties, but the following is used repeatedly.)

[20] We are in fact not sure whether this case can arise — nonetheless it causes no difficulty in the algorithm if it does.

Lemma 9.3.2. *Let $K/\mathbb{Q}$ be a number field with a real place $K \hookrightarrow \mathbb{R}$. Let $L/\mathbb{Q}$ be a CM field. Let $\Phi \subseteq \mathrm{Hom}_{\mathbb{Q}\text{-}alg.}(L, \mathbb{C})$ be a CM type of L. Let A/K be an abelian variety admitting a map $L \hookrightarrow \mathrm{End}_K^0(A)$. Write*

$$\mathrm{Lie}(A/\mathbb{C}) \cong \bigoplus_{\sigma \in \Phi} \sigma^{a_\sigma} \oplus \overline{\sigma}^{b_\sigma}$$

as complex representations of L — thus $a_\sigma, b_\sigma \in \mathbb{N}$. Then: $a_\sigma = b_\sigma$ for all $\sigma \in \Phi$.

Proof of Lemma 9.3.2. L preserves $\mathrm{Lie}(A/K)$ by hypothesis, so that the traces of $L \curvearrowright \mathrm{Lie}(A/\mathbb{C})$ all lie in K. For each $\sigma \in \Phi$, use Minkowski to produce an $x \in \mathfrak{o}_L$ with $\sigma(x)$ large and almost purely imaginary, and $\tau(x)$ tiny for every other $\sigma \neq \tau \in \Phi$. Since $\sum_{\tau \in \Phi} a_\tau \cdot \tau(x) + b_\tau \cdot \overline{\tau(x)} \in K \hookrightarrow \mathbb{R}$, it follows that $a_\sigma = b_\sigma$. □

These a_σ, b_σ are r_ν, s_ν in Shimura's notation (see e.g. Section 4 of Shimura's [93]).

Now we turn to proving Lemma 9.3.1.

Proof of Lemma 9.3.1. Write $A \sim_K \prod_{i=1}^m B_i^{\times n_i}$ with B_i/K K-simple and pairwise non-K-isogenous. If $m > 1$ then, because $E \hookrightarrow \mathrm{End}_K^0(B_i^{\times n_i})$ for all i (since it is a field and the identity map is in the image), choosing i for which $n_i \dim B_i$ is minimal shows that some $B_i^{\times n_i}/K$ admits sufficiently many complex multiplications over K (see e.g. Chapter 1 of Chai-Conrad-Oort [37], specifically its Theorems $1.3.1.1$ and $1.3.4$). But this is impossible by Lemma 9.3.2 (in this case $a_\sigma + b_\sigma = 1$ for all $\sigma \in \Phi$). Hence $m = 1$ and thus A/K is a power, i.e. $A \sim_K B^{\times n}$.

Let $D := \mathrm{End}_K^0(B)$. We claim that D is a subfield of E of degree $\dim B$. Let C be the commutant of E in $\mathrm{End}_K^0(A)$. Again by the same CM argument as above we see that C is a division algebra (all its nonzero elements are isogenies, else either their kernels or images would be abelian varieties for which the above CM argument would apply) with centre E (any larger and the above CM argument would apply to A). Let $\tilde{d}$ be its index over E.

Let F be the centre of D. By the Albert classification [1,2] it follows that F is imaginary CM or totally real. Let $F^+ \subseteq F$ denote the maximal totally real subfield of F. Let d be the index of D over F and $d^+ := d \cdot [F : F^+]$. Again by the Albert classification [1,2], it follows that if $F = F^+$ then $d \leq 2$. Note also that, by Lemma $1.3.7.1$ (attributed to Tate) in Chai-Conrad-Oort [37] and the above CM argument, it follows that $d \cdot [F : \mathbb{Q}] < 2 \dim B$, since the former is the degree of a maximal commutative subfield of D.

Tensoring (over $\mathbb{Q}$) the inclusion $C \subseteq M_n(D)$ up to $\mathbb{C}$ one finds the inequality

$$\tilde{d} \cdot [E : \mathbb{Q}] \leq nd^+ \cdot [F^+ : \mathbb{Q}] = nd \cdot [F : \mathbb{Q}].$$

Because $[E : \mathbb{Q}] = n \dim B$, we deduce that

$$\tilde{d} \cdot \dim B \leq d \cdot [F : \mathbb{Q}] < 2 \dim B.$$

Thus $\tilde{d} = 1$, i.e. $C = E$. But F is the centre of D, so that $F \subseteq C = E$. Thus by hypothesis it follows that $F = F^+$ is totally real, and so $d \leq 2$ by the Albert classification [1,2]. Also, again by the Albert classification [1,2], $d \cdot [F : \mathbb{Q}] \mid \dim B$. Because we already have the inequality $\dim B \leq d \cdot [F : \mathbb{Q}]$ it follows that $[F : \mathbb{Q}] = \frac{\dim B}{d}$.

We thus need only rule out the cases where $d = 2$, since $d = 1$ would imply that $D = F$ as desired. To do this note that if $d = 2$ that $\dim_{\mathbb{Q}} D = 2 \dim B$, so that (again by Albert) B/K has maximal totally indefinite quaternionic multiplication, or maximal totally definite quaternionic multiplication. The former is not possible by a theorem of Shimura (here we again use that K has a real place), see Theorem 0 (with $n = 1$ in his notation) of [94].

As for the latter, if B/K has totally definite quaternionic multiplication by D over K, then certainly so does $B/\mathbb{C}$, and then we apply Shimura's argument in

Section 4.3 of [93] (i.e. his proof that his Proposition 14 implies his Proposition 15 in [93]) to produce an imaginary quadratic extension L/F with $L \hookrightarrow D$ and $L \curvearrowright \mathrm{Lie}(B/\mathbb{C})$ of generalized CM-type — i.e., there is a CM type $\Phi \subseteq \mathrm{Hom}_{\mathbb{Q}\text{-alg.}}(L, \mathbb{C})$ for which the L-representation $\mathrm{Lie}(B/\mathbb{C})$ decomposes into $\bigoplus_{\sigma \in \Phi} \sigma^{\oplus 2}$. In Shimura's notation in [93], this is to say that all the $r_\nu = 2$ and the $s_\nu = 0$ (a priori $r_\nu = s_\nu = 1$ was also possible for some ν).

Then since $L \hookrightarrow D$ and $D = \mathrm{End}^0_K(B)$, it follows that L acts by K-endomorphisms of B/K, so that Lemma 9.3.2 applies and we deduce the desired contradiction. □

9.3.2 The residual images of relevant compatible families of two-dimensional Galois representations.

The following is an explicit form of a theorem of Dimitrov. We note that it is also an improvement on Lemma $7.1.3$ of the ten-author paper [3], which concludes the same for a Dirichlet density 1 subset of the primes.

Lemma 9.3.3 (Cf. Propositions 3.1 and 3.8 of Dimitrov's [44], and Lemma $7.1.3$ of the ten-author paper [3].). *Let $E/\mathbb{Q}$ be a number field. Let $K/\mathbb{Q}$ be a number field that is neither $\mathbb{Q}$ nor imaginary quadratic. Let $N \in \mathbb{Z}^+$. Let ℓ be a prime of $\mathbb{Z}$ prime to N. Let $\lambda \subseteq \mathfrak{o}_E$ with $\lambda|(\ell)$ be a prime of E.*

Then: there is an explicit (thus effectively computable) constant $C_{E,K,N,\lambda} \in \mathbb{Z}^+$, depending explicitly and only on E, K, N, and λ, such that the following holds. Let $p \geq C_{E,K,N,\lambda}$ be a prime of $\mathbb{Z}$. Let $\mathfrak{p} \subseteq \mathfrak{o}_E$ be a prime of E with $\mathfrak{p}|(p)$. Let $\rho_\lambda, \rho_\mathfrak{p}$ be representations

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\lambda})$$
$$\rho_\mathfrak{p} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\mathfrak{p}})$$

satisfying the following:

- *we have*

$$\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_{\mathfrak{m}})) = \mathrm{tr}(\rho_{\mathfrak{p}}(\mathrm{Frob}_{\mathfrak{m}})) \in \mathfrak{o}_E$$

  and

$$|\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_{\mathfrak{m}}))|_v \leq 10^{10} \cdot (\mathrm{Nm}\,\mathfrak{m})^{10^{10}}$$

  for all $v|\infty$ and primes $\mathfrak{m} \subseteq \mathfrak{o}_K$ of K with $\mathrm{Nm}\,\mathfrak{m}$ prime to N and λ, and with $\mathrm{Nm}\,\mathfrak{m} \leq C_{E,K,N,\lambda}$,

- *$\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell$ is irreducible, unramified at primes not dividing $N \cdot \ell$, and not of the form $\mathrm{Ind}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}^{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}(\chi)$ with L/K quadratic and $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/L) \to \overline{\mathbb{Q}}_\ell^\times$ a character,*

- *$\rho_{\mathfrak{p}} \otimes_{\mathfrak{o}_{E,\mathfrak{p}}} E_{\mathfrak{p}}$ has conductor dividing $N \cdot p^\infty$ and $\rho_{\mathfrak{p}}$ is crystalline with Fontaine-Laffaille/Hodge-Tate weights $\{0,-1\}$ under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$.*

Then: writing the mod-$\mathfrak{p}$ residual representation as $\overline{\rho}_{\mathfrak{p}} := \rho_{\mathfrak{p}} \otimes_{\mathfrak{o}_{E,\mathfrak{p}}} \mathfrak{o}_{E,\mathfrak{p}}/\mathfrak{p}$,

$$\mathrm{SL}_2(\mathbb{F}_p) \subseteq \overline{\rho}_{\mathfrak{p}}(\mathrm{Gal}(\overline{\mathbb{Q}}/K)).$$

Thus in particular the Taylor-Wiles hypothesis at $\mathfrak{p}$, namely that

$$(\overline{\rho}_{\mathfrak{p}} \otimes_{\mathfrak{o}_E/\mathfrak{p}} \overline{\mathbb{F}}_p)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K(\zeta_p))} : \mathrm{Gal}(\overline{\mathbb{Q}}/K(\zeta_p)) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$$

is irreducible, holds.

The last line follows because $\mathrm{SL}_2(\mathbb{F}_p)$ is perfect, so that the composition

$$\mathrm{SL}_2(\mathbb{F}_p) \hookrightarrow \overline{\rho}_{\mathfrak{p}}(\mathrm{Gal}(\overline{\mathbb{Q}}/K)) \twoheadrightarrow \overline{\rho}_{\mathfrak{p}}(\mathrm{Gal}(\overline{\mathbb{Q}}/K))/\overline{\rho}_{\mathfrak{p}}(\mathrm{Gal}(\overline{\mathbb{Q}}/K(\zeta_p))$$

is trivial (since the latter is surjected upon by $\mathrm{Gal}(K(\zeta_p)/K) \hookrightarrow \mathbb{F}_p^\times$), and $\mathrm{SL}_2(\mathbb{F}_p) \hookrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is evidently irreducible (consider the torus and then the unipotents).

We note that we have taken K to not be $\mathbb{Q}$ or an imaginary quadratic field so that $\mathfrak{o}_K^\times$ is of infinite order — of course we may always arrange that this is the case by first passing to a totally real cubic extension of any number field input into an algorithm in this chapter. We will leave this implicit when using this lemma.

Proof. The proof of the lemma is the same as Dimitrov's arguments for Hilbert modular forms, except that we change the end of his proof of Proposition 3.1 in his [44] in order to obtain an explicit bound (which is necessary in Chapter 10). Let us indicate[21] precisely how.

Dimitrov's argument to prove his Proposition 3.1 of his [44] goes as follows. If $\overline{\rho}_\mathfrak{p}$ is reducible, then (because we may read its Fontaine-Laffaille/Hodge-Tate weights off from those of $\rho_\mathfrak{p}$, since the category of crystalline representations is closed under passing to subquotients and the Fontaine-Laffaille functor is exact) it follows that either $\mathfrak{p}$ divides a nonzero constant only depending on E and N (that Dimitrov specifies), or else $\mathrm{tr} \circ \overline{\rho}_\mathfrak{p}$ is the sum of the reductions of algebraic Hecke

[21] For completeness, we give a slight modification of Dimitrov's proof of absolute irreducibility here. First, $0 \to \pi \cdot \rho_\mathfrak{p} \to \rho_\mathfrak{p} \to \overline{\rho}_\mathfrak{p} \to 0$, where π is a uniformizer of $\mathfrak{p}$. Because subquotients of crystalline representations are crystalline, it follows that $\overline{\rho}_\mathfrak{p}$ is crystalline. Because the Fontaine-Laffaille functor is exact, we find that $\overline{\rho}_\mathfrak{p}$ has Fontaine-Laffaille weights $\{0, -1\}$ at all places above p.

Suppose now that $\overline{\rho}_\mathfrak{p} \otimes_{\mathfrak{o}_E/\mathfrak{p}} \overline{\mathbb{F}}_p$ is absolutely reducible, into e.g. $\mathrm{tr} \circ \overline{\rho}_\mathfrak{p} = \alpha + \beta$ with $\alpha, \beta : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \overline{\mathbb{F}}_p^\times$. By crystallineness and the fact that subquotients of crystalline representations are again crystalline, it follows that α and β are crystalline, so that e.g. α has Fontaine-Laffaille weight either 0 or -1 under each embedding $K \hookrightarrow \overline{\mathbb{Q}}_p$. Let S_α be the finite set of embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$ at which α has Fontaine-Laffaille weight -1, and similarly for S_β, so that $S_\alpha \cup S_\beta$ is the finite set of embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$, and $S_\alpha \cap S_\beta = \emptyset$.

The conductors of α and β also divide $N \cdot p^\infty$. Regarding them as characters of the idèle class group $\mathbb{I}_K/K^\times$, and thus the (narrow) ray class group $\mathrm{Cl}_{N\cdot p^\infty}(K)$, we see that they must e.g. be trivial on global units $\varepsilon \in \mathfrak{o}_K^\times$ with $\varepsilon \equiv 1 \pmod{N}$. Thus e.g.

$$1 \equiv \alpha(\varepsilon) \equiv \prod_{i \in S_\alpha} i(\varepsilon)^{-1} \pmod{\mathfrak{p}'}$$

for $\mathfrak{p}'|\mathfrak{p}$ a prime of an explicit extension E'/E with $[E' : E] \leq 2$. Evidently (by explicitly constructing a suitable ε via Minkowski) this cannot occur for all ε unless p is explicitly bounded, or else one of S_α or S_β is empty. Without loss of generality $S_\alpha = \emptyset$, which is to say that $\psi := \alpha$ and $\psi' := \beta \cdot \chi_p^{-1}$, which via the canonical lifting character we regard as maps $\psi, \psi' : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathfrak{o}_{E',\mathfrak{p}'}^\times$ with E'/E explicit with $[E' : E] \leq 2$ and $\mathfrak{p}'|\mathfrak{p}$, have Fontaine-Laffaille weight 0 under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$. Recalling that $\prod_{\mathfrak{q}|(p)} \mathfrak{o}_\mathfrak{q}^\times \to \mathrm{Cl}_{N\cdot p^\infty}(K) \to \mathrm{Cl}_N(K) \to 0$, it follows that ψ and ψ' factor through the ray class group $\mathrm{Cl}_N(K)$, so that they are indeed finite order. The rest of the argument is below.

characters of the form $\psi, \psi' \cdot \chi_p$, where $\psi, \psi' : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathfrak{o}_{E',\mathfrak{p}'}^\times$ have conductor dividing N and are of finite order (thus we may have written their codomain as $(\mathfrak{o}_{E'}/\mathfrak{p}')^\times$, where E'/E is an explicit extension of degree $[E' : E] \leq 2$ and $\mathfrak{p}'|\mathfrak{p}$ is a prime of E'), and χ_p is the p-adic cyclotomic character. Note that all these characters fit into compatible families. Thus we find that

$$\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_\mathfrak{m})) = \mathrm{tr}(\rho_\mathfrak{p}(\mathrm{Frob}_\mathfrak{m})) \equiv \psi(\mathrm{Frob}_\mathfrak{m}) + \psi'(\mathrm{Frob}_\mathfrak{m}) \cdot \mathrm{Nm}\,\mathfrak{m} \pmod{\mathfrak{p}'}$$

for all primes $\mathfrak{m} \subseteq \mathfrak{o}_K$ with $\mathrm{Nm}\,\mathfrak{m}$ prime to N and such that $\mathrm{Nm}\,\mathfrak{m} \leq C_{E,K,N,\lambda}$. We take $C_{E,K,N,\lambda}$ sufficiently large so that the tuple of traces at this set of primes determines the irreducible Galois representation $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} E_\lambda$ up to isomorphism (see Lemma 7.2.1 of Chapter 7, especially the comments on explicitly bounding the set using an explicit form of the Chebotarev density theorem).

But note also that both sides of the congruence are elements of $\mathfrak{o}_E$ that are $\leq 10^{10} \cdot (\mathrm{Nm}\,\mathfrak{m})^{10^{10}}$ at all infinite places. Thus once $\mathrm{Nm}\,\mathfrak{p}$ is (explicitly) sufficiently large the congruence must be an equality. Therefore, because the traces of the two λ-adic representations $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} E_\lambda$ and $\psi \oplus \psi' \cdot \chi_\ell$ match at these $\mathfrak{m}$, it follows that $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell \cong \psi \oplus \psi' \cdot \chi_\ell$ is reducible. Contradiction. □

The difference between this argument and Dimitrov's is that he appeals to Chebotarev's density theorem to similarly conclude that the above congruence (thus equality once $\mathfrak{p}$ is sufficiently large in terms of $\mathfrak{m}$) cannot hold for all but finitely many $\mathfrak{m}$, whereas we use Faltings' Lemma to restrict to an explicit finite set of $\mathfrak{m}$. This difference is of course minor, but it is crucial for us to have an explicit bound rather than an "almost all" result.

9.3.3 Relevant Galois representations that are induced arise from CM abelian varieties.

Finally, we note that the situation that is relevant to us but left out by the hypotheses of Lemma 9.3.3, namely when one of the Galois representations is induced, is quite simple: the two-dimensional λ-adic Galois representation corresponds to the λ-adic Galois representation associated to the Weil restriction from an imaginary quadratic extension L/K of (the isogeny class of) a CM abelian variety A/L. Moreover, because the conductor of the two-dimensional Galois representation (and thus of $\mathrm{Res}^L_K(A)/K$) is bounded in terms of S it follows that L and thus A are restricted to explicit finite sets. The point is simply that the two-dimensional λ-adic Galois representation must be induced from an algebraic Hecke character. Indeed, since these are classified, we quickly see that the corresponding quadratic extension L/K is imaginary CM. Because the Galois representations relevant to us have integral traces and determinant with all Hodge-Tate weights -1, it follows that the algebraic Hecke character will correspond to a CM abelian variety (whose isogeny class is) defined over L, as desired.

Lemma 9.3.4. *Let $E/\mathbb{Q}$ be a number field. Let $K/\mathbb{Q}$ be a number field. Let S be a finite set of places of K. Let ℓ be a prime of $\mathbb{Z}$ prime to S. Let $\lambda \subseteq \mathfrak{o}_E$ with $\lambda|(\ell)$ be a prime of E.*

Let $\rho_\lambda : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\lambda})$ be a representation that:

- *is unramified outside S and primes over ℓ,*
- *has $\det \rho_\lambda \cdot \chi_\ell^{-1}$ of finite order, where χ_ℓ is the ℓ-adic cyclotomic character,*
- *has $\mathrm{tr}(\rho_\lambda(\mathrm{Frob}_\mathfrak{p})) \in \mathfrak{o}_E$ for all primes $\mathfrak{p} \subseteq \mathfrak{o}_K$ with $\mathrm{Nm}\,\mathfrak{p}$ prime to S and ℓ,*
- *is of the form $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell \cong \mathrm{Ind}^{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\tilde{L})}(\tilde{\chi})$, where $\tilde{L}/K$ is quadratic and $\tilde{\chi} : \mathrm{Gal}(\overline{\mathbb{Q}}/\tilde{L}) \to \overline{\mathbb{Q}}_\ell^\times$ is a character.*

Then: there is an imaginary CM field $L'/\mathbb{Q}$, a quadratic extension L/K that is unramified outside S and primes over ℓ with $L' \subseteq L$, and a CM type $\Phi \subseteq \mathrm{Hom}_{\mathbb{Q}\text{-}alg.}(L', \mathbb{C})$ for which, writing $\chi^{(\Phi)}$ for the algebraic Hecke character corresponding to the reflex CM type of (L', Φ), $\tilde{\chi}^{(\Phi)} := \chi^{(\Phi)} \circ \mathrm{Nm}_{L/L'}$, and $\tilde{\chi}_\ell^{(\Phi)}$ for the corresponding compatible family of ℓ-adic representations $\tilde{\chi}_\ell^{(\Phi)} : \mathrm{Gal}(\overline{\mathbb{Q}}/L) \to \overline{\mathbb{Q}}_\ell^\times$,

$$\mathrm{Ind}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}^{\mathrm{Gal}(\overline{\mathbb{Q}}/K)} \tilde{\chi}_\ell^{(\Phi)} \cong \rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell.$$

In particular, if we also have that $K/\mathbb{Q}$ is totally real with $[K : \mathbb{Q}]$ odd, then, for $v_K|\infty$, there is a K-isogeny factor A/K of the Jacobian $\mathrm{Jac}\, C_{v_K}\left(\prod_{\mathfrak{p}\in S} \mathfrak{p}^{10^{10g}\cdot[K:\mathbb{Q}]}\right)$ of the Shimura curve (with full level-$\left(\prod_{\mathfrak{p}\in S} \mathfrak{p}^{10^{10g}\cdot[K:\mathbb{Q}]}\right)$ structure) corresponding to the quaternion algebra split at all infinite places of K other than v_K and unramified at all finite places, for which $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell$ is the 2-dimensional λ-adic $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$-representation corresponding to A/K.

Proof. Let $L := \tilde{L}$.

It follows from the main result in Henniart's [56][22] that $\tilde{\chi} = \tilde{\xi}_{\lambda'}$ is the λ'-adic realization of an algebraic Hecke character $\tilde{\xi}$ of L, where $\tilde{\xi}(\mathrm{Frob}_\mathfrak{q}) \in E'^\times$ for all $\mathfrak{q}$ with $\mathrm{Nm}\,\mathfrak{q}$ prime to S and ℓ, E'/E is an extension with $[E' : E] \leq 2$, and $\lambda' \subseteq \mathfrak{o}_{E'}$ is a prime of E' with $\lambda'|\lambda$.

By the classification of algebraic Hecke characters (see e.g. Proposition 1.12 of Fargues' [48], Section 3.4 of Serre's [91], or Section 3 in Chapter 0 of Schappacher's [90]), it follows that, writing $L' \subseteq \tilde{L}$ for the maximal imaginary CM subfield of L,

[22] See also page III-20 of Serre's [91] — Henniart's [56] removes the compositum-of-quadratic-fields hypothesis by using a stronger input from transcendence theory due to Waldschmidt (see page 119 of [105]). However, we must comment on the rationality hypothesis, even though it seems it is usually skipped in the literature. Let $\mathfrak{p} \subseteq \mathfrak{o}_K$ be split in L/K, say into $\mathfrak{p} =: \mathfrak{P} \cdot \tilde{\mathfrak{P}}$, and such that $\tilde{\chi}(\mathrm{Frob}_\mathfrak{P}) \neq \tilde{\chi}(\mathrm{Frob}_{\tilde{\mathfrak{P}}})$ (this must occur if $\tilde{\chi}$ does not extend to a character on $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$, and in the latter case the argument is easy). Write E'/E for a number field containing the roots of the characteristic polynomial of $\rho_\lambda(\mathrm{Frob}_\mathfrak{p})$. Then, since the projectors onto the distinct eigenspaces of $\rho_\lambda(\mathrm{Frob}_\mathfrak{p})$ have coefficients in E', it follows that $\tilde{\chi}(\mathrm{Frob}_\mathfrak{q}) \in E'^\times$ for all primes $\mathfrak{q} \subseteq \mathfrak{o}_L$ of L that are prime to S and ℓ.

or $\mathbb{Q}$ if there is no such,

$$\chi = \psi \cdot (\xi \circ \mathrm{Nm}_{L/L'})$$

for ψ a finite-order character of L and ξ an algebraic Hecke character of L'.

Write (see Section 3 in Chapter 0 of Schappacher's [90]) $(n_\sigma)_{\sigma : L' \hookrightarrow \mathbb{C}} \in \mathbb{Z}^{\mathrm{Hom}_{\mathbb{Q}\text{-alg.}}(L',\mathbb{C})}$, thus $n_\sigma \in \mathbb{Z}$ for all $\sigma : L' \hookrightarrow \mathbb{C}$, for the infinity-type of ξ. Let $w \in \mathbb{Z}$ be the weight of ξ — i.e. such that

$$n_\sigma + n_{\varepsilon \circ \sigma} = w$$

for all $\sigma : L' \hookrightarrow \mathbb{C}$ and ε complex conjugations on L'. Thus

$$|\xi(\mathfrak{P})| = (\mathrm{Nm}\,\mathfrak{P})^{\frac{w}{2}}$$

for all primes $\mathfrak{P}$ of L' with $\mathrm{Nm}\,\mathfrak{P}$ prime to S and ℓ.

Since

$$\rho_\lambda|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)} \cong \left(\mathrm{Ind}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}^{\mathrm{Gal}(\overline{\mathbb{Q}}/K)} \tilde{\xi}_{\lambda'}\right)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)} \simeq \tilde{\xi}_{\lambda'} \oplus (\tilde{\xi}_{\lambda'} \circ \text{conj.}_\tau),$$

where $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ is a lift of the nontrivial element of $\mathrm{Gal}(L/K)$, it follows that

$$\mathrm{Nm}\,\mathfrak{P} = |(\det \rho_\lambda)(\mathfrak{P})| = |\tilde{\chi}(\mathrm{Nm}_{L/L'}(\mathfrak{P}))| \cdot |\tilde{\chi}(\mathrm{Nm}_{L/L'}(\mathfrak{P})^\tau)| = (\mathrm{Nm}\,\mathfrak{P})^w$$

for all primes $\mathfrak{P}$ of L with $\mathrm{Nm}\,\mathfrak{P}$ prime to S and ℓ. Thus $w = 1$.

It follows that $L' \neq \mathbb{Q}$. Indeed, because the algebraic Hecke characters of $\mathbb{Q}$ are all of the form (finite order) $\cdot$ $(z \mapsto |z_\infty|^k)$ for $k \in \mathbb{Z}$, and such characters have infinity-type (k), all algebraic Hecke characters of $\mathbb{Q}$ have even weight.

So $\tilde{\xi}$ is an algebraic Hecke character of an imaginary CM field. Moreover, from the above calculation $\tilde{\xi}$ has weight (as an algebraic Hecke character) 1.

We next use the fact that the $\rho_\lambda(\mathrm{Frob}_\mathfrak{p})$, and thus the $\rho'_\lambda(\mathrm{Frob}_\mathfrak{P})$ by Gauss's Lemma, have integral characteristic polynomials. By Corollary 3 on page II-36 of Serre's [91], it follows that the algebraic Hecke character χ has infinity-type $(n_\sigma)_{\sigma : L' \hookrightarrow \mathbb{C}}$ with all $n_\sigma \geq 0$. Since it has weight 1, so that $n_\sigma + n_{\varepsilon\circ\sigma} = 1$ for all complex conjugations ε on L', it follows that $n_\sigma \in \{0, 1\}$ for all $\sigma : L' \hookrightarrow \mathbb{C}$.

Let $\Phi := \{\sigma : L' \hookrightarrow \mathbb{C} \,|\, n_\sigma = 1\}$. Then $\Phi \subseteq \mathrm{Hom}_{\mathbb{Q}\text{-alg.}}(L', \mathbb{C})$ is a CM type of L', and indeed evidently $\tilde{\xi} = \chi^{(\Phi)}$ is the algebraic Hecke character corresponding to the reflex CM type of (L', Φ).

Finally, because the induced representation is unramified outside S and primes above ℓ, it follows that L/K is unramified outside S and primes above ℓ (evident by hand, or else see Section 7.1 of Hida's [57] and recall the equality of conductor and level arising from the equality of ε-factors in Section 0.5 of Carayol's [35]). The first statement follows.

As for the second statement, write f_Φ for the automorphic induction (see Yoshida's Princeton PhD thesis [109] or Arthur-Clozel [9]) to GL_2/K of $\chi^{(\Phi)}$. Note that f_Φ is a parallel weight 2 cuspidal Hilbert modular eigencuspform with level dividing[23] $\mathfrak{m}_S := \prod_{\mathfrak{p}\in S} \mathfrak{p}^{10^{10g}\cdot[K:\mathbb{Q}]}$ by Yoshida's thesis [109] (see also Section 7.1 of Hida's [57]). The corresponding (i.e., to its Jacquet-Langlands transfer) K-quotient of $\mathrm{Jac}\, C_{v_K}(\mathfrak{m}_S)$ has the claimed properties. □

9.4 Proof of termination and correctness.

Let us now turn to the proof of termination and correctness of the algorithm specified in Section 9.2.1. In this section we prove the following theorem.

[23] This is of course a vast overestimate.

Theorem 9.4.1. *Let $K/\mathbb{Q}$ be a totally real field of odd degree. Let S be a finite set of places of K. Let $F/\mathbb{Q}$ be a totally real field. Let $\mathfrak{o} \subseteq \mathfrak{o}_F$ be an order in F. Then: on input $(\mathfrak{o}, K, S)$, the algorithm specified in Section 9.2.1 terminates with output $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$.*

Evidently this implies Theorem 9.1.1.

Again, as noted in Section 9.2.1, essentially the same algorithm allows one to determine the abelian varieties A/K with good reduction outside S and admitting a map $L \hookrightarrow \mathrm{End}^0_K(A)$ with $[L : \mathbb{Q}] = \dim A$ when $L/\mathbb{Q}$ is CM. (The theorem asserts the same for $L/\mathbb{Q}$ totally real.)

9.4.1 Proof of Theorem 9.4.1.

We break the proof into two parts for clarity. First we prove that, if the algorithm terminates, its output is correct. After that we will prove that the algorithm always terminates.

9.4.1.1 Proof of correctness.

Proof of correctness assuming termination. We first deal with the proof of correctness. Because of the last step of the algorithm in Section 9.2.1, it suffices to show that any abelian variety A/K with $\mathfrak{o}$-multiplication over K and good reduction outside S has height bounded by the H computed in Step 5 of the algorithm.

Write $F := \mathrm{Frac}\,\mathfrak{o}$. By Lemma 9.3.1, such an A is of the form $A \sim_K B^{\times n}$ with B/K K-simple and $n := \frac{\dim A}{\dim B} \leq [F : \mathbb{Q}]$, and such that $E \simeq \mathrm{End}^0_K(B)$, where $E \subseteq F$ is such that $[E : \mathbb{Q}] = \dim B$.

The K-endomorphism algebra $\mathfrak{o}' := \mathrm{End}_K(B)$ is thus an order in $\mathfrak{o}_E$. Let, for each prime λ of $\mathfrak{o}'$ with $\lambda|(\ell)$, say, $\rho_\lambda : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}'_\lambda)$ be the Galois representation corresponding to the λ-adic Tate module $T_\lambda(B) := \varprojlim B[\lambda^n]$. By construction, ρ_λ is odd, weight two (in the terminology of Snowden [95]), pure of weight 1, unramified outside S and primes above ℓ, and has $(\det \rho_\lambda) \cdot \chi_\ell^{-1}$ a finite-order character, where $\lambda | (\ell)$ and χ_ℓ is the ℓ-adic cyclotomic character. By Brumer-Kramer [34] it moreover has conductor dividing $\mathfrak{m}_K := \prod_{\mathfrak{p} \in S} \mathfrak{p}^{10^{10g} \cdot [K:\mathbb{Q}]}$. Because $E \simeq \operatorname{End}_K^0(B)$, we also have that $\rho_\lambda \otimes_{\mathfrak{o}_{E,\lambda}} \overline{\mathbb{Q}}_\ell$ is irreducible.

We first claim that we are done so long as the set $\mathcal{F}$ produced in Step 2 is such that there is an $L \in \mathcal{F}$ for which B/L is modular — i.e. there is a parallel weight 2 cuspidal Hilbert modular eigencuspform f over L with $L(s, f) = L(s, \rho_\lambda)$ up to finitely many Euler factors and for λ with $\operatorname{Nm} \lambda$ prime to S.

Indeed, in that case by the equality of ε-factors in Section 0.5 of Carayol's [35] it follows that the conductors of the λ-adic representations associated to f match its level. Because, by Brumer-Kramer [34], the conductor of A/L certainly divides $\mathfrak{m}_L := \prod_{\mathfrak{p} \subseteq \mathfrak{o}_L : (\mathfrak{p}, S) \cdot (\mathfrak{p}, \lambda) \neq 1} \mathfrak{p}^{10^{10g} \cdot [L:\mathbb{Q}]}$ (defined as in Step 3 of the algorithm in Section 9.2.1), it follows that the level of f divides $\mathfrak{m}_L$.

By the Jacquet-Langlands correspondence [59] (see also Theorem 3.9 of [43]), it follows that f transfers to a quaternionic modular form of parallel weight 2 and level dividing $\mathfrak{m}_L$ on $B_{v_L}^\times / L$, the units of the quaternion algebra over L split at all finite places and at v_L and ramified at all other infinite places. Thus by e.g. Theorem 4.4 of Hida's [57] it follows that there is an L-isogeny factor B_f/L of $\operatorname{Jac} C_{v_L}(\mathfrak{m}_L)$ with $a_\mathfrak{p}(B_f) = a_\mathfrak{p}(f) = a_p(B) := \operatorname{tr}(\rho_\lambda(\operatorname{Frob}_\mathfrak{p}))$ for all $\mathfrak{p} \notin S$ and $\mathfrak{p} \nmid \operatorname{Nm} \lambda$. We deduce that B_f/L is L-isogenous to B/L, so that B/L is an L-isogeny factor of $\operatorname{Jac} C_{v_L}(\mathfrak{m}_L)$. Thus by Poincaré complete reducibility it follows that there is a B'/L for which $\operatorname{Jac} C_{v_L}(\mathfrak{m}_L) \sim_L B \times B'$. Consequently $\operatorname{Jac} C_{v_L}(\mathfrak{m}_L)^{\times n} \sim_L A \times (B')^{\times n}$. Hence by Masser-Wüstholz (see Theorem 7.2.2) and Bost (see Theorem 7.2.3):

$$\begin{aligned} h(A) - \text{Bost}(n \cdot \dim \operatorname{Jac} C_{v_L}(\mathfrak{m}_L)) &\leq h(A) - \text{Bost}(n \cdot \dim B') \\ &\leq h(A) + h(B'^{\times n}) \end{aligned}$$

$$\begin{aligned} &= h(A \times (B')^{\times n}) \\ &\leq \text{MasserWüstholz}(\operatorname{Jac} C_{v_L}(\mathfrak{m}_L)^{\times n}, L). \end{aligned}$$

Thus $h(A) \leq H$, as desired.

Thus it suffices to show that there is an $L \in \mathcal{F}$ for which B/L is modular.

Let us first deal with a special case. If ρ_λ is the induction of a character from a quadratic extension, then by Lemma 9.3.4 it follows that B/K is already modular, and indeed the conclusion of Lemma 9.3.4 gives us the input required for the above argument over K (recall that $K \in \mathcal{F}$). So we are done if ρ_λ is induced. Thus from now on we assume that ρ_λ is *not* induced.

Now we follow the RelevantExtensions routine.

Let $\psi_\lambda := (\det \rho_\lambda) \cdot \chi_\ell^{-1}$. Then certainly $\psi_\lambda \in \Psi_\lambda$. Moreover we have already listed properties of ρ_λ which imply that $(\rho_\lambda \bmod \lambda, \psi_\lambda) \in R_\lambda$. So we are reduced to showing the correctness of the output of the TaylorMoretBaillyExtensions subroutine.

But Snowden's proof[24] of Theorem $5.1.2$ in [95] amounts exactly to a proof of correctness of the TaylorMoretBaillyExtensions subroutine! Moreover, the output is exactly a field, say L, over which, by Snowden's Theorem $5.1.2$ in [95], B/L is modular. Moreover L/K is Galois.

Now we follow Snowden's proof of Proposition $9.4.1$ in [95]. Let $H \subseteq \operatorname{Gal}(L/K)$ be a 2-Sylow subgroup. Note that $H \simeq \operatorname{Gal}(L/L^H)$ is solvable. By solvable descent for GL_2 (see e.g. Lapid-Rogawski [65], Langlands [64], or Arthur-Clozel/Clozel-Rajan (who fix the well-known gap) [9,39]), it follows that B/L^H is also modular.

[24] There is one caveat: Snowden never actually explicitly checks that Galois representations associated to GL_2-type abelian varieties have definite type functions — but he does prove it! See his Proposition 2.6.1 in [95], and replace "potentially modular" by "pure of weight 1". The proof of the proposition just uses purity (which he deduces in the potentially modular case from the Ramanujan bound for Hilbert modular forms) to rule out having an extension of the cyclotomic character by the trivial character over a finite extension.

Moreover, $L^H \in \mathcal{F}$ by construction. Thus we have shown that B/K is modular over a field in $\mathcal{F}$ in all cases. By the above discussion this completes the proof. □

9.4.1.2 Proof of termination.

Proof of termination. For only a few steps in the algorithms is termination not self-evident, but we will comment on each step nonetheless.

We begin with the IntegralPointsOnHilbertModularVarieties algorithm. Suppose for the moment that the RelevantExtensions algorithm always terminates in finite time. Then it is evident that Steps $1, 3, 4, 5, 6$ terminate in finite time, and we have supposed that Step 2 terminates in finite time. So it suffices to show that the RelevantExtensions algorithm always terminates in finite time.

So we move next to the RelevantExtensions algorithm.

Evidently Step 1 terminates. The characters in Step 2 are simply those factoring through the canonical lifting character $(\mathfrak{o}_E/\lambda)^\times \to \mathfrak{o}_{E,\lambda}^\times$ (lifting the canonical surjection), since these are the finite-order elements of $\mathfrak{o}_{E,\lambda}^\times$ (use the logarithm). Consequently Step 2 is computed in the usual way, which we have been leaving implicit throughout: by finding number fields of bounded degree and bounded ramification using Minkowski. In any case, it evidently terminates.

Similarly in Step 3 we produce a set of Galois representations of bounded image and ramification, again by Minkowski. So evidently Step 3 terminates.

Suppose for a moment that the TaylorMoretBaillyExtensions algorithm always terminates. Then evidently Step 4 always terminates. Moreover Steps 5 and 6 of course terminate, giving the termination of the RelevantExtensions algorithm.

So it remains to show that the TaylorMoretBaillyExtensions algorithm always terminates.

Step 1 terminates (Snowden even gives an explicit construction). Steps 2 and 3 of course terminate. Step 4 again terminates because Snowden gives an explicit construction. Step 5 evidently terminates. Suppose for the moment that Step 6 terminates. Then Steps 7 and 8 evidently terminate, so that we are done. Thus we have reduced the whole proof to checking that Step 6 terminates.

But the termination of Step 6 is Moret-Bailly's theorem (see Theorem 1.3 in part II of [72]). Specifically, we are asking that a search for an element of a nonempty recursively enumerable set (the set is nonempty precisely by Moret-Bailly's theorem, and it is recursively enumerable because we are dealing with algebraic numbers) will terminate in finite time, which is evident.[25] This concludes the proof. □

9.5 Modifications for $[K:\mathbb{Q}]$ even.

For $K/\mathbb{Q}$ totally real of even degree, not all parallel weight 2 Hilbert modular eigencuspforms admit such an easy construction of their associated abelian varieties — specifically, since K now has an even number of infinite places, we can no longer simply transfer to a quaternion algebra split at all but one infinite place[26], since one needs local conditions at some finite prime to hold for f. And indeed it is not yet known for these remaining f that there is an abelian variety A/K of $\mathrm{GL}_2(\mathfrak{o}_F)$-type (with $F/\mathbb{Q}$ the coefficient field of f) over K with $\rho_{A,\lambda} \cong \rho_{f,\lambda}$ for $\lambda \subseteq \mathfrak{o}_F$ a prime of F with sufficiently large norm.

However the entire rest of the analysis of the previous sections goes through for such $K/\mathbb{Q}$. Thus we see that we may run identical algorithms, except that in

[25] Of course one can be more explicit here, and we have already commented that Rumely's proof of Fekete-Szegő, or Moret-Bailly's original proof of Rumely's theorem, are indeed far more explicit, but for our sake this argument suffices. See also the comments on Step 6 of the TaylorMoretBaillyExtensions algorithm.

[26] The number of split places at infinity is the dimension of the corresponding quaternionic Shimura variety, and so to get to a Shimura curve there must be exactly one — in general if one transfers to a quaternion algebra one finds ℓ-adic representations formed by a tensor induction corresponding to the various split places at infinity (e.g. if one takes the cohomology of the Hilbert modular variety itself one finds the full $2^{[K:\mathbb{Q}]}$-dimensional tensor induction in its degree-$[K:\mathbb{Q}]$ cohomology). Of course the total number of split places of the quaternion algebra must be even, which is why evenness of $[K:\mathbb{Q}]$ is an issue.

Step 5 of the algorithm in Section 9.2.1 we must produce a height bound $h_{L,S}$ on the abelian variety A_f associated to a parallel weight 2 Hilbert modular eigencuspform over L of level dividing $\mathfrak{m}_L$ — that is to say, we must prove that either there is no A_f/L associated to f, or else $h(A_f) \leq h_{L,S}$ with $h_{L,S}$ explicitly computable in terms of L and S.

Now, assuming the absolute Hodge conjecture, Blasius in [25] has proven that such f have associated abelian varieties A_f/L. Thus our first remark is that the algorithm of Section 9.2.1 is easily modified to produce an algorithm that one can prove terminates with correct output (namely $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$) assuming the absolute Hodge conjecture. In other words it is easy to modify the algorithm in Section 9.2.1 to treat the case of $[K:\mathbb{Q}]$ even, but, in doing so naïvely, one gives up its most interesting aspect, namely that one can prove it terminates with correct output *unconditionally* when $[K:\mathbb{Q}]$ is odd.

The purpose of this section is to sketch the proof of a slightly more interesting result, namely Theorem 9.1.2. Specifically, we will explain how to modify the algorithm of Section 9.2.1 so that one can prove unconditionally that the algorithm always terminates, and either outputs $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$ *or else an unconditional disproof of the absolute Hodge conjecture*.

The crux of the argument will be an upper bound for the height of an A_f/L associated to an f on GL_2/L, if it exists. By following Blasius's construction (though redoing things integrally), we will prove the following:

Theorem 9.5.1. *There is an explicitly computable function $(L,\mathfrak{n}) \mapsto h_{L,\mathfrak{n}}$ with the following property. For all totally real number fields $L/\mathbb{Q}$ and all parallel weight 2 Hilbert modular eigencuspforms f over L (with coefficient field $F/\mathbb{Q}$, say) of level dividing $\mathfrak{n}$, the absolute Hodge conjecture implies that there is an A_f/L of $\mathrm{GL}_2(\mathfrak{o}_F)$-type over L with $h(A_f) \leq h_{L,\mathfrak{n}}$ and $\rho_{A_f,\lambda} \cong \rho_{f,\lambda}$ for all primes $\lambda \subseteq \mathfrak{o}_F$ of F with sufficiently large norm.*

In each particular case (and, at the end of the algorithm in Section 9.2.1 we are reduced to checking this for the explicit finite set of parallel weight 2 Hilbert modular eigencuspforms of level dividing $\mathfrak{m}_L$), the existence of such an A_f/L is verified or falsified by an explicit finite search. If, in the course of the algorithm, we fail to find such an abelian variety, then we have an (unconditional) disproof of the absolute Hodge conjecture, by virtue of Theorem 9.5.1.[27] So we simply output this f and a certificate that none of the finitely many $[F : \mathbb{Q}]$-dimensional abelian varieties A/L with $h(A) \leq h_{L,\mathfrak{m}_L}$ are associated to f, and this serves as a certificate for the falsehood of the absolute Hodge conjecture.

Otherwise this finite search always finds an A_f/L, and then one concludes in the same way as in Steps 5 and 6 — specifically, one uses Masser-Wüstholz and $h(A_f) \leq h_{L,\mathfrak{m}_L}$ for all relevant f to produce a height bound on all points of $\mathcal{H}_{\mathfrak{o}}(\mathfrak{o}_{K,S})$, and then enumeration of bounded height points concludes the algorithm.

So we see that, to prove Theorem 9.1.2, it suffices to prove Theorem 9.5.1. Let us describe how.

The key point will be that Lemma 9.3.3 is explicit, and we will use this in the following way: two Galois representations valued in $\mathrm{GL}_2(\mathfrak{o}_{F,\lambda})$ with irreducible residual representations (valued in $\mathrm{GL}_2(\mathfrak{o}_F/\lambda)$) which are isomorphic after tensoring up to F_λ are already isomorphic. Indeed, irreducibility of the residual representation forces there to be a unique Galois-stable lattice up to scaling.

The key difficulty (which is somewhat unrelated) is the fact that the absolute Hodge conjecture is completely nonquantitative — it asserts that $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ acts through a finite quotient on a Hodge class, but of course it asserts nothing about the order of this quotient.

Let us now sketch a proof of Theorem 9.5.1. We give only a sketch to avoid dealing with (explicitly) small primes.

[27] Of course one expects that this outcome never occurs!

Sketch of proof of Theorem 9.5.1. Let f be a parallel weight 2 Hilbert modular eigencuspform over L with level dividing $\mathfrak{n}$. We may assume that f is not CM (i.e. that f is not the automorphic induction of an algebraic Hecke character associated to a CM abelian variety with isogeny class defined over an imaginary CM extension of L), since otherwise it is simple to construct A_f/L. Write F for the number field generated by the Hecke eigenvalues of f. Note that $F/\mathbb{Q}$ is CM. In our case of interest $F/\mathbb{Q}$ is totally real, so that in this sketch we will assume $F/\mathbb{Q}$ is totally real as well.

Let $\sigma : L \hookrightarrow \mathbb{R}$.

Blasius in fact constructs a motive associated not to f, but rather to (a transfer to a unitary group over L compact at all places but σ of) its symmetric square lifting $\mathrm{Sym}^2 f$ to GL_3/L. Using the Kuga-Satake construction he produces a corresponding abelian variety over $\mathbb{C}$, then over $\overline{\mathbb{Q}}$ by the absolute Hodge conjecture and a countability argument, and then over L by a restriction of scalars argument. Crucially, the absolute Hodge conjecture also allows him to read off the λ-adic representations of the abelian variety, so that one can conclude by comparing to the representations associated to f.

We would like to bound the height of this abelian variety, in which case we cannot work up to isogeny (i.e. with motives with $\mathbb{Q}$-coefficients) as in Blasius — we must work with integral coefficients throughout. Recalling the comparison of the Faltings and naïve (i.e. that arising from the Baily-Borel/Satake compactification) height, as well as the fact that all points in question are S-integral (with S an explicit finite set of places of L), it will suffice to bound various expressions in terms of the periods (at places at infinity and in S) of this A_f/L. Because we are only giving a sketch, for simplicity we will only discuss the places at infinity.

The aforementioned motive (over L with $\mathbb{Q}$-coefficients) associated to $\mathrm{Sym}^2 f$ lies in H^2 of a Picard modular surface, say X_σ/L. Its corresponding projector π is an F-linear combination of (two-dimensional) Hecke correspondences in $X_\sigma \times X_\sigma$. Let $k \in \mathbb{Z}^+$ be such that $\theta := k \cdot \pi$ is an $\mathfrak{o}_F$-linear combination of such — note that k can be explicitly computed in terms of discriminants of characteristic polynomials of Hecke operators. We find that $\theta^2 = k \cdot \theta$. We will use θ to work integrally.

Via θ we construct a rank-3 Hodge structure $\tilde{\Lambda}_\sigma$, namely[28] the lattice $\tilde{\Lambda}_\sigma := \theta_* \cdot H^2(X_\sigma(\mathbb{C}), \mathbb{Z})(1)$ (note the twist), the polarization arising from the cup product on $H^2(X_\sigma(\mathbb{C}), \mathbb{Z})(1)$, and the Hodge decomposition arising from $H^2(X_\sigma(\mathbb{C}), \mathbb{C})(1) \simeq H^{2,0}(X_\sigma(\mathbb{C}))(1) \oplus H^{1,1}(X_\sigma(\mathbb{C}))(1) \oplus H^{0,2}(X_\sigma(\mathbb{C}))(1)$.

We then consider the even Clifford algebra of $\tilde{\Lambda}_\sigma$ (note that we are working in the category of $\mathfrak{o}_F$-modules), $C^+(\tilde{\Lambda}_\sigma)$. Blasius proves in the course of his argument that $C^+(\tilde{\Lambda}_\sigma \otimes_{\mathbb{Z}} \mathbb{Q})$ is the split quaternion algebra over F — specifically, he proves that $C^+(\tilde{\Lambda}_\sigma \otimes_{\mathbb{Z}} \mathbb{Q}) \cong \mathrm{End}_{\mathbb{Q}}(V_\sigma)$, with V a rank-2 F-Hodge structure of type $\{(1,0),(0,1)\}$. Now, by the classification (just act on a nonzero vector) of maximal orders in split quaternion algebras, it follows that the order $C^+(\tilde{\Lambda}_\sigma)$ lies inside a maximal order of the form $\mathrm{End}_{\mathfrak{o}_F}(\Lambda_\sigma)$, with $\Lambda_\sigma \subseteq V_\sigma$ a rank-2 $\mathfrak{o}_F$-Hodge structure of type $\{(1,0),(0,1)\}$ — i.e. $\Lambda_\sigma \subseteq V_\sigma$ is an $\mathfrak{o}_F$-lattice. Note that one can (by computing a discriminant) explicitly bound the index of $C^+(\tilde{\Lambda}_\sigma)$ in $\mathrm{End}_{\mathfrak{o}_F}(\Lambda_\sigma)$.

By Riemann there is an abelian variety $A_\sigma/\mathbb{C}$ with

$$H^1(A_\sigma(\mathbb{C}), \mathbb{Z}) \cong \Lambda_\sigma$$

as Hodge structures. Following Blasius, by the absolute Hodge conjecture (and a countability argument) it follows that without loss of generality A_σ is defined over $\overline{\mathbb{Q}}$. Let L'/E be a sufficiently large number field so that A_σ is defined over L'.

[28] We omit all analytifications — thus $X_\sigma(\mathbb{C})$ should read $(X_\sigma)^{\mathrm{an.}}(\mathbb{C})$, etc. Note that we use the embedding $\sigma : L \hookrightarrow \mathbb{R} \hookrightarrow \mathbb{C}$ to base change X_σ up to $\mathbb{C}$ — using another embedding would (by a very special case of a theorem of Borovoi and Milne [30, 69] on conjugation of Shimura varieties) amount to replacing σ by another embedding $L \hookrightarrow \mathbb{R}$.

Now, the above isomorphism of Hodge structures ultimately (via Künneth) corresponds to a Hodge class specifying the map (remember that $\tilde{\Lambda}_\sigma$ has rank 3 over $\mathfrak{o}_F$)

$$H^2(X_\sigma(\mathbb{C}), \mathbb{Z})^{\otimes 2} \twoheadrightarrow \tilde{\Lambda}_\sigma^{\otimes 2} \to C^+(\tilde{\Lambda}_\sigma) \hookrightarrow \mathrm{End}_{\mathfrak{o}_F}(\Lambda_\sigma) \simeq \Lambda_\sigma \otimes_{\mathfrak{o}_F} \Lambda_\sigma^\vee,$$

say $\omega \in H^6((X_\sigma \times X_\sigma \times A_\sigma \times A_\sigma^\vee)(\mathbb{C}), \mathbb{Z})$. By the absolute Hodge conjecture we see that $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ acts on ω through a finite-index quotient. Without loss of generality L'/E is sufficiently large so that $\mathrm{Gal}(\mathbb{C}/L') \cdot \omega = \omega$ and so that A_σ is defined over L'. As usual, via the comparison isomorphisms $H^*_{\text{sing.}}(X_\sigma(\mathbb{C}), \mathbb{Z}_\ell) \simeq H^*_{\text{ét.}}((X_\sigma)_{/\overline{\mathbb{Q}}}, \mathbb{Z}_\ell)$ one sees that ω has ℓ-adic realizations, say ω_ℓ, for all ℓ.

We repeat that in this sketch we will ignore (explicitly) sufficiently small primes (e.g. those dividing k, or the index of the order generated by the Hecke eigenvalues of f in $\mathfrak{o}_F$, or the index of $C^+(\tilde{\Lambda}_\sigma) \subseteq \mathrm{End}_{\mathfrak{o}_F}(\Lambda_\sigma)$, etc.). For ℓ (explicitly) sufficiently large, ω_ℓ furnishes an isomorphism (writing i_λ for the embedding $\mathfrak{o}_F \hookrightarrow \mathfrak{o}_{F,\lambda}$ corresponding to the prime $\lambda \subseteq \mathfrak{o}_F$ of F with $\lambda | (\ell)$)

$$\begin{aligned}
\bigoplus_{\lambda|(\ell)} \mathrm{Sym}^2 \rho_{A_\sigma,\lambda} &\cong \mathrm{Sym}^2 H^1_{\text{ét.}}((A_\sigma)_{/\overline{\mathbb{Q}}}, \mathbb{Z}_\ell) \\
&\cong \bigoplus_{\lambda|(\ell)} i_\lambda(\theta)_* \cdot H^2_{\text{ét.}}((X_\sigma)_{/\overline{\mathbb{Q}}}, \mathbb{Z}_\ell)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')} \\
&\cong \bigoplus_{\lambda|(\ell)} \mathrm{Sym}^2 \rho_{f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')},
\end{aligned}$$

and indeed we see similarly that $i_\lambda(\theta)_* \cdot H^2_{\text{ét.}}((X_\sigma)_{/\overline{\mathbb{Q}}}, \mathbb{Z}_\ell)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')} \cong \mathrm{Sym}^2 \rho_{A_\sigma,\lambda}$, so that $\mathrm{Sym}^2 \rho_{A_\sigma,\lambda} \cong \mathrm{Sym}^2 \rho_{f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}$ for all λ with (explicitly) sufficiently large norm. Note that we have absolutely no control on L', but on the other hand we know the Hodge structure of A_σ exactly (here we implicitly use a theorem of Borovoi-Milne [30, 69] on the conjugation of Shimura varieties).

Our goal is to compare the Hodge structures of A_σ/L' and A_f/L. What allows us to do this is the following, which is the main point. By Lemma 9.3.3 (we emphasize that it is completely essential that this lemma gives an explicitly computable bound), for (explicitly) sufficiently large $\lambda \subseteq \mathfrak{o}_F$, we have that $\overline{\rho}_{f,\lambda}$ is irreducible as a representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ (and indeed its image contains $\mathrm{SL}_2(\mathbb{F}_\ell)$ where $\lambda|(\ell)$). This means that the rational representation $\rho_{f,\lambda} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda$ has a unique (up to scaling) invariant lattice! In other words, because (by definition of A_f)

$$\rho_{A_f,\lambda} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda \cong \rho_{f,\lambda} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda,$$

it follows automatically that one has

$$\rho_{A_f,\lambda} \cong \rho_{f,\lambda}$$

— i.e., the isomorphism holds integrally. (We continue to ignore explicitly small primes in this sketch, but note that one instead concludes for small λ that there are maps $\rho_{f,\lambda} \twoheadrightarrow \rho_{A_f,\lambda}$ and $\rho_{A_f,\lambda} \twoheadrightarrow \rho_{f,\lambda}$ with kernels of explicitly bounded size.)

Therefore we see that the base change A_f/L' satisfies

$$\mathrm{Sym}^2 \rho_{A_\sigma,\lambda} \cong \mathrm{Sym}^2 \rho_{A_f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}.$$

Moreover, taking determinants of this isomorphism, we see that $\det \rho_{A_\sigma,\lambda}$ and $\det \rho_{A_f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}$ differ by a cubic character valued in $\mathfrak{o}_F^\times$. Since in our case $F/\mathbb{Q}$ is totally real, it follows that

$$\det \rho_{A_\sigma,\lambda} \cong \det \rho_{A_f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}.$$

Let us first see that

$$\rho_{A_\sigma,\lambda} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda \cong (\rho_{A_f,\lambda} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}.$$

Write

$$\begin{aligned}\rho &:= \rho_{A_\sigma,\lambda},\\ \widetilde{\rho} &:= \rho_{A_f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')},\\ \rho^0 &:= \rho \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda,\\ \widetilde{\rho}^0 &:= \widetilde{\rho} \otimes_{\mathfrak{o}_{F,\lambda}} F_\lambda.\end{aligned}$$

We first claim that $\rho^0 \cong \widetilde{\rho}^0$. To see this, first note that (by cuspidality and the fact that f is not CM) $\widetilde{\rho}^0$ is irreducible. Similarly, $\mathrm{Sym}^2\rho^0 \cong \mathrm{Sym}^2\widetilde{\rho}^0$ is also irreducible (because f is not CM — i.e. by cuspidality on GL_3/L).

Our claim is that

$$\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\rho^0, \widetilde{\rho}^0) \neq 0,$$

in other words that the four-dimensional F_λ-adic representation $\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)$ has nontrivial invariants.

To see this, first note that $(\rho^0)^* \simeq \rho^0 \otimes (\det \rho^0)^{-1}$, and similarly for $\widetilde{\rho}^0$ (and recall that $\det \rho \cong \det \widetilde{\rho}$, so that e.g. $(\rho^0)^{\otimes 2} \simeq \mathrm{Sym}^2\rho^0 \oplus \det \rho^0 \cong (\widetilde{\rho}^0)^{\otimes 2}$). Next observe that

$$\begin{aligned}\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)^{\otimes 2} &\simeq (\rho^0)^{\otimes 2} \otimes (\widetilde{\rho}^0)^{\otimes 2} \otimes (\det \rho^0)^{-2}\\ &\simeq (\mathrm{Sym}^2\rho^0 \oplus \det \rho^0) \otimes (\mathrm{Sym}^2\widetilde{\rho}^0 \oplus \det \widetilde{\rho}^0) \otimes (\det \rho^0)^{-2}\\ &\cong \mathrm{triv} \oplus (\mathrm{Sym}^2\rho^0 \otimes (\det \rho^0)^{-1})^{\oplus 2} \oplus \mathrm{End}_{F_\lambda}(\mathrm{Sym}^2\rho^0 \otimes (\det \rho^0)).\end{aligned}$$

Thus, because the first (obvious) and last (because of the identity endomorphism) terms have trivial summands, it follows that:

$$\dim_{F_\lambda} \mathrm{End}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)) \geq 2.$$

In other words, $\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)$ has at least two irreducible summands.

On the other hand (arguing exactly as in Blasius), the six-dimensional representation $\Lambda^2 \mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)$ decomposes as:

$$\begin{aligned}\Lambda^2 \mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0) &\simeq \Lambda^2(\rho^0 \otimes \widetilde{\rho}^0 \otimes (\det \rho^0)^{-1}) \\ &\cong (\mathrm{Sym}^2 \rho^0 \otimes (\det \rho^0)^{-1})^{\oplus 2}\end{aligned}$$

(the "cross-term" in the previous isomorphism). By hypothesis both three-dimensional summands are irreducible, whereas if $\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)$ were to decompose into at least three irreducible summands (and thus certainly decompose into the direct sum of two two-dimensional representations), $\Lambda^2 \mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)$ would have at least two one-dimensional summands, a contradiction.

Therefore

$$\dim_{F_\lambda} \mathrm{End}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\mathrm{Hom}_{F_\lambda}(\rho^0, \widetilde{\rho}^0)) = 2,$$

i.e.

$$\dim_{F_\lambda}(\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\rho^0, \widetilde{\rho}^0)) = 1,$$

so that there is a $\mathrm{Gal}(\overline{\mathbb{Q}}/L')$-equivariant F_λ-linear isomorphism $\rho^0 \cong \widetilde{\rho}^0$.

Thus we have seen that $\rho^0 \cong \widetilde{\rho}^0$. In particular (for simplicity in this sketch we will only deal with the case that A_f/L is geometrically simple, and where $\mathfrak{o}_F \simeq \mathrm{End}_{\overline{\mathbb{Q}}}(A_f)$, the general case follows similarly by replacing A_f by the Serre tensor product $A_f \otimes_{\mathrm{End}_L(A_f)} \mathfrak{o}_F$ and decomposing A_f into an n-th power with $n \leq [F : \mathbb{Q}]$) A_σ/L' and A_f/L' are L'-isogenous.

Now let us work integrally. Consider the $\mathfrak{o}_F$-module $\mathrm{Hom}_{L'}(A_\sigma, A_f)$. A choice of L'-isogeny $A_\sigma \sim_{L'} A_f$ shows that, as an $\mathfrak{o}_F$-module, $\mathrm{Hom}_{L'}(A_\sigma, A_f) \cong I$ for an explicit representative $I \subseteq \mathfrak{o}_F$ of an element of $\mathrm{Cl}(\mathfrak{o}_F)$.

By replacing A_f with the Serre tensor product $A_f \otimes_{\mathfrak{o}_F} I^{-1}$ if necessary (note that I lies in an explicit set of representatives for $\mathrm{Cl}(\mathfrak{o}_F)$ — alternatively note that

$A_f \sim_L A_f \otimes_{\mathfrak{o}_F} I^{-1}$, so that by Masser-Wüstholz it suffices to bound the height of the latter as well), we may assume that $\mathrm{Hom}_{L'}(A_\sigma, A_f) \cong \mathfrak{o}_F$ as $\mathfrak{o}_F$-modules. Let $\varphi : A_\sigma \twoheadrightarrow A_f$ be a generator of $\mathrm{Hom}_{L'}(A_\sigma, A_f)$.

We claim that φ is an isomorphism (were we including the analysis at small primes we would conclude that φ is an L'-isogeny with explicitly bounded degree, which of course suffices for a height bound). To see this it suffices to show that $\deg \varphi = 1$, i.e. that φ induces isomorphisms $\varphi : \rho_{A_\sigma,\lambda} \simeq \rho_{A_f,\lambda}$ for all λ.

Again, we will only deal with λ (explicitly) sufficiently large.

Now because $\mathrm{Sym}^2\rho^0$ is irreducible, it follows that (this is a restatement of Schur's lemma)

$$\dim_{F_\lambda}(\mathrm{End}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\mathrm{Sym}^2\rho^0)) = 1.$$

Therefore all isomorphisms $\mathrm{Sym}^2\rho^0 \cong \mathrm{Sym}^2\widetilde{\rho}^0$ must be F_λ-multiples of a single such — for example, $\mathrm{Sym}^2\varphi$! However we know there is an (integral) isomorphism $\mathrm{Sym}^2\rho \cong \mathrm{Sym}^2\widetilde{\rho}$. Thus (by scaling away units) there is an $n \in \mathbb{Z}$ with $\pi^n \cdot \mathrm{Sym}^2\varphi : \mathrm{Sym}^2\rho \simeq \mathrm{Sym}^2\widetilde{\rho}$ an isomorphism, where $\pi \in \mathfrak{o}_{F,\lambda}$ is a uniformizer of λ.

However because $\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\rho, \widetilde{\rho}) = \mathfrak{o}_{F,\lambda} \cdot \varphi$ (this was a manifestation of the Tate conjecture), we see that there must be a $v \in \mathfrak{o}_{F,\lambda}^{\oplus 2}$ (aka ρ) with $\pi^{-1} \cdot \varphi(v) \notin \mathfrak{o}_{F,\lambda}^{\oplus 2}$ (aka $\widetilde{\rho}$) — indeed, otherwise $\pi^{-1} \cdot \varphi$ would be in $\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\rho, \widetilde{\rho})$.

By explicit calculation or otherwise one sees that similarly

$$\pi^{-1} \cdot (\mathrm{Sym}^2\varphi)(v^2) \notin \mathfrak{o}_{F,\lambda}^{\oplus 3} \text{ (aka } \mathrm{Sym}^2\widetilde{\rho}\text{)}.$$

Thus (since $\mathrm{Sym}^2\varphi$ does preserve integrality)

$$\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/L')}(\mathrm{Sym}^2\rho, \mathrm{Sym}^2\widetilde{\rho}) = \mathfrak{o}_{F,\lambda} \cdot \mathrm{Sym}^2\varphi.$$

Thus $\mathrm{Sym}^2\varphi : \mathrm{Sym}^2\rho \simeq \mathrm{Sym}^2\widetilde{\rho}$ must be an isomorphism.

Thus $\varphi : \rho \to \widetilde{\rho}$ must be an isomorphism. Indeed, φ is automatically surjective (since the corresponding isogeny $\varphi : A_\sigma \twoheadrightarrow A_f$ is). Moreover if $\varphi(v) = 0$ then $(\mathrm{Sym}^2\varphi)(v^2) = 0$, so that $v^2 = 0$ in $\mathrm{Sym}^2\rho$, so that (by explicit calculation) $v = 0$.

In other words $\deg\varphi$, the degree of the isogeny $\varphi : A_\sigma \twoheadrightarrow A_f$, must be prime to λ. Since this holds for all (recall that we are ignoring explicitly small primes) λ, it follows that $\deg\varphi = 1$ and so $\varphi : A_\sigma \simeq A_f$ is an L'-isomorphism. (Taking explicitly small primes into account would lead us to conclude that $\deg\varphi$ is explicitly bounded, which also suffices.)

We conclude that $A_\sigma(\mathbb{C}) \cong A_f(\mathbb{C})$, where we have used (an extension of) the embedding $\sigma : L \hookrightarrow \mathbb{R} \hookrightarrow \mathbb{C}$ to base change to $\mathbb{C}$. Therefore we can read the periods of A_f/L corresponding to σ off from those of A_σ — that is, the periods of A_f/L corresponding to σ are just those of Λ_σ, which we have determined explicitly, or else a Serre tensor product thereof with an explicit representative of one of the finitely many ideal classes of $\mathfrak{o}_F$. Doing this for all $\sigma : L \hookrightarrow \mathbb{R}$, we have therefore explicitly bounded the (contribution at infinity — recall that we are ignoring primes in S — of the) naïve height of A_f/L induced by the Baily-Borel/Satake compactification of A_g, and thus the stable Faltings height $h(A_f)$, as desired.

This completes the sketch of the proof of Theorem 9.5.1. □

Chapter 10

$C(K)$: Cohen-Wolfart.

Abstract.

We give an algorithm that, on input (C, K) with C/K a smooth projective hyperbolic curve over a CM field K admitting a Belyi map over K with sufficiently divisible ramification indices, outputs $C(K)$, along with an unconditional certificate of correctness of the output. Assuming the existence of motives associated to "parallel weight two" automorphic representations of GL_2 over a CM field, we prove this algorithm always terminates in finite time.

The key point is to use the arguments of Chapter 9 and a construction of Cohen-Wolfart to produce a family (namely a pullback of a hypergeometric family) of $\mathrm{GL}_2(\mathbb{Q}(\zeta_N))$-type abelian varieties $A \to C$ defined over K and with endomorphisms by $\mathbb{Z}[\zeta_N]$ defined over $K(\zeta_N)$. We of course replace our use of potential modularity results over totally real fields in Chapter 9 with the recent breakthrough providing such results over imaginary CM fields.

10.1 Introduction.

So we have seen how to algorithmically find the rational points on (smooth projective hyperbolic) curves over number fields by assuming standard motivic conjectures (Chapter 7) or strong modularity conjectures (remarked in e.g. Chapter 9). We have seen that the latter modularity conjectures can be weakened if we restrict to hyperelliptic curves, thanks to a theorem of Bogomolov-Tschinkel (Chapter 8). We have also seen that we can obtain completely unconditional algorithms if our curve instead happens to be defined over an odd-degree totally real field and also to admit a map to some Hilbert modular variety that is defined over an odd-degree totally real field (Chapter 9). But it is not clear that these constitute progress on our original goal, namely to give effective height bounds in Faltings' Theorem, at least over e.g. $\mathbb{Q}$. Is there an easily-checked criterion that implies that one can find the rational points on a given curve using Theorems 9.1.1 and 9.1.2 of Chapter 9?

The purpose of this chapter is twofold. First, in a sense we explain why the results of Chapter 9 are interesting absent such a criterion. Second, we reduce the full problem of giving effective bounds in Faltings' Theorem for an explicit and large class of curves over a CM field (in particular, over $\mathbb{Q}$) to a statement about "parallel weight 2" cuspidal automorphic representations of GL_2 over a CM field. This statement follows from the existence of abelian varieties associated to such automorphic representations (or even motives associated to some functorial lift thereof, since we may argue as in Section 9.5 of Chapter 9), for example, so that the problem of effectivizing Faltings' Theorem for such curves over CM fields reduces to a completely standard conjecture in the Langlands program.

10.1.1 Main theorem.

In this chapter we prove the following theorem.

Theorem 10.1.1. *There is a finite-time algorithm that, on input $(\mathfrak{o}, K, S)$ with $\mathfrak{o}$ an order in a CM field $F/\mathbb{Q}$, $K/\mathbb{Q}$ a CM field, and S a finite set of places of K, outputs a finite set Π of pairs (L, π), where L/K is an odd-degree CM extension totally split at infinity (thus*

$L/\mathbb{Q}$ is CM), and π is a weight zero[1] *cuspidal automorphic representation of* GL_2/L*, such that, for all $[F:\mathbb{Q}]$-dimensional abelian varieties A/K with good reduction outside S and admitting $\mathfrak{o} \hookrightarrow \mathrm{End}_K(A)$, there is an $(L,\pi) \in \Pi$ for which one has that:*

- *the ring of Hecke eigenvalues of π is a subring of $\mathfrak{o}$,*
- *and that*

$$(\rho_{A,\lambda} \otimes_{\mathfrak{o}_\lambda} F_\lambda)|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)} \cong \rho_{\pi,\lambda} \otimes_{\mathfrak{o}_\lambda} F_\lambda$$

  for all (explicitly) sufficiently large λ, where $\rho_{A,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_\lambda)$ are the λ-adic representations associated to $T_\lambda(A) := \varprojlim A[\lambda^n]$, and $\rho_{\pi,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/L) \to \mathrm{GL}_2(\mathfrak{o}_\lambda)$ are the λ-adic representations associated to π.

Of course, if one knew the existence of abelian varieties A_π/L associated[2] to such π, then, by simply searching through abelian varieties B/L of larger and larger height until we find A_π/L, we would be able to produce a height bound (via Masser-Wüstholz) for all A/K of $\mathrm{GL}_2(\mathfrak{o})$-type over K with good reduction outside S. By the construction of Cohen-Wolfart discussed in Section 10.2, this would then produce a computable height bound on all points of $C(K)$ given $(K, C/K)$ with C/K Belyi-hyperbolic over K, in the following sense.

Definition 10.1.2. *Let $K/\mathbb{Q}$ be a number field. Let $e_0, e_1, e_\infty \in \mathbb{Z}^+$ be such that $\frac{1}{e_0} + \frac{1}{e_1} + \frac{1}{e_\infty} < 1$. A smooth projective hyperbolic curve C/K is called*[3] Belyi-hyperbolic over K (with corresponding Belyi map β and parameters (e_0, e_1, e_∞)) *if $\beta : C \to \mathbb{P}^1$ is defined over K, is a Belyi map* [13] *(i.e. is ramified only over 0, 1, and ∞), and is such that all ramification indices over 0 are divisible by e_0, all ramification indices over 1 are divisible by e_1, and all ramification indices over ∞ are divisible by e_∞.*

[1] We switch to this nomenclature (which is a ρ-shift from our use of "parallel weight 2") in order to match the ten-author paper [3].

[2] Here we allow the possibility that $\rho_{A_\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q} \cong (\rho_{\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q})^{\oplus d}$ with $d \in \mathbb{Z}^+$ (or even just that there is a nonzero map $\rho_{\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q} \to \rho_{A_\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q}$), since it makes no difference to the algorithms.

[3] We will often abbreviate the statement that C/K is smooth, projective, and Belyi-hyperbolic over K to the statement that C/K is a smooth projective Belyi-hyperbolic curve.

We summarize the above remark as follows.[4]

Corollary 10.1.3. *There is a finite-time algorithm that, on input $(K, C/K)$ with $K/\mathbb{Q}$ CM and C/K a smooth projective Belyi-hyperbolic curve, outputs $(\beta, e_0, e_1, e_\infty, a, b, c, N, A, B, C, \Pi)$, where $(\beta, e_0, e_1, e_\infty, a, b, c, N, A, B, C)$ are as in Definition 10.1.2 and Section 10.2, and Π is a finite set of pairs (L, π) with L/K a CM extension totally split at infinity and π a weight zero cuspidal automorphic representation of GL_2/L, such that, for every $P \in C(K)$, writing A_P/K for the hypergeometric abelian variety with parameters (N, A, B, C) corresponding to $\beta(P) \in \mathbb{P}^1(e_0, e_1, e_\infty)$, there is an $(L_P, \pi_P) \in \Pi$ for which*[5]

$$L(s, \rho_{A_P/K(\zeta_N),\lambda}) = L(s, \pi_P)$$

for all (explicitly) sufficiently large primes $\lambda \subseteq \mathbb{Z}[\zeta_N]$ of $\mathbb{Q}(\zeta_N)$.

In particular, there is an algorithm that, on input $(K, C/K)$ with $K/\mathbb{Q}$ CM and C/K a smooth projective Belyi-hyperbolic curve, outputs $C(K)$, along with unconditional proof of correctness of the output. Assuming the existence of abelian varieties associated[6] *to weight zero cuspidal automorphic representations of GL_2 over CM fields, this algorithm always terminates in finite time.*

We note that, by a similar search through motives and by mimicking the proof of Theorem 9.1.2 given in Section 9.5 of Chapter 9, one should be able to derive the same conclusion (modulo allowing for the output of an unconditional disproof of a standard conjecture, just as in Theorem 9.1.2 as compared to Theorem 9.1.1) from

[4] We must comment on the following. From Cohen-Itzykson-Wolfart's [41], it would seem that every smooth projective curve C/K is Belyi-hyperbolic over K as defined above, and thus that the hypergeometric family construction we proceed to give (and thus our results about rational points via potential modularity) go through for every such C/K. However unfortunately there is a mistake in [41]: implicit in the last line of the statement of Proposition 1 in [41] is an incorrect step. In their notation, the quotient $H\backslash\mathfrak{h}$ is not necessarily isomorphic to $\mathscr{C}$ — it is in fact an orbifold such that, upon forgetting multiplicities, one gets $\mathscr{C}$. This invalidates the argument given to prove their Theorem 1.

[5] Note that this determines $L(s, A_P/K)$ explicitly in terms of π_P.

[6] Again, we allow the possibility that $\rho_{A_\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q} \cong (\rho_{\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q})^{\oplus d}$ with $d \in \mathbb{Z}^+$ (or even just that there is a nonzero map $\rho_{\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q} \to \rho_{A_\pi,\lambda} \otimes_{\mathbb{Z}} \mathbb{Q}$), since it makes no difference to the algorithms.

the existence of motives associated to suitable functorial lifts of π as well — in the proof of Theorem 9.1.2 we used the existence of motives associated to $\mathrm{Sym}^2\pi$ when $L/\mathbb{Q}$ is totally real.[7]

10.1.2 Main idea.

It is clear what to do to prove Theorem 10.1.1, and arguably the main idea of this chapter is to use the construction of Cohen-Wolfart detailed in Section 10.2 to deduce Corollary 10.1.3. Nonetheless let us explain the proof of Theorem 10.1.1.

The point is that we may simply imitate the algorithm in Section 9.2.1 of Chapter 9 in this context, because Lemma 9.3.3 still applies (we could also use only Lemma $7.1.3$ of the ten-author paper [3] and then imitate the algorithm sketched in Section 9.1.3 instead). We must of course replace the TaylorMoretBaillyExtensions subroutine, which closely followed Section 5 of Snowden's [95], with a similar subroutine closely following Section 7 of the ten-author paper [3], but the core of the argument is the same: an extension produced by invoking the Moret-Bailly theorem (for example $F_2^+/F_1^+F^+$, in the notation of the proof of Corollary $7.1.11$ of the ten-author paper [3] — see their page 186, where they invoke Proposition $3.1.1$ of [12]) is computable in finite time.

10.2 A construction of Cohen-Wolfart.

Let us now detail the construction of Cohen-Wolfart (see specifically Section 3.3 of their [40], though note that they work over $\overline{\mathbb{Q}}$).

[7] It is tempting to guess that one can bound the periods of an A_π/L in terms of the subspace corresponding to π in the singular cohomology of the corresponding locally symmetric space, for example (in the totally real case this would correspond to using the full tensor induction $(\otimes\text{-Ind})_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}^{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}\rho_{\pi,\lambda}$ instead of $\mathrm{Sym}^2\rho_{\pi,\lambda}$ for the arguments of Section 9.5 of Chapter 9).

Let $K/\mathbb{Q}$ be a number field. Let C/K be a smooth projective hyperbolic curve that is Belyi-hyperbolic over K. Let $\beta : C \to \mathbb{P}^1$ be the corresponding Belyi map, and $e_0, e_1, e_\infty \in \mathbb{Z}^+$ the corresponding divisors of the respective ramification indices. Without loss of generality $e_0 \leq e_1 \leq e_\infty$. We regard this as a map $C \to \mathbb{P}^1(e_0, e_1, e_\infty)$, where $\mathbb{P}^1(e_0, e_1, e_\infty)$ is an orbifold $\mathbb{P}^1$ with $0, 1, \infty$ given multiplicities

$$p := \frac{1}{e_0}, q := \frac{1}{e_1}, r := \frac{1}{e_\infty},$$

respectively. In other words $\mathbb{P}^1(e_0, e_1, e_\infty) := \mathfrak{h}/\Delta(e_0, e_1, e_\infty)$, the quotient of the upper-half plane by the standard triangle group.

The point is that $\mathbb{P}^1(e_0, e_1, e_\infty)$ has a standard family (first noticed perhaps by Weil, but certainly by Deligne-Mostow, Wolfart, Wüstholz, Darmon, etc.) of abelian varieties on it, namely the hypergeometric family, which we will now detail.[8]

Specifically, let $a, b, c \in \mathbb{Q}^+$ solve the following system of equations:

$$p = |1 - c|, q = |c - a - b|, r = |a - b|.$$

Let N be the least common multiple of the denominators of a, b, and c, when written in lowest terms. Let

$$A := N \cdot (1 - b), B := N \cdot (1 + b - c), C := N \cdot a.$$

The hypergeometric family corresponding to (e_0, e_1, e_∞) is then defined as follows: given $z \in \mathbb{P}^1 - \{0, 1, \infty\}$, let $X_z^{(N,A,B,C)}/\mathbb{Q}(z)$ be the desingularization of the curve

$$C_z^{(N,A,B,C)} : y^N = x^A \cdot (1 - x)^B \cdot (1 - z \cdot x)^C.$$

[8] See Archinard's [8] for a careful and thorough discussion of the construction.

Let $A_z/\mathbb{Q}(z)$ be connected component of the intersection of the kernels of $\operatorname{Jac} X_z^{(N,A,B,C)} \to \operatorname{Jac} X_z^{(d,A,B,C)}$ (arising from the obvious maps $C_z^{(N,A,B,C)} \to C_z^{(d,A,B,C)}$ via $(x,y) \mapsto (x, y^{\frac{N}{d}})$) over all $d|N$ with $d < N$.

Then $\dim A_z = \varphi(N)$ and $\mathbb{Z}[\zeta_N] \hookrightarrow \operatorname{End}_{\mathbb{Q}(\zeta_N,z)}(A_z)$, with ζ_N acting by $(x,y) \mapsto (x, \zeta_N \cdot y)$ at the level of $C_z^{(N,A,B,C)}$.

In other words A_z is of $\mathrm{GL}_2(\mathbb{Z}[\zeta_N])$-type over $\mathbb{Q}(\zeta_N, z)$! Moreover by a period calculation (see e.g. the ends of Sections 3.1 or 3.2 of Cohen-Wolfart's [40] — the key point is that Schwarz triangle maps continue over $0, 1, \infty$) one sees that the family, which we have a priori only defined over $\mathbb{P}^1 - \{0, 1, \infty\}$, extends to a family over $\mathbb{P}^1(e_0, e_1, e_\infty)$.

Pulling back by the Belyi map β, we find that over our smooth projective Belyi-hyperbolic C/K there is a family $A \to C$ of abelian varieties of $\mathrm{GL}_2(\mathbb{Z}[\zeta_N])$-type over $K(\zeta_N)$.[9] We note that, just as in Chapters 7, 8, and 9, the corresponding map $C \to \mathcal{A}_{\varphi(N)}$ extends to S-integral models for S an explicitly sufficiently large finite set of places of K.

Therefore, to determine $C(K)$, it suffices to determine instead the finitely many $\varphi(N)$-dimensional abelian varieties A/K with good reduction outside S and admitting a map $\mathbb{Z}[\zeta_N] \hookrightarrow \operatorname{End}_{K(\zeta_N)}(A)$. This is tantalizingly close to being the situation considered in Chapter 9. The crucial difference is that, even if we take e.g. $K = \mathbb{Q}$, we must still work with representations $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_N)) \to \mathrm{GL}_2(\mathbb{Z}[\zeta_N]_\lambda)$, because the endomorphism ζ_N of our abelian varieties is only defined over $\mathbb{Q}(\zeta_N)$. Because this then implies that the corresponding automorphic forms are on GL_2 over an imaginary CM field, we will not be able to construct abelian varieties asso-

[9] It is worth noting that these abelian varieties are even more constrained: because of the monodromy of the hypergeometric differential equation, their Mumford-Tate groups lie in the units of the quaternion algebra spanned by $\Delta(e_0, e_1, e_\infty)$ over the trace field of its squares. This further constrains their Galois representations by e.g. Deligne's absolute Hodge theorem, though we do not see a way to take advantage of this extra structure because it occurs over a finite extension (which one can compute by mimicking the arguments in the next chapter, particularly the use of the identity of finite-field hypergeometric functions).

ciated to the corresponding automorphic representations, even though the potential modularity step (thanks to the ten-author paper [3]) is still available.

Nonetheless this explains why the results of Chapter 9 are arguably interesting: we find that *many* curves C/K over a number field support a family of abelian varieties over K of GL_2-type over $K(\zeta_N)$ for some N. While Chapter 9 only allows one to take a totally real extension of scalars (and thus not $K \mapsto K(\zeta_N)$), it is not a priori clear that one cannot do better by modifying the Cohen-Wolfart construction slightly.

10.2.1 Example: Fermat curves.

To orient the reader, and in order to present a crucial point of intuition, we explain the above construction in the particular case of Fermat curves.

First, for $n \geq 4$, the Fermat curve $F_n : x^n + y^n = 1$ is Belyi-hyperbolic over $\mathbb{Q}$: the degree-n^2 map $\beta_n : (x, y) \mapsto x^n$ has all ramification indices over $0, 1, \infty$ equal to n. Thus we may (and will) take $e_0 := e_1 := e_\infty := n$, and so $p = q = r = \frac{1}{n}$. A solution of the system defining a, b, c is given by e.g.

$$a := \frac{n-1}{2n}, b := \frac{n-3}{2n}, \text{ and } c := \frac{n-1}{n}.$$

Let us take n to be an odd prime for clarity. Then $N = n$. Correspondingly,

$$A = \frac{n+3}{2}, B = \frac{n-1}{2}, \text{ and } C = \frac{n-1}{2}.$$

Thus, given e.g. $P =: (u, v) \in F_n(K)$ with $u, v \neq 0, \infty$, we obtain A_P/K as a quotient of the Jacobian of the (desingularization of the) curve

$$y^n = x^{\frac{n+3}{2}} \cdot (1-x)^{\frac{n-1}{2}} \cdot (1 - u^n \cdot x)^{\frac{n-1}{2}}.$$

Moreover A_P acquires endomorphisms by $\mathbb{Z}[\zeta_n]$ over $K(\zeta_n)$, and, because the hypergeometric family is over the projective curve F_n, A_P has explicitly bounded conductor.

However of course it is not clear how to prove a priori that no such P exists when $K = \mathbb{Q}$! But for our purposes it is crucial that the conductor of A_P/K be bounded — i.e. that the family not degenerate at any point of F_n. Note also a key weakness of the technique: as we have seen in Chapter 9, we would be in much better shape were A_P/K of GL_2-type over a totally real field, rather than over the imaginary CM field $K(\zeta_n)$.

We must of course eventually discuss Wiles' proof of Fermat's Last Theorem, so let us do so here. Recall that Wiles uses the Frey curves: the family of elliptic curves over the *punctured* Fermat curve $F_n - \{u^n \in \{0, 1, \infty\}\}$ given by $(u, v) \mapsto E_{(u,v)} : y^2 = x \cdot (x + u^n) \cdot (x - v^n)$.[10] Because the goal is to show that $(F_n - \{u^n \in \{0, 1, \infty\})(\mathbb{Q}) = \emptyset$, this family is perfectly tailored to the problem. However from the perspective of the technique we give here Wiles' proof is a remarkable gambit: by giving up compactness and using a family of abelian varieties over the *punctured* curve (that actually degenerates at the punctures), one gives up a priori control of the conductors of the fibres. To partially rectify this one uses level-lowering to find a congruence between $L(s, E_{(u,v)}/\mathbb{Q})$ and the L-function of a bounded-level modular form, so that at least one has control over the conductor of some congruent form. The miracle that makes the gambit successful is, of course, the nonexistence of nonzero cusp forms of that level — and it is not clear what one would have learned had there been e.g. a five-dimensional space of such forms!

We comment also on Darmon's use of hypergeometric abelian varieties for the solution of generalized Fermat equations. For the moment take $K/\mathbb{Q}$ totally real.

[10] Note that the map $(u, v) \mapsto u^n$ taking $F_n \to \mathbb{P}^1(n, n, n)$ restricts to a map $F_n - \{u^n \in \{0, 1, \infty\}\} \to \mathbb{P}^1 - \{0, 1, \infty\} = \mathbb{P}^1(\infty, \infty, \infty)$, and that the hypergeometric family of abelian varieties over $\mathbb{P}^1 - \{0, 1, \infty\}$ formed by (formally) taking $e_0 := e_1 := e_\infty := \infty$ is of course the Legendre family of elliptic curves.

As we have mentioned, our abelian varieties A_P/K are only GL_2-type over $K(\zeta_n)$, and not over e.g. $K(\zeta_n)^+ = K(\cos\left(\frac{2\pi}{n}\right))$. Darmon realized that, at least when e.g. $e_1 = \infty$ among other hypotheses, one can quotient the curves $X_z^{(N,A,B,C)}$ by an explicit involution so as to produce a family of $\frac{\varphi(N)}{2}$-dimensional abelian varieties with endomorphisms by $\mathbb{Z}[\cos\left(\frac{2\pi}{n}\right)]$ defined over $K(\cos\left(\frac{2\pi}{n}\right))$. However in so doing one trades away compactness, and therefore all a priori control on the conductor of an A_P/K, so that our techniques are not relevant.

10.3 The algorithm and its subroutines.

Let us now precisely specify the algorithm alluded to above. We note that we will be considerably more terse in this chapter, since we are simply imitating the algorithms in Section 9.2 of Chapter 9.

10.3.1 ExplicitPotentialModularity($\mathfrak{o}, K, S$):

Input: $\mathfrak{o}, K, S$, with $\mathfrak{o}$ an order in a CM number field $\mathrm{Frac}\,\mathfrak{o} =: F/\mathbb{Q}$, $K/\mathbb{Q}$ a CM field, and S a finite set of places of K.[11]

Output: Π, a finite set of pairs (L, π) with L/K an odd-degree CM extension totally split at infinity, and π a weight zero cuspidal automorphic representation of GL_2/L, such that, for all $[F : \mathbb{Q}]$-dimensional abelian varieties A/K with good reduction outside S and admitting $\mathfrak{o} \hookrightarrow \mathrm{End}_K(A)$, there is an $(L, \pi) \in \Pi$ for which one has that

$$L(s, \rho_{A,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}) = L(s, \pi)$$

for all (explicitly) sufficiently large $\lambda \subseteq \mathfrak{o}$, where $\rho_{A,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_\lambda)$ is the representation corresponding to $T_\lambda(A) := \varprojlim A[\lambda^n]$, the λ-adic Tate module of A.

[11] We note that we may of course without loss of generality assume that K contains an imaginary quadratic field and that S is the full finite set of primes lying over a finite set of primes of $\mathbb{Z}$. This is a minor technical point related to the stabilization of the trace formula and we will ignore it below.

Algorithm:

1. Let $F := \mathrm{Frac}\,\mathfrak{o}$. Let $g := [F : \mathbb{Q}]$. Let $\mathfrak{m}_K := \prod_{\mathfrak{p}\in S} \mathfrak{p}^{10^{10g}\cdot[K:\mathbb{Q}]}$.

2. $\mathcal{F}$:= output of RelevantExtensionsOverCMFields$(F, K, \mathrm{Nm}\,\mathfrak{m}_K)$.

3. Let, for $L \in \mathcal{F}$, $\mathfrak{m}_L := \prod_{\mathfrak{p}\subseteq\mathfrak{o}_L : (\mathfrak{p}, \mathrm{Nm}\,S)\neq 1} \mathfrak{p}^{10^{10g}\cdot[L:\mathbb{Q}]}$, a product over primes of L dividing $\mathrm{Nm}\,\mathfrak{q}$ for some $\mathfrak{q} \in S$.

4. Let

$$\Pi := \left\{ (L, \pi) \;\middle|\; \begin{array}{c} L \in \mathcal{F}, \pi \text{ a weight zero cuspidal automorphic representation} \\ \text{of } \mathrm{GL}_2/L \text{ of level dividing } \mathfrak{m}_L \end{array} \right\}.$$

5. Output Π.

10.3.1.1 Explanation in words.

This is the evident adaptation of the algorithm of Section 9.2.1 of Chapter 9, without, of course, the height bound (since there is no analogue of the Jacquet-Langlands transfer/cohomology of Shimura curves construction of abelian varieties associated to our π). Again we crudely bound the level of a π associated to (the base change to L of) an $[F : \mathbb{Q}]$-dimensional GL_2-type abelian variety A/K with good reduction outside S using Brumer-Kramer [34] and the fact that the conductors of the λ-adic representations associated to π agree with the form's level (here instead of using Carayol's [35] we invoke results on local-global compatibility in the ten-author paper [3] and Varma's Princeton PhD thesis [102]).

10.3.2 RelevantExtensionsOverCMFields(E, K, s):

Input: E, K, s, with $E/\mathbb{Q}$ and $K/\mathbb{Q}$ CM, and $s \in \mathbb{Z}^+$.

Output: $\mathcal{F}$, a finite set of odd-degree CM extensions L/K that are totally split at infinity such that, for all primes λ of E with $\mathrm{Nm}\,\lambda$ prime to s and representations $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\lambda})$ arising from abelian varieties A/K with good reduction at primes not dividing s and admitting an embedding $E \hookrightarrow \mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Q}$, there is an $L \in \mathcal{F}$ such that $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}$ is the λ-adic representation of a weight zero cuspidal automorphic representation of GL_2/L (with coefficient ring a subring of $\mathfrak{o}_E$).

Algorithm:

1. Let $\lambda_0 \subseteq \mathfrak{o}_E$ be a prime of E with $\mathrm{Nm}\,\lambda_0$ prime to s. Let $\lambda \subseteq \mathfrak{o}_E$ with $\mathrm{Nm}\,\lambda$ prime to s and satisfying $\mathrm{Nm}\,\lambda \geq C_{E,K,s,\lambda_0}$, in the notation of Lemma 9.3.3 of Chapter 9. Let ℓ be the prime of $\mathbb{Z}$ with $\lambda|(\ell)$. Let $S := \{\mathfrak{p}|(s \cdot \ell) : \mathfrak{p} \subseteq \mathfrak{o}_K\}$.

2. Let Ψ_λ be the set of finite-order characters
$$\psi : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathfrak{o}_{E,\lambda}^\times$$
of K that are unramified outside S and primes above ℓ.

3. Let
$$R_\lambda := \left\{ (\rho, \psi) \;\middle|\; \begin{array}{c} \rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_E/\lambda) \text{ odd and unramified outside } S \cup \{\mathfrak{q}|(\ell)\}, \\ \psi \in \Psi_\lambda, \mathrm{SL}_2(\mathbb{F}_\ell) \subseteq \rho(\mathrm{Gal}(\overline{\mathbb{Q}}/K)), \det\rho \equiv \psi \cdot \chi_\ell \pmod{\lambda} \end{array} \right\},$$
where χ_ℓ is the ℓ-adic cyclotomic character.[12]

4. Let, for each $(\rho, \psi) \in R_\lambda$,
$$\tilde{\mathcal{F}}_{(\rho,\psi)} := \text{output of TaylorMoretBaillyExtensionsOverCMFields}(\rho, \psi, K, E, \lambda).$$

[12] Note that if $K/\mathbb{Q}$ is imaginary CM then all ρ are automatically odd.

Let

$$\mathcal{F}_{(\rho,\psi)} := \left\{ \tilde{L}_{(\rho,\psi)}^{H_{(\rho,\psi)}} : \tilde{L}_{(\rho,\psi)} \in \tilde{\mathcal{F}}_{(\rho,\psi)}, H_{(\rho,\psi)} \subseteq \mathrm{Gal}(\tilde{L}_{(\rho,\psi)}/K) \text{ a 2-Sylow subgroup} \right\}.$$

5. Let $\mathcal{F} := \{K\} \cup \bigcup_{(\rho,\psi)\in R_\lambda} \mathcal{F}_{(\rho,\psi)}$.
6. Output $\mathcal{F}$.

10.3.2.1 Explanation in words.

Because Lemma 9.3.3 of Chapter 9 applies verbatim to this situation, the only part of the algorithm of Section 9.2.2 of Chapter 9 that we have needed to change is the use of the TaylorMoretBaillyExtensions subroutine, which follows Section 5 of Snowden's [95] and is thus special to Hilbert modular eigencuspforms.

10.3.3 TaylorMoretBaillyExtensionsOverCMFields($\rho, \psi, K, E, \mathfrak{q}$):

Input: $\bar{\rho}, \psi, K, E, \mathfrak{q}$, with $\mathfrak{q}|(p)$.

Output: $\mathcal{F}$, a finite set of CM Galois extensions L/K that are totally split at infinity for which, for all odd and finitely ramified $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_{E,\mathfrak{q}})$ with Hodge-Tate weights $\{0, -1\}$ under all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$ and such that $\rho \bmod \mathfrak{q} \cong \bar{\rho}$ and $\det \rho = \psi \cdot \chi_p$ with χ_p the p-adic cyclotomic character, one has that there is an $L \in \mathcal{F}$ for which $\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}$ is the $\mathfrak{q}$-adic representation of a weight zero cuspidal automorphic form of GL_2/L.

Algorithm: Rather than copying over the proof of Theorem $7.1.10$ (and thus Corollary $7.1.11$) of the ten-author paper [3] as we did in Section 9.2.3 of Chapter 9 with Snowden's proof of Theorem $5.1.2$ of his [95], we will simply observe that each subsequent extension used in their proof is computable. The only nontrivial point is, again, that the extension (in their notation, $F_2^+/F_1^+F^+$) guaranteed to exist by the Moret-Bailly theorem is evidently computable.

10.3.3.1 Explanation in words.

We have simply referred to the proof of Theorem 7.1.10 in the ten-author paper [3], so there is nothing to explain. We repeat that Moret-Bailly's proof (see Section 4 of [71]) and Rumely's proof [85,86] of Rumely's theorem are both constructive — see the comments in Section 9.2.3.1 of Chapter 9.

10.4 Proof of termination and correctness.

Let us now turn to the proof of termination and correctness of the algorithm specified in Section 10.3.1. In this section we prove the following theorem.

Theorem 10.4.1. *Let $K/\mathbb{Q}$ be CM. Let S be a finite set of places of K. Let $F/\mathbb{Q}$ be CM. Let $\mathfrak{o} \subseteq \mathfrak{o}_F$ be an order in F. Then:*

- *The algorithm specified in Section 10.3.1 terminates on input $(\mathfrak{o}, K, S)$,*
- *and, letting*

  $$\Pi := \text{output of ExplicitPotentialModularity}(\mathfrak{o}, K, S),$$

  for all $[F : \mathbb{Q}]$-dimensional abelian varieties A/K with good reduction outside S that admit a map $\mathfrak{o} \hookrightarrow \operatorname{End}_K(A)$, there is an $(L, \pi) \in \Pi$ such that

  $$L(s, \rho_{A,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}) = L(s, \pi).$$

Evidently this implies Theorem 10.1.1.

10.4.1 Proof of Theorem 10.4.1.

We break the proof into two parts for clarity. First we prove that, if the algorithm terminates, its output is correct. After that we will prove that the algorithm always terminates.

Note that, given the discussion in Chapter 9, and because of the fact that the algorithm of Section 10.3.3 simply refers to the ten-author paper [3], the proofs of both termination and correctness are essentially self-evident. Nonetheless we will give them.

10.4.1.1 Proof of correctness.

Proof of correctness assuming termination. Let A/K be an abelian variety of dimension $\dim A = [F : \mathbb{Q}]$ with good reduction outside S and admitting a map $\mathfrak{o} \hookrightarrow \mathrm{End}_K(A)$. The claim is that there is an $(L, \pi) \in \Pi$ for which $L(s, \rho_{A,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L)}) = L(s, \pi)$. That this holds follows from combining Corollary $7.1.11$ of the ten-author paper [3] (because we follow their proof in the algorithm in Section 10.3.3), Lemma 9.3.3 of Chapter 9, and, to bound the level of a weight zero cuspidal automorphic representation corresponding to an abelian variety A/L with good reduction outside S by $\mathfrak{m}_L$, Brumer-Kramer [34] and the local-global compatibility proved in the ten-author paper [3] and in Varma's Princeton PhD thesis [102]. □

10.4.1.2 Proof of termination.

Proof of termination. In fact this is evident. Certainly Steps $1, 3$, and 5 in the ExplicitPotentialModularity algorithm terminate. Step 4 is a search for weight zero cuspidal automorphic representations of GL_2/L of bounded conductor, so it terminates (implicitly we specify them by the first few Dirichlet coefficients of their L-functions). So we must prove that Step 2, i.e. the RelevantExtensionsOverCM-Fields algorithm, terminates.

But this is again evident (in the same way as in Chapter 9) — we need only comment on Step 4, i.e. the TaylorMoretBaillyExtensionsOverCMFields algorithm, and again the point is that the Moret-Bailly theorem guarantees that we are searching for an element of a nonempty set that is recursively enumerable, so that the algorithm terminates. □

10.5 Remarks.

We have already remarked that, in order to conclude that $(K, C/K) \mapsto C(K)$ is computable (where $K/\mathbb{Q}$ is CM and C/K is Belyi-hyperbolic over K), one may assume something weaker than the existence of abelian varieties associated[13] to weight zero cuspidal automorphic representations of GL_2 over CM fields — specifically, one may assume the existence of motives attached to suitable functorial lifts of π (e.g. $\mathrm{Sym}^2\pi$ on GL_3/L or $\mathrm{Ind}_L^{L^+}\pi$ on GSp_4/L^+, etc.) as well.

In fact one may assume even less.[14] To make the idea clear, let us consider the familiar case of hyperelliptic curves $C_k : y^2 = x^n - k$ over $\mathbb{Q}$ with $n \geq 5$ odd — note that C_k is Belyi-hyperbolic over $\mathbb{Q}$ via the map $(x, y) \mapsto \frac{x^n}{k}$. Let Π be the output of the algorithm implicit in Corollary 10.1.3 when run on input $(\mathbb{Q}, C_k/\mathbb{Q})$. Rather than using Π to bound the heights of the points in $C_k(\mathbb{Q})$, which we could do if we knew there was an A_π/L associated to any π with $(L, \pi) \in \Pi$, we instead seek to compute a finite set S of primes of $\mathbb{Q}$ for which

$$C_k(\mathbb{Q}) \hookrightarrow \mathcal{C}_k^{\mathrm{aff.}}(\mathbb{Z}[S^{-1}])$$

[13] Recall the footnotes in Section 10.1.1.

[14] To be clear, we are not sure how to prove a hypothesis like the below without e.g. the existence of a global object, like a motive associated to π, controlling all the abelian varieties $A_\mathfrak{q}/L_\mathfrak{q}$ associated to the local representations of π.

— in other words, to bound the primes in the denominators of elements of $C_k(\mathbb{Q})$. Of course this amounts to bounding the primes p at which there is a $P \in C_k(\mathbb{Q})$ with $P \equiv \infty \pmod{p}$, where $\infty \in C_k(\mathbb{Q})$ is the unique point at infinity of C_k. (Recall that of course $\mathcal{C}_k^{\text{aff.}}(\mathfrak{o}_{K,S})$ is easily computable by Baker.)

We first simply compute the image of $\infty \in C_k(\mathbb{Q})$ under the Cohen-Wolfart map to find a $\varphi(N)$-dimensional abelian variety $A_0/\mathbb{Q}$ with $\mathbb{Z}[\zeta_N] \hookrightarrow \operatorname{End}_{\mathbb{Q}(\zeta_N)}(A_0)$. Let now $(L, \pi) \in \Pi$ be such that $L(s, T_\lambda(A_0)|_{\operatorname{Gal}(\overline{\mathbb{Q}}/L)}) \neq L(s, \pi)$. By modifying the algorithms we have given to also take a parameter $M \in \mathbb{Z}^+$ and to attempt to falsify the following statement for time M (much like we did in the SeemspDivisible subroutine in Section 7.3.5 of Chapter 7), we may assume without loss of generality (by a day/night argument) that the representations $\rho_{\pi,\mathfrak{p}} : \operatorname{Gal}(\overline{\mathbb{Q}}/L) \to \operatorname{GL}_2(\mathbb{Z}[\zeta_N]_\mathfrak{p})$ have the property that, for all $\mathfrak{q} | \operatorname{Nm} \mathfrak{p}$, $\rho_{\pi,\mathfrak{p}}|_{\operatorname{Gal}(\overline{\mathbb{Q}}/L_\mathfrak{q})}$ is the Galois representation of the Tate module $T_\mathfrak{p}(A_\mathfrak{q})$ of a $\varphi(N)$-dimensional abelian variety $A_\mathfrak{q}/L_\mathfrak{q}$ admitting a map $\mathbb{Z}[\zeta_N] \hookrightarrow \operatorname{End}_{L_\mathfrak{q}(\zeta_N)}(A_\mathfrak{q})$.

Thus we see that, were there a finite-time algorithm that, on input $(A_0/K, L, \pi)$, outputs an integer $T_{(A_0/K,L,\pi)} \in \mathbb{Z}^+$ such that, for all $\mathfrak{q}$ with $\operatorname{Nm} \mathfrak{q} \geq T_{(A_0/K,L,\pi)}$, the abelian $\mathfrak{o}_L/\mathfrak{q}$-varieties $A_0 \pmod{\mathfrak{q}}$ and $A_\mathfrak{q} \pmod{\mathfrak{q}}$ are not isomorphic, we could conclude. This is the aforementioned hypothesis — note that it is obvious if all the $A_\mathfrak{q}/L_\mathfrak{q}$ are base changes of a single A/L, since then $T_{(A_0/K,L,\pi)}$ is controlled by $h(A)$ (two distinct integers $m \neq n$ (or distinct S-integral K-points on $\mathcal{A}_g$) can only be congruent at primes $p \leq |m-n|$). We note also that one can rephrase this in terms of p-divisible groups if one prefers ($\bigoplus_{\mathfrak{p}|(p)} \rho_{\pi,\mathfrak{p}}|_{\operatorname{Gal}(\overline{\mathbb{Q}}/L_\mathfrak{q})}$, regarded as valued in $\operatorname{GL}_{2\varphi(N)}(\mathbb{Z}_p)$, corresponds to a p-divisible group over $L_\mathfrak{q}$, and thus, by Raynaud, over $\mathfrak{o}_{L,\mathfrak{q}}$. Applying the $D_{\text{crys.}}$ functor we find a filtered φ-module corresponding to the p-divisible group over $\mathfrak{o}_{L,\mathfrak{q}}$, and forgetting the filtration gives the p-divisible group associated to the special fibre).

Nonetheless a fully unconditional result for all Belyi-hyperbolic curves over CM fields seems just out of reach.

Chapter 11

$x^6 + 4y^3 = 1$: an example.

Abstract.

We give a finite-time algorithm that, on input $K/\mathbb{Q}$ a totally real field, outputs either the finite set $\{(x, y) : x, y \in K, x^6 + 4y^3 = 1\}$, or else — and this case can only occur when $[K : \mathbb{Q}]$ is even — an unconditional counterexample to the absolute Hodge conjecture.

11.1 Introduction.

The purpose of this chapter is to give a simple, self-contained, and explicit example application of the ideas detailed in the previous chapters.

11.1.1 Main theorem.

We will prove the following. Let, for $a \in \overline{\mathbb{Q}}$, $C_a/\mathbb{Q}(a)$ be the curve with affine model $x^6 + 4y^3 = a^2$.

Theorem 11.1.1. *There is a finite-time algorithm that, on input (K, a) with $K/\mathbb{Q}$ a totally real field and $a \in K^\times$, outputs either $C_a(K)$, or else — and this case can only occur when $[K : \mathbb{Q}]$ is even — an unconditional disproof of the absolute Hodge conjecture.*

The caveat about the absolute Hodge conjecture should be familiar from Chapter 9 and arises simply because it is not yet known unconditionally that there is an abelian variety associated to *every* parallel weight 2 Hilbert modular eigencuspform over an even-degree totally real field.

Note that in the particular case of $K/\mathbb{Q}$ totally real of odd degree and $a \in K^\times$ we are asserting that there is a (to be clear: completely unconditional) finite-time algorithm computing $C_a(K)$. In fact because in principle one could phrase the main theorems of Chapter 9 as explicit height bounds, this asserts the existence of (completely unconditional) explicit height bounds on the K-rational points on the curves C_a/K for $K/\mathbb{Q}$ totally real of odd degree.

11.1.2 The technique.

To prove Theorem 11.1.1, we will construct a non-isotrivial family $A \to C_a$ of abelian surfaces over the smooth projective genus 4 curve C_a/K with affine model $x^6 + 4y^3 = a^2$ which has the property that each fibre over a point in $C_a(K)$ is an abelian surface over K with bounded conductor (since the family is a family of abelian surfaces over the whole of the projective curve C_a), has quaternionic multiplication over $K(\zeta_3)$, and, for $K/\mathbb{Q}$ totally real, is of GL_2-type over K.

This implies that, to find $C_a(K)$ for such K, we need only produce the GL_2-type abelian surfaces over K with bounded conductor. In fact it is simple to produce an explicit finite set $\mathcal{F}$ for which each of the relevant abelian surfaces is of $\mathrm{GL}_2(F)$-type over K for some $F \in \mathcal{F}$, and then we may simply apply the main theorems of Chapter 9 to conclude.

Finally, we will check that we are not proving something tautological by explicitly constructing infinitely many totally real points on C_1 — these are of course guaranteed to exist by Moret-Bailly's theorem, but we will give an explicit construction nonetheless. We will also produce a totally real cubic field $K/\mathbb{Q}$ over which there are infinitely many $a \in K^\times$ for which $C_a(K) \neq \emptyset$ to show that the theorem is nontrivial over odd-degree totally real fields — where one may ignore the caveat about the absolute Hodge conjecture — as well.

We note that we chose the example $C_1 : x^6 + 4y^3 = 1$ (and its twists C_a) because of Deines-Fuselier-Long-Swisher-Tu's [42], where they discuss exactly the hypergeometric family of abelian varieties associated to the arithmetic triangle group[1] $\Delta(3,6,6)$ which we use here. We hope it is evident just how much the existence of this chapter owes to their [42].

11.2 The hypergeometric family.

For us the relevant hypergeometric family will be the one corresponding to the hypergeometric function $z \mapsto {}_2F_1\left(\begin{array}{cc|c} \frac{1}{6} & \frac{1}{3} & \\ & & z \\ & \frac{5}{6} & \end{array}\right)$. Let, for $\lambda \neq 0, 1, \infty$, $X_\lambda/\mathbb{Q}(\lambda)$ be the desingularization of the curve

$$y^6 = x^4(1-x)^3(1-\lambda \cdot x).$$

Note that each fibre of this family $X \to \mathbb{P}^1 - \{0, 1, \infty\}$ is a genus 3 curve. Let $\operatorname{Jac} X_\lambda/\mathbb{Q}(\lambda)$ be the family of Jacobians, and let $A_\lambda/\mathbb{Q}(\lambda)$ be the family of Prym varieties corresponding to the map to the (desingularization of the) curve E_λ : $y^3 = x^4(1-x)^3(1-\lambda \cdot x)$. In other words, there is an isogeny

$$\operatorname{Jac} X_\lambda \sim_{\mathbb{Q}(\lambda)} E_\lambda \times A_\lambda,$$

[1] For comparison to Chapter 8, note that $\Delta(3,6,6)$ is of index 2 inside $\Delta(2,6,6)$ — indeed, for $x, y \in \Delta(2,6,6)$ generators with $x^2 = y^6 = (xy)^6 = \mathrm{id}$, note that $(xy)^2$ and y^{-1} generate a subgroup isomorphic to $\Delta(3,6,6)$.

defined over $\mathbb{Q}(\lambda)$, with $A_\lambda/\mathbb{Q}(\lambda)$ an abelian surface, and $E_\lambda : y^2 = x^3 + 16\lambda^2$ an elliptic curve over $\mathbb{Q}(\lambda)$.

Thanks to the automorphism $(x, y) \mapsto (x, \zeta_6 \cdot y)$ of $y^6 = x^4(1-x)^3(1-\lambda\cdot x)$, it follows that

$$\mathbb{Z}[\zeta_3] \hookrightarrow \mathrm{End}_{\mathbb{Q}(\zeta_3,\lambda)}(A_\lambda).$$

Thus each fibre is of $\mathrm{GL}_2(\mathbb{Z}[\zeta_3])$-type over $\mathbb{Q}(\zeta_3, \lambda)$, and indeed this was all we used in Chapter 10.

However because the relevant triangle group, namely $\Delta(3, 6, 6)$, is arithmetic, we get even more. Specifically, the base change $A_\lambda/\overline{\mathbb{Q}(\lambda)}$ admits quaternionic multiplication by the indefinite quaternion algebra over $\mathbb{Q}$ of discriminant 6.

In fact we will see that the quaternionic multiplication is defined over $\mathbb{Q}\left(\zeta_3, (\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2)^{\frac{1}{6}}\right)$, but let us leave this to Section 11.3.

Let us now explain how to form a family of abelian surfaces over our $C_a/\mathbb{Q}(a^2)$ using this construction.

Specifically, writing, for $P =: (x, y) \in C_a$,

$$f_a(P) := \frac{x^6}{a^2},$$

we claim that the pullback family

$$P \mapsto A_{f_a(P)},$$

which is a priori defined over $C_a - \{P \in C_a : f_a(P) \in \{0, 1, \infty\}\}$, extends to a family over all of C_a. In Chapter 10 we cited Cohen-Wolfart to see this in the general case, but let us check it by hand here.[2]

[2] Our argument will basically amount to an explicit continuation of particular Schwarz triangle functions to their singular points and so will ultimately be no different from the general argument given by Cohen-Wolfart in [40].

It suffices to show that the pullback family

$$P \mapsto \operatorname{Jac} X_{f_a(P)}$$

extends to all of C_a. This is a map from a punctured curve to the moduli space A_3. By the Borel extension theorem (here basically Riemann extension) it extends to a map to the Baily-Borel/Satake compactification $C_a \to A_3^{\mathrm{B.B.}}$. Let us now check at the level of periods that the family does not degenerate[3] at those P for which $f_a(P) \in \{0, 1, \infty\}$. It is evident that the pullback family $P \mapsto E_{f_a(P)}$, which is isotrivial (all fibres are isomorphic over $\mathbb{Q}(f_a(P), a^{\frac{1}{3}})$ to the elliptic curve $y^2 = x^3 + 1$), does not degenerate. So it suffices to check this for the family $P \mapsto A_{f_a(P)}$.

Following Deines-Fuselier-Long-Swisher-Tu's [42], we see that the relevant periods are those of the differentials

$$\omega_+ := \frac{dx}{\sqrt[6]{x^4(1-x)^3(1-\lambda \cdot x)}}$$

and

$$\omega_- := \frac{dx}{\sqrt[6]{x^2(1-x)^3(1-\lambda \cdot x)^5}}$$

over Pochhammer contours $\gamma_{0,1}, \gamma_{\frac{1}{\lambda},\infty}$ around the pairs $\{0, 1\}$ and $\{\frac{1}{\lambda}, \infty\}$ of singular points of the model $y^6 = x^4(1-x)^3(1-\lambda \cdot x)$.

These are of course all values of hypergeometric functions, and indeed Example 3 in Deines-Fuselier-Long-Swisher-Tu's [42] evaluates the periods explicitly: we find that, for $\lambda \neq 0, 1, \infty$, the period lattice of A_λ is given by

$$\mathbb{Z}[\zeta_3] \cdot v + \mathbb{Z}[\zeta_3] \cdot w \subseteq \mathbb{C}^2,$$

[3] We leave implicit the fact that the family extends *holomorphically* over those points to a family over C — this will follow from Riemann extension and the below analysis upon noting that $x^6 = \lambda := f_a(P)$ and $y^3 = \frac{1-\lambda}{4}$, so that the 2×4 complex matrices that we produce over points in small punctured disks around the various punctures will remain bounded and vary holomorphically in (x, y).

with $\zeta_3 \curvearrowright \mathbb{C}^2$ via $\mathrm{diag}(\zeta_3, \zeta_3^{-1}) \in \mathrm{GL}_2(\mathbb{C})$, and with

$$v := \begin{pmatrix} B\left(\frac{1}{3}, \frac{1}{2}\right) \cdot {}_2F_1 \left(\begin{matrix} \frac{1}{6} & \frac{1}{3} \\ & \frac{5}{6} \end{matrix} \middle| \lambda \right) \\ B\left(\frac{2}{3}, \frac{1}{2}\right) \cdot {}_2F_1 \left(\begin{matrix} \frac{5}{6} & \frac{2}{3} \\ & \frac{7}{6} \end{matrix} \middle| \lambda \right) \end{pmatrix},$$

$$w := \begin{pmatrix} 0 & \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}} \\ 2 \cdot \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{-\frac{1}{6}} & 0 \end{pmatrix} \cdot v,$$

where we have written $B(\cdot, \cdot)$ for the usual beta function

$$B(a, b) := \int_0^1 x^a (1-x)^b \frac{dx}{x(1-x)} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

and ${}_2F_1 \left(\begin{matrix} a & b \\ & c \end{matrix} \middle| z \right)$ for the usual Gauss hypergeometric function

$$B(b, c-b) \cdot {}_2F_1 \left(\begin{matrix} a & b \\ & c \end{matrix} \middle| z \right) = \int_0^1 x^b (1-x)^{c-b} (1 - z \cdot x)^{-a} \frac{dx}{x(1-x)}.$$

Let us check that the family of abelian varieties does not degenerate as $\lambda \to 0$. Of course we may apply an automorphism of $\mathbb{C}^2$ to the period lattice, and so we apply the automorphism $\mathrm{diag}\left(1, \frac{1}{2} \cdot \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right) \in \mathrm{GL}_2(\mathbb{C})$ and then take the limit as $\lambda \to 0$. We obtain the lattice[4]

$$\mathbb{Z}[\zeta_3] \cdot \begin{pmatrix} B\left(\frac{1}{3}, \frac{1}{2}\right) \cdot {}_2F_1 \left(\begin{matrix} \frac{1}{6} & \frac{1}{3} \\ & \frac{5}{6} \end{matrix} \middle| 0 \right) \\ 0 \end{pmatrix} + \mathbb{Z}[\zeta_3] \cdot \begin{pmatrix} 0 \\ B\left(\frac{1}{3}, \frac{1}{2}\right) \cdot {}_2F_1 \left(\begin{matrix} \frac{1}{6} & \frac{1}{3} \\ & \frac{5}{6} \end{matrix} \middle| 0 \right) \end{pmatrix},$$

[4] Of course ${}_2F_1 \left(\begin{matrix} \frac{1}{6} & \frac{1}{3} \\ & \frac{5}{6} \end{matrix} \middle| 0 \right) = 1$, etc., but we have left it as such for clarity.

and so we find, as expected, that the limiting abelian variety over $\mathbb{C}$ is CM — and indeed simply $E_1^{\times 2}/\mathbb{C}$, the square of the elliptic curve $E_1 : y^2 = x^3 + 1$ over $\mathbb{C}$ with complex multiplication by $\mathbb{Z}[\zeta_3]$.

One gets the same as $\lambda \to 1$ after scaling away the singularity of ${}_2F_1\left(\begin{array}{cc|c} \frac{5}{6} & \frac{2}{3} & \\ & & \lambda \\ & \frac{7}{6} & \end{array}\right)$ at $\lambda = 1$ by e.g. changing variables by $\mathrm{diag}\left(1, \left(\frac{1-\lambda}{4}\right)^{\frac{1}{3}}\right) \in \mathrm{GL}_2(\mathbb{C})$, so that the family also does not degenerate (and indeed has the same limit, namely $E_1^{\times 2}/\mathbb{C}$) at $\lambda = 1$.

Finally one instead cancels the zeroes of ${}_2F_1\left(\begin{array}{cc|c} \frac{1}{6} & \frac{1}{3} & \\ & & \lambda \\ & \frac{5}{6} & \end{array}\right)$ and ${}_2F_1\left(\begin{array}{cc|c} \frac{5}{6} & \frac{2}{3} & \\ & & \lambda \\ & \frac{7}{6} & \end{array}\right)$ as $\lambda \to \infty$ (via $\mathrm{diag}(\lambda^{\frac{1}{6}}, \lambda^{\frac{2}{3}}) \in \mathrm{GL}_2(\mathbb{C})$) to see that the family does not degenerate (and has the same limit of $E_1^{\times 2}/\mathbb{C}$) at $\lambda = \infty$ as well.

So the map $C_a \to A_3^{\mathrm{B.B.}}$ via $P \mapsto \mathrm{Jac}\, X_{f_a(P)}$ indeed lands inside A_3, whence we have the desired family $A \to C_a$.

11.3 Endomorphisms of the hypergeometric family.

At the level of periods it is obvious that the abelian surfaces $A_\lambda/\mathbb{Q}(\lambda)$ have, at least over $\overline{\mathbb{Q}(\lambda)}$, quaternionic multiplication by the indefinite quaternion algebra over $\mathbb{Q}$ of discriminant 6: one evidently has endomorphisms by $\mathbb{Z}[\zeta_3]$, and one also has the endomorphism

$$\begin{pmatrix} 0 & \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}} \\ 2 \cdot \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{-\frac{1}{6}} & 0 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}),$$

which squares to multiplication by 2. However to run our argument we must be much more precise about the field of definition of the quaternionic multiplication of the A_λ. Thankfully this too is made simple by the analysis done in Deines-Fuselier-Long-Swisher-Tu's [42].

First and foremost, as we have already noted in both Chapter 10 and above, it is evident that the endomorphism ζ_3 is defined over $\mathbb{Q}(\zeta_3, \lambda)$, and so $\mathbb{Z}[\zeta_3] \hookrightarrow \mathrm{End}_{\mathbb{Q}(\zeta_3,\lambda)}(A_\lambda)$.

We claim that indeed the base change $A_\lambda/\mathbb{Q}\left(\zeta_3, \lambda, \left[\lambda \cdot \left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)$ has quaternionic multiplication by an order in the indefinite quaternionic algebra over $\mathbb{Q}$ of discriminant 6. To see this we will argue indirectly by using Faltings' proof of the Tate conjecture for abelian varieties.[5]

For $\lambda \in \overline{\mathbb{Q}}$, because $A_\lambda/\mathbb{Q}(\lambda)$ is of $\mathrm{GL}_2(\mathbb{Z}[\zeta_3])$-type over $\mathbb{Q}(\zeta_3, \lambda)$, to each prime $\mathfrak{p} \subseteq \mathbb{Z}[\zeta_3]$ there is a corresponding 2-dimensional Galois representation

$$\rho_{\mathfrak{p},\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_3, \lambda)) \to \mathrm{GL}_2(\mathbb{Z}[\zeta_3]_{\mathfrak{p}}),$$

and indeed the 4-dimensional Galois representation $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_3, \lambda)) \to \mathrm{GL}_2(\mathbb{Z}_p)$ (here $\mathfrak{p}|(p)$) corresponding to the Tate module $T_p(A_\lambda)$ decomposes as $\rho_{\mathfrak{p},\lambda} \oplus \overline{\rho}_{\mathfrak{p},\lambda}$, where the $\bar{\cdot}$ denotes the action of complex conjugation on $\mathbb{Z}[\zeta_3]$ and thus $\mathbb{Z}[\zeta_3]_{\mathfrak{p}}$.

Again following Deines-Fuselier-Long-Swisher-Tu's [42], we see (from counting points on the reduction of X_λ and its twists modulo $\mathfrak{q}$ in the usual way) that, for all $n \in \mathbb{Z}^+$,

$$\begin{aligned}&\{\mathrm{tr}(\rho_{\mathfrak{p},\lambda}(\mathrm{Frob}_{\mathfrak{q}}^n)), \mathrm{tr}(\overline{\rho}_{\mathfrak{p},\lambda}(\mathrm{Frob}_{\mathfrak{q}}^n))\} \\ &= \left\{ -\overline{\eta}(-1) \cdot (\mathrm{Nm}\,\mathfrak{q})^n \cdot {}_2F_1\left(\begin{matrix}\eta & \eta^2 \\ & \overline{\eta}\end{matrix} \;\middle|\; \lambda\right)_{(\mathrm{Nm}\,\mathfrak{q})^n}, -\eta(-1) \cdot (\mathrm{Nm}\,\mathfrak{q})^n \cdot {}_2F_1\left(\begin{matrix}\overline{\eta} & \overline{\eta}^2 \\ & \eta\end{matrix} \;\middle|\; \lambda\right)_{(\mathrm{Nm}\,\mathfrak{q})^n} \right\}.\end{aligned}$$

[5] We will of course only take $\lambda \in \overline{\mathbb{Q}}$ so that we may apply Faltings' proof of the Tate conjecture in this context. Presumably one can simply write down the extra endomorphism explicitly as a correspondence on X_λ (using e.g. the explicit transformation $z \mapsto \frac{1-z}{1-\lambda \cdot z}$ that proves the Euler transformation formula for hypergeometric functions which underlies the endomorphism, at least at the level of periods), but we do not yet see how.

Here $\eta : \mathbb{F}_{(\mathrm{Nm}\,\mathfrak{q})^n} \twoheadrightarrow \mu_6 \subseteq \mathfrak{o}_{\mathbb{Q}(\zeta_3,\lambda)}^\times$ is the sextic character lifting $a \mapsto a^{\frac{(\mathrm{Nm}\,\mathfrak{q})^n-1}{6}} \in \mu_6 \pmod{\mathfrak{q}}$, and, for $\alpha, \beta, \gamma : \mathbb{F}_q^\times \to \mathbb{C}^\times$, writing triv for the trivial character,

$$ {}_2F_1\left(\begin{array}{cc|c} \alpha & \beta & \\ & & z \\ & \gamma & \end{array}\right)_q := \mathrm{triv}(z) \cdot \frac{(\beta\cdot\gamma)(-1)}{q} \sum_{x\in\mathbb{F}_q} \beta(x)\cdot(\beta^{-1}\cdot\gamma)(1-x)\cdot(\alpha^{-1})(1-z\cdot x) $$

is the usual Gauss hypergeometric function over $\mathbb{F}_q$.

By the $\mathbb{F}_q$-analogue of the Euler hypergeometric identity — proven in the same way via the change of variables $x \mapsto \frac{1-x}{1-z\cdot x}$ — and an evaluation of a ratio of Jacobi sums[6], it follows that:

$$ {}_2F_1\left(\begin{array}{cc|c} \eta & \eta^2 & \\ & & \lambda \\ & \overline{\eta} & \end{array}\right)_{(\mathrm{Nm}\,\mathfrak{q})^n} = \eta\left(\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right)\cdot {}_2F_1\left(\begin{array}{cc|c} \overline{\eta} & \overline{\eta}^2 & \\ & & \lambda \\ & \eta & \end{array}\right)_{(\mathrm{Nm}\,\mathfrak{q})^n}. $$

Moreover, we of course have that $\eta(-1) = \overline{\eta}(-1)$.

Therefore, upon restricting to $\mathrm{Gal}\left(\overline{\mathbb{Q}}/\mathbb{Q}\left(\zeta_3, \lambda, \left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)\right)$, we find that:

$$ \left.(\rho_{\mathfrak{p},\lambda}\otimes_{\mathbb{Z}}\mathbb{Q})\right|_{\mathrm{Gal}\left(\overline{\mathbb{Q}}/\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)\right)} \cong \left.(\overline{\rho}_{\mathfrak{p},\lambda}\otimes_{\mathbb{Z}}\mathbb{Q})\right|_{\mathrm{Gal}\left(\overline{\mathbb{Q}}/\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)\right)}, $$

[6] See e.g. Proposition 9 of Deines-Fuselier-Long-Swisher-Tu's [42]. At the infinite place the evaluation of the resulting ratio of Jacobi sums amounts to the statement that

$$ \frac{\Gamma\left(\frac{1}{3}\right)\cdot\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{6}\right)\cdot\Gamma\left(\frac{2}{3}\right)} = 4^{-\frac{1}{3}}, $$

which follows from the usual duplication formula $\Gamma(z)\cdot\Gamma\left(z+\frac{1}{2}\right) = 2^{1-2z}\cdot\Gamma\left(\frac{1}{2}\right)\cdot\Gamma(2z)$ at $z = \frac{1}{6}$. The corresponding statement at a finite prime is proven in the same way — the duplication formula is a particular case of the Γ-function multiplication theorem, whose finite-field analogue is the Hasse-Davenport relation. Note that this is the provenance of the constant 4 in the models for our curves $C_a/\mathbb{Q}$.

since their Frobenius traces match. In other words

$$\mathrm{End}^0_{\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)}(A_\lambda)\otimes_{\mathbb{Q}}\mathbb{Q}_p \simeq \mathrm{End}_{\mathrm{Gal}\left(\overline{\mathbb{Q}}/\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)\right)}\left((\rho_{\mathfrak{p},\lambda}\oplus\overline{\rho}_{\mathfrak{p},\lambda})\otimes_{\mathbb{Z}}\mathbb{Q}\right)$$

is strictly larger than $\mathbb{Q}[\zeta_3]\otimes_{\mathbb{Q}}\mathbb{Q}_p$. Because we know from our discussion of periods that $\mathrm{End}^0_{\mathbb{C}}(A_\lambda)$ contains the indefinite quaternion algebra over $\mathbb{Q}$ of discriminant 6, it follows from the Albert classification that either the base change $A_\lambda/\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)$ is not simple and thus the square of an elliptic curve with potential CM, or else A_λ acquires quaternionic multiplication by an order in the indefinite quaternion algebra over $\mathbb{Q}$ of discriminant 6 over $\mathbb{Q}\left(\zeta_3,\lambda,\left[\lambda\cdot\left(\frac{1-\lambda}{4}\right)^2\right]^{\frac{1}{6}}\right)$, as desired.

Naturally if L/K is a quadratic extension and A/K is an abelian variety with A/L L-simple, then

$$\dim_{\mathbb{Q}}\mathrm{End}^0_K(A)\geq\frac{1}{2}\cdot\dim_{\mathbb{Q}}\mathrm{End}^0_L(A).$$

Indeed, writing $\mathrm{id}\neq\sigma\in\mathrm{Gal}(L/K)$ for the nontrivial automorphism, one has the $\mathbb{Q}$-linear diagonalization $\mathrm{End}^0_L(A)=\mathrm{End}^0_L(A)^+\oplus\mathrm{End}^0_L(A)^-$ in the usual way, and indeed $\mathrm{End}^0_L(A)^+=\mathrm{End}^0_K(A)$. Thus either $\mathrm{End}^0_K(A)=\mathrm{End}^0_L(A)$ or else, writing $0\neq\psi\in\mathrm{End}^0_L(A)^-$ for an element with $\sigma\cdot\psi=-\psi$, one finds a surjective $\mathbb{Q}$-linear map

$$\mathrm{End}^0_K(A)\twoheadrightarrow\mathrm{End}^0_L(A)^-$$

via $\varphi\mapsto\varphi\circ\psi$. Hence e.g. if $\dim_{\mathbb{Q}}\mathrm{End}^0_L(A)\geq 2\cdot\dim A$ it follows that A/K is of GL_2-type over K.

It follows for example that, for $L/\mathbb{Q}$ a CM field with maximal totally real subfield L^+ and $P=:(x,y)\in C_a(L)$ with $f_a(P)=\frac{x^6}{a^2}\in L^+$, the abelian surface $A_{f_a(P)}/L^+$ is of GL_2-type over $L(\zeta_3)^+$, the totally real subfield of the compositum

$L(\zeta_3)$. This follows because the defining equation $x^6 + 4y^3 = a^2$ implies that

$$f_a(P) \cdot \left(\frac{1 - f_a(P)}{4}\right)^2 = \frac{(x \cdot y)^6}{a^6}$$

is a sixth power in L, so that

$$\mathbb{Q}\left(\zeta_3, f_a(P), \left[f_a(P) \cdot \left(\frac{1 - f_a(P)}{4}\right)^2\right]^{\frac{1}{6}}\right) = L(\zeta_3).$$

Applying this when $L/\mathbb{Q}$ is totally real gives[7] that each $P \in C_a(L)$ gives rise to an $A_{f_a(P)}/L$ which is of GL_2-type over $L = L(\zeta_3)^+$.

11.4 Proof of Theorem 11.1.1.

Having produced the family $A \to C_a$ and controlled the field of definition of its quaternionic multiplication, we are now ready to prove Theorem 11.1.1.

Proof of Theorem 11.1.1. Let $K/\mathbb{Q}$ be a totally real field. Let $a \in K^\times$. Let $\mathcal{C}_a/\mathfrak{o}_K$ be the minimal proper regular model of the curve C_a/K. There is an explicit finite set S_a of places of K containing all infinite places of K such that the map $C_a \to A_3$ via $P \mapsto \operatorname{Jac} X_{f_a(P)}$ extends to a finite-to-one map of S_a-integral models $\mathcal{C}_a \to \mathcal{A}_3$. Enlarging S_a to another explicit finite set of places of K if necessary, it follows that it suffices to find in finite time the image of $C_a(K) \hookrightarrow \mathcal{C}_a(\mathfrak{o}_{K,S_a}) \to \mathcal{A}_3(\mathfrak{o}_{K,S_a})$.

In other words we find that, for each $P \in C_a(K)$, the abelian surface $A_{f_a(P)}/K$ is of GL_2-type over K and has good reduction outside S_a. By Brumer-Kramer [34]

[7] One could also see this via the equivalent argument that, thanks to the identity of Frobenius traces we used above, $\rho_{\mathfrak{p},\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/L(\zeta_3))}$ is fixed up to isomorphism via the outer automorphism of $\mathrm{Gal}(\overline{\mathbb{Q}}/L(\zeta_3))$ induced by complex conjugation (which acts by conjugating ζ_3, thanks to our explicit formula for the automorphism at the level of the curve X_λ), whence it is the restriction of a 2-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/L(\zeta_3)^+)$. Said another way, the automorphic forms produced by the argument of Chapter 10 are all base changes of Hilbert modular forms over the relevant totally real subfields.

it follows that the conductor of $A_{f_a(P)}/K$ divides the explicit integer

$$N_{K,a} := \prod_{\mathfrak{p}\in S_a} (\mathrm{Nm}\,\mathfrak{p})^{10^{10^{10}}\cdot[K:\mathbb{Q}]}.$$

Let us immediately dispatch the reducible or potentially CM cases. First, using Baker it is easy to find all products of elliptic curves, and thus products of pairs thereof, over K with good reduction outside S_a. So we may (and will) assume that all our abelian surfaces $A_{f_a(P)}/K$ are K-simple.

Similarly, by an observation of Silverberg one has that, if $A_{f_a(P)}/\overline{\mathbb{Q}}$ has sufficiently many complex multiplications, then so does $A_{f_a(P)}/K(A_{f_a(P)}[2\cdot 3\cdot 5\cdot 7])$. However the extension $K(A_{f_a(P)}[2\cdot 3\cdot 5\cdot 7])/K$ is of explicitly bounded degree and ramification, and so lies in an explicit finite set. Thus if $A_{f_a(P)}/K$ is potentially CM, it must be CM over an explicit finite extension $\tilde{K}/K$. However the $\tilde{K}$-simple abelian varieties over $\tilde{K}$ with CM defined over $\tilde{K}$ are explicitly computable, since they correspond to one of an explicit finite set of CM types (namely the reflex types of the CM types of CM subfields of $\tilde{K}$). So we may (and will) assume that all our abelian surfaces $A_{f_a(P)}/K$ are not potentially CM, and thus e.g. $\mathrm{End}^0_{K(\zeta_3)}(A_{f_a(P)}) = \mathrm{End}^0_{\overline{\mathbb{Q}}}(A_{f_a(P)})$.

Let us now produce an explicit finite set $\mathcal{F}$ of quadratic fields $F/\mathbb{Q}$ for which, for each $P\in C_a(K)$, there is an $F_P\in\mathcal{F}$ such that $A_{f_a(P)}/K$ is of $\mathrm{GL}_2(F_P)$-type over K.

Let $q\in\mathbb{Z}^+$ be a prime that is explicitly sufficiently large with respect to both $N_{K,a}$ and $K/\mathbb{Q}$. Let Σ_q be the finite set of primes of K lying over $q\in\mathbb{Z}^+$. Let

$$T := \text{output of FaltingsPrimeList}(2, K, S_a\cup\Sigma_q, q^2).$$

Let $\tilde{T}$ be the finite set of primes of $K(\zeta_3)$ lying over a prime in T.

Let

$$\Lambda := \left\{ (a_{\mathfrak{P}})_{\mathfrak{P} \in \tilde{T}} \in \mathbb{Z}^{\tilde{T}} : \forall \mathfrak{P} \in \tilde{T}, |a_{\mathfrak{P}}| \leq 2 \cdot \sqrt{\operatorname{Nm} \mathfrak{P}} \right\}.$$

Let Ψ be the explicit finite set of characters $\psi : \operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \mathbb{F}_{q^2}^{\times}$, regarded as finite-order characters $\operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \mu_{q^2-1} \hookrightarrow \overline{\mathbb{Q}}_q^{\times}$ via the canonical lifting character, that are unramified outside $S_a \cup \Sigma_q$.

For each $\vec{a} \in \Lambda$, $\psi \in \Psi$, and $\mathfrak{P} \in \tilde{T}$, write $\mathfrak{p} \subseteq \mathfrak{o}_K$ for the prime of K below $\mathfrak{P} \subseteq \mathfrak{o}_{K(\zeta_3)}$ (thus $\mathfrak{p} \in T$). If $\mathfrak{p}$ is split in $K(\zeta_3)$, write $b_{\mathfrak{p}}^{(\vec{a},\psi)} := a_{\mathfrak{P}}$. Otherwise (thus $\mathfrak{p}$ is inert in $K(\zeta_3)$ by construction) let $b_{\mathfrak{p}}^{(\vec{a},\psi)} \in \overline{\mathbb{Q}}$ be a solution of the equation

$$(b_{\mathfrak{p}}^{(\vec{a},\psi)})^2 - 2 \cdot \psi(\mathfrak{p}) \cdot \operatorname{Nm} \mathfrak{p} = a_{\mathfrak{P}}.$$

Write, for each $\vec{a} \in \Lambda$ and $\psi \in \Psi$,

$$F_{(\vec{a},\psi)} := \mathbb{Q}(\{b_{\mathfrak{p}}^{(\vec{a},\psi)}\}_{\mathfrak{P} \in \tilde{T}}).$$

Let

$$\mathcal{F} := \left\{ F_{(\vec{a},\psi)} \,|\, \vec{a} \in \Lambda, \psi \in \Psi : [F_{(\vec{a},\psi)} : \mathbb{Q}] \leq 2 \right\}.$$

We claim that, for all $P \in C_a(K)$, there is an $F \in \mathcal{F}$ such that the abelian surface $A_{f_a(P)}/K$ is of $\operatorname{GL}_2(F)$-type over K.

Write $E/\mathbb{Q}$ for the quadratic extension for which $A_{f_a(P)}/K$ is of $\operatorname{GL}_2(E)$-type over K, so that our claim is that $E \in \mathcal{F}$. Write $\mathfrak{q} \subseteq \mathfrak{o}_E$ with $\mathfrak{q}|(q)$ for a prime of E over q. Note that $\operatorname{Nm} \mathfrak{q} \,|\, q^2$. Let $\rho_{P,\mathfrak{q}} : \operatorname{Gal}(\overline{\mathbb{Q}}/K) \to \operatorname{GL}_2(\mathfrak{o}_{E,\mathfrak{q}})$ be the corresponding 2-dimensional Galois representation. By construction $\rho_{P,\mathfrak{q}}$ is unramified outside $S_a \cup \Sigma_q$. Moreover, because the Hodge-Tate weights of $\rho_{P,\mathfrak{q}}$ at all primes over q are $\{0, -1\}$, it follows that $\psi_P := \det \rho_{P,\mathfrak{q}} \cdot \chi_q^{-1}$ is of finite order, where χ_q is the q-adic cyclotomic character. However evidently ψ_P is also unramified outside $S_a \cup \Sigma_q$, so that $\psi_P \in \Psi$.

Recall that $\rho_{P,\mathfrak{q}}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/K(\zeta_3))}$ has $\mathbb{Z}$-integral Frobenius traces (by the identity of finite-field analogues of hypergeometric functions we used above). Thus by purity $\vec{a}_P := (\mathrm{tr}(\rho_{P,\mathfrak{q}}(\mathrm{Frob}_{\mathfrak{P}})))_{\mathfrak{P}\in\tilde{T}} \in \Lambda$.

Finally the polynomial identity $\mathrm{tr}(A^2) = \mathrm{tr}(A)^2 - 2\cdot\det A$ for two-by-two matrices (and the fact that $\mathrm{Frob}_{\mathfrak{P}} = \mathrm{Frob}_{\mathfrak{p}}$ or $\mathrm{Frob}_{\mathfrak{p}}^2$ if $\mathfrak{p}$ is split or inert, respectively) implies that, for all $\mathfrak{p}\in T$,

$$\mathrm{tr}(\rho_{P,\mathfrak{q}}(\mathrm{Frob}_{\mathfrak{p}})) = \pm b_{\mathfrak{p}}^{(\vec{a}_P,\psi_P)}.$$

Therefore $F^{(P)} := F_{(\vec{a}_P,\psi_P)} = \mathbb{Q}(\{\mathrm{tr}(\rho_{P,\mathfrak{q}}(\mathrm{Frob}_{\mathfrak{p}}))\}_{\mathfrak{p}\in T})$. It remains to show that $E = F^{(P)}$. Certainly because $\mathrm{tr}(\rho_{P,\mathfrak{q}}(\mathrm{Frob}_{\mathfrak{p}})) \in E$ for all $\mathfrak{p}\in T$ we have that $F^{(P)} \subseteq E$, so that it suffices to show that $F^{(P)} \neq \mathbb{Q}$. But were $F^{(P)} = \mathbb{Q}$, then the conjugate representation $\overline{\rho}_{P,\mathfrak{q}} \otimes_{\mathbb{Z}} \mathbb{Q} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(E_{\mathfrak{q}})$ would have Frobenius traces matching those of $\rho_{P,\mathfrak{q}} \otimes_{\mathbb{Z}} \mathbb{Q}$ at primes $\mathfrak{p}\in T$. Hence by Faltings' Lemma (Lemma 7.2.1 in Chapter 7) the two would be isomorphic, which would imply that

$$\mathrm{End}^0_K(A_{f_a(P)}) \otimes_{\mathbb{Q}} \mathbb{Q}_q \simeq \mathrm{End}_{\mathrm{Gal}(\overline{\mathbb{Q}}/K)}((\rho_{P,\mathfrak{q}} \oplus \overline{\rho}_{P,\mathfrak{q}}) \otimes_{\mathbb{Z}} \mathbb{Q}) \simeq M_2(E_{\mathfrak{q}}),$$

whence by the Albert classification either $A_{f_a(P)}/K$ is not K-simple, or else it has quaternionic or complex multiplication defined over K. Since we have already dealt with the first and third cases, and because it is impossible for an abelian surface to have quaternionic multiplication defined over a field with a real place (consider $\mathrm{End}^0_K(A_{f_a(P)}) \curvearrowright H^1(A_{f_a(P)}(\mathbb{R}),\mathbb{Q})$), we obtain the desired contradiction.

Thus $E = F^{(P)}$ and so indeed $E \in \mathcal{F}$, as desired.

Finally we apply Theorems 9.1.1 and 9.1.2 of Chapter 9 for each[8] $F \in \mathcal{F}$ to obtain bounds on $h(A_{f_a(P)})$, whence we conclude. □

[8] Note that if A/K has $\mathfrak{o} \hookrightarrow \mathrm{End}_K(A)$ and $\mathfrak{o} \subseteq \mathfrak{o}'$ is of finite index, then the Serre tensor product $A' := A \otimes_{\mathfrak{o}} \mathfrak{o}'$ has $\mathfrak{o}' \hookrightarrow \mathrm{End}_K(A')$ and also $A \sim_K A'$. Thus via Masser-Wüstholz it suffices to apply Theorems 9.1.1 and 9.1.2 of Chapter 9 to maximal orders only.

To keep the example self-contained, let us explain the key points that we are implicitly using by citing Chapter 9, at least in the case of $[K:\mathbb{Q}]$ odd. Let now $F \in \mathcal{F}$, and let us restrict ourselves to searching for $\mathrm{GL}_2(F)$-type abelian surfaces over K which are K-simple and not potentially CM.

Suppose for the moment that we have produced an explicit finite set $\mathcal{L}$ of odd-degree extensions of K for which, for all K-simple, non-potentially-CM, $\mathrm{GL}_2(F)$-type abelian surfaces A/K with good reduction outside S_a, there is an $L \in \mathcal{L}$ for which the base changed $\mathrm{GL}_2(F)$-type abelian surface A/L is modular, in the sense that it corresponds to a parallel weight 2 Hilbert modular eigencuspform on GL_2/L. Then we are done: we deduce (via Jacquet-Langlands transfer) that A/L is an L-isogeny factor of the square of the Jacobian of an explicit Shimura curve with explicit level structure (depending on $N_{K,a}$, which by local-global compatibility bounds the level of the corresponding parallel weight 2 Hilbert eigencuspform on GL_2/L), whence we conclude via the usual combination of Masser-Wüstholz and Bost (i.e. Theorems 7.2.2 and 7.2.3 of Chapter 7): if B/L is an L-isogeny factor of C/L, then by Poincaré complete reducibility there is a B'/L with $C \sim_L B \times B'$, whence $h(B) = h(B \times B') - h(B')$ is bounded explicitly in terms of $h(C)$, $[L:\mathbb{Q}]$, and $\dim B$.

So it suffices to determine such a finite set of extensions $\mathcal{L}$. To do so we first observe, via Lemma 9.3.3 of Chapter 9, which is an explicit form of a result of Dimitrov, that, once $q \gg_{F,K,N_{K,a}} 1$ is sufficiently large, writing $\mathfrak{q}|(q)$ for a prime of F above the prime $q \in \mathbb{Z}^+$, the residual representations $\overline{\rho}_{\mathfrak{q},A/K} : \mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_F/\mathfrak{q})$ of all $\mathrm{GL}_2(F)$-type abelian surfaces A/K with good reduction outside S_a have image containing $\mathrm{SL}_2(\mathbb{F}_q)$ and thus satisfying the Taylor-Wiles hypothesis.

Choosing such a q, we determine the finitely many possible residual representations $\mathrm{Gal}(\overline{\mathbb{Q}}/K) \to \mathrm{GL}_2(\mathfrak{o}_F/\mathfrak{q})$ via Minkowski — after all they correspond to number fields of bounded degree (e.g. $\leq q^{10^{10}}$) and ramification (since they are unramified outside $S_a \cup \Sigma_q$). We then follow Snowden's [95] proof of Taylor's potential modularity theorem, which produces from the finitely many possible residual representations a finite set $\mathcal{L}'$ of totally real Galois extensions of K for which all K-simple, non-potentially-CM, $\mathrm{GL}_2(F)$-type abelian surfaces A/K are modular over an $L' \in \mathcal{L}'$. Each step is explicit, though there is one worth commenting on: in the key step of the argument Taylor uses Moret-Bailly's theorem to prove that a certain twist (depending on a given residual representation) of a Hilbert modular variety with level structure has a point over a totally real Galois extension of K — in our case one need only observe that one can therefore find such a point in finite time by simply e.g. enumerating rational points of larger and larger height and degree.

We then note, as Snowden does, that, by solvable descent for GL_2, writing $H \subseteq \mathrm{Gal}(L'/K)$ for a 2-Sylow subgroup (which is therefore solvable) and $L := (L')^H$, modularity of an A/K over L' implies modularity over L, which now has $[L : K]$ odd. Thus we determine an explicit finite set $\mathcal{L}$. As we saw above, by Masser-Wüstholz and Bost, this suffices.

11.5 Nontriviality of the example.

Naturally it would not be interesting to provide a finite-time algorithm determining the totally real points on $C := C_1$ if e.g. C only had finitely many totally real points at all. So we must check something to see that the example we have given is not tautological.

Now, by again applying Moret-Bailly's theorem, one knows that, because $C(\mathbb{R})$ is evidently infinite, there are infinitely many totally real points on C. However one can easily explicitly construct such totally real points, as follows.

One checks that, over e.g. the totally real cubic field $K := \mathbb{Q}[t]/(t^3 - 4t + 2)$, the elliptic curve over $\mathbb{Q}$ with plane cubic model $E : x^3 + 4y^3 = 1$ and marked point $(1, 0)$, which is isomorphic to $E_{0,-108} : y^2 = x^3 - 108$ over $\mathbb{Q}$ via $\alpha : E \simeq E_{0,-108}$ taking

$$(x, y) \mapsto \left(12 \cdot \frac{y}{1-x}, 18 \cdot \frac{1+x}{1-x}\right),$$

has positive rank, with an infinite-order point given by the image of[9]

$$P := (4 - \rho^2, 8 - 2\rho^2) \in E_{0,4}(K)$$

under the 3-isogeny $\beta : E_{0,4} \to E_{0,-108}$ via

$$(x, y) \mapsto \left(\frac{x^3 + 16}{x^2}, \frac{y \cdot (x^3 - 32)}{x^3}\right),$$

where we have written $\rho \in K$ for the image of $t \in \mathbb{Q}[t]$.

Now, because P, and thus $Q := \beta(P)$, is nontorsion, it follows from the pigeonhole principle that there are $n_i \in \mathbb{Z}^+$ with $n_1 < n_2 < \cdots$ such that, for all $\sigma : K \hookrightarrow \mathbb{R}$, $\sigma(n_i \cdot Q) \in E_{0,-108}(\mathbb{R})$ converges to $\alpha((1, 0)) = \infty \in E_{0,-108}(\mathbb{Q})$ as $i \to \infty$. Writing $(x_i, y_i) := \alpha^{-1}(n_i \cdot Q) \in E(K)$, it in particular follows that $x_i \in K$ is totally positive once $i \gg 1$. Of course because Q is nontorsion it also follows that the set $\{x_i\}_{i \in \mathbb{Z}^+} \subseteq K$ is infinite.

The desired infinite set of totally real points on C is simply

$$(\sqrt{x_i}, y_i) \in C(K(\sqrt{x_i})).$$

Finally let us show in similar fashion that there are infinitely many twists C_a with an odd-degree totally real point, so that Theorem 11.1.1 is nontrivial even over odd-degree totally real fields.

[9] Indeed, $(8 - 2t^2)^2 - ((4 - t^2)^3 + 4) = (t^3 - 4t + 2) \cdot (t^3 - 4t - 2)$ in $\mathbb{Z}[t]$.

We will instead use the totally real cubic field $L := \mathbb{Q}[t]/(t^3 - 5t + 1)$. One checks that the elliptic curve $E_{0,16}/\mathbb{Q}$ with Weierstrass model $E_{0,16} : y^2 = x^3 + 16$ has positive rank over L. Let $R \in E_{0,16}(L)$ be a nontorsion point. Write, for each $n \in \mathbb{Z}^+$, $(a_n, b_n) := n \cdot R \in E_{0,16}(L)$. Thus the set $\{b_n\}_{n \in \mathbb{Z}^+} \subseteq L$ is infinite, and one has that

$$1^6 + 4 \cdot \left(\frac{a_n}{4}\right)^3 = \left(\frac{b_n}{4}\right)^2,$$

so that $\left(1, \frac{a_n}{4}\right) \in C_{\frac{b_n}{4}}(L) \neq \emptyset$. Of course if one prefers one can clear denominators and use that L has class number one to write $a_n =: \frac{s_n}{d_n^2}, b_n =: \frac{t_n}{d_n^3}$ with $d_n, s_n, t_n \in \mathfrak{o}_L$ and $(d_n, s_n) = (d_n, t_n) = (1)$, so that the equality reads

$$(2d_n)^6 + 4 \cdot (s_n)^3 = (2t_n)^2,$$

i.e. $(2d_n, s_n) \in C_{2t_n}(L) \neq \emptyset$, but this is simply aesthetics.

Thus we see that the example was not completely trivial. We therefore conclude the chapter and this thesis.

Bibliography.

[1] A. Adrian Albert, *Algebras of degree 2^e and pure Riemann matrices*, Ann. of Math. (2) **33** (1932), no. 2, 311–318. MR1503054

[2] ______, *A solution of the principal problem in the theory of Riemann matrices*, Ann. of Math. (2) **35** (1934), no. 3, 500–515. MR1503176

[3] Patrick B. Allen, Frank Calegari, Ana Caraiani, Toby Gee, David Helm, Bao V. Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack A. Thorne, *Potential automorphy over CM fields* (2018), available at `https://arxiv.org/abs/1812.09999`.

[4] Levent Alpoge, *The average number of integral points on elliptic curves is bounded* (2014), available at `https://arxiv.org/abs/1412.1047`.

[5] ______, *The average number of rational points on genus two curves is bounded* (2018), available at `https://arxiv.org/abs/1804.05859`.

[6] Levent Alpoge, Manjul Bhargava, and Ari Shnidman, *Monogenic cubic fields*, forthcoming.

[7] Levent Alpoge and Wei Ho, *The second moment of the number of integral points on elliptic curves is bounded* (2018), available at `https://arxiv.org/abs/1807.03761`.

[8] Natália Archinard, *Hypergeometric abelian varieties*, Canad. J. Math. **55** (2003), no. 5, 897–932. MR2005278

[9] James Arthur and Laurent Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies, vol. 120, Princeton University Press, Princeton, NJ, 1989. MR1007299

[10] A. Baker, *Bounds for the solutions of the hyperelliptic equation*, Proc. Cambridge Philos. Soc. **65** (1969), 439–444. MR234912

[11] Henry Frederick Baker, *An introduction to the theory of multiply periodic functions*, Cambridge University Press, Cambridge, 1907. Digitized in 2007, original from Cabot Library at Harvard University.

[12] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), no. 2, 501–609. MR3152941

[13] G. V. Belyĭ, *Galois extensions of a maximal cyclotomic field*, Izv. Akad. Nauk SSSR Ser. Mat. **43** (1979), no. 2, 267–276, 479. MR534593

[14] Manjul Bhargava, *Higher composition laws*, ProQuest LLC, Ann Arbor, MI, 2001. Thesis (Ph.D.)–Princeton University. MR2702004

[15] ______, *The density of discriminants of quartic rings and fields*, Ann. of Math. (2) **162** (2005), no. 2, 1031–1063. MR2183288

[16] Manjul Bhargava and Benedict H. Gross, *The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point*, Automorphic representations and L-functions, 2013, pp. 23–91. MR3156850

[17] Manjul Bhargava and Wei Ho, *Coregular spaces and genus one curves*, Camb. J. Math. **4** (2016), no. 1, 1–119. MR3472915

[18] ______, *On average sizes of Selmer groups and ranks in families of elliptic curves having marked points* (2020), available at `http://www-personal.umich.edu/~weiho/papers/Selmer-averages-families.pdf`.

[19] Manjul Bhargava and Arul Shankar, *The average number of elements in the* 4*-selmer groups of elliptic curves is* 7 (2013), available at `https://arxiv.org/abs/1007.0052`.

[20] ______, *The average size of the* 5*-selmer group of elliptic curves is* 6*, and the average rank is less than* 1 (2013), available at `https://arxiv.org/abs/1312.7859`.

[21] ______, *Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves*, Ann. of Math. (2) **181** (2015), no. 1, 191–242. MR3272925

[22] ______, *Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank* 0, Ann. of Math. (2) **181** (2015), no. 2, 587–621. MR3275847

[23] Manjul Bhargava, Arul Shankar, and Xiaoheng Wang, *Geometry-of-numbers methods over global fields II: Coregular representations*, forthcoming.

[24] B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves. I*, J. Reine Angew. Math. **212** (1963), 7–25. MR146143

[25] Don Blasius, *Elliptic curves, Hilbert modular forms, and the Hodge conjecture*, Contributions to automorphic forms, geometry, and number theory, 2004, pp. 83–103. MR2058605

[26] Fedor Bogomolov and Yuri Tschinkel, *Unramified correspondences*, Algebraic number theory and algebraic geometry, 2002, pp. 17–25. MR1936365

[27] Oskar Bolza, *Darstellung der rationalen ganzen Invarianten der Binärform sechsten Grades durch die Nullwerthe der zugehörigen ϑ-Functionen*, Math. Ann. **30** (1887), no. 4, 478–495. MR1510458

[28] Enrico Bombieri, Andrew Granville, and János Pintz, *Squares in arithmetic progressions*, Duke Math. J. **66** (1992), no. 3, 369–385. MR1167100

[29] Enrico Bombieri and Walter Gubler, *Heights in Diophantine geometry*, New Mathematical Monographs, vol. 4, Cambridge University Press, Cambridge, 2006. MR2216774

[30] M. V. Borovoĭ, *Langlands' conjecture concerning conjugation of connected Shimura varieties*, 1983/84, pp. 3–39. Selected translations. MR732450

[31] Jean-Benoît Bost, *Périodes et isogenies des variétés abéliennes sur les corps de nombres (d'après D. Masser et G. Wüstholz)*, 1996, pp. Exp. No. 795, 4, 115–161. Séminaire Bourbaki, Vol. 1994/95. MR1423622

[32] Jean-François Boutot, Lawrence Breen, Paul Gérardin, Jean Giraud, Jean-Pierre Labesse, James Stuart Milne, and Christophe Soulé, *Variétés de Shimura et fonctions L*, Publications Mathématiques de l'Université Paris VII [Mathematical Publications of the University of Paris VII], vol. 6, Université de Paris VII, U.E.R. de Mathématiques, Paris, 1979. MR680404

[33] Tim Browning and Roger Heath-Brown, *The geometric sieve for quadrics* (2020), available at `https://arxiv.org/abs/2003.09593`.

[34] Armand Brumer and Kenneth Kramer, *The conductor of an abelian variety*, Compositio Math. **92** (1994), no. 2, 227–248. MR1283229

[35] Henri Carayol, *Sur les représentations l-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. École Norm. Sup. (4) **19** (1986), no. 3, 409–468. MR870690

[36] J. W. S. Cassels and E. V. Flynn, *Prolegomena to a middlebrow arithmetic of curves of genus 2*, London Mathematical Society Lecture Note Series, vol. 230, Cambridge University Press, Cambridge, 1996. MR1406090

[37] Ching-Li Chai, Brian Conrad, and Frans Oort, *Complex multiplication and lifting problems*, Mathematical Surveys and Monographs, vol. 195, American Mathematical Society, Providence, RI, 2014. MR3137398

[38] Claude Chevalley and André Weil, *Un théorème d'arithmétique sur les courbes algébriques*, C. R. Acad. Sci. Paris **195** (1932), 570–572.

[39] L. Clozel and C. S. Rajan, *Solvable base change* (2018), available at `https://arxiv.org/abs/1806.02513`.

[40] Paula Cohen and Jürgen Wolfart, *Modular embeddings for some nonarithmetic Fuchsian groups*, Acta Arith. **56** (1990), no. 2, 93–110. MR1075639

[41] Paula Beazley Cohen, Claude Itzykson, and Jürgen Wolfart, *Fuchsian triangle groups and Grothendieck dessins. Variations on a theme of Belyĭ*, Comm. Math. Phys. **163** (1994), no. 3, 605–627. MR1284798

[42] Alyson Deines, Jenny G. Fuselier, Ling Long, Holly Swisher, and Fang-Ting Tu, *Generalized Legendre curves and quaternionic multiplication*, J. Number Theory **161** (2016), 175–203. MR3435724

[43] Lassina Dembélé and John Voight, *Explicit methods for Hilbert modular forms*, Elliptic curves, Hilbert modular forms and Galois deformations, 2013, pp. 135–198. MR3184337

[44] Mladen Dimitrov, *Galois representations modulo p and cohomology of Hilbert modular varieties*, Ann. Sci. École Norm. Sup. (4) **38** (2005), no. 4, 505–551. MR2172950

[45] J.-H. Evertse, *On equations in S-units and the Thue-Mahler equation*, Invent. Math. **75** (1984), no. 3, 561–584. MR735341

[46] J.-H. Evertse and J. H. Silverman, *Uniform bounds for the number of solutions to $Y^n = f(X)$*, Math. Proc. Cambridge Philos. Soc. **100** (1986), no. 2, 237–248. MR848850

[47] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366. MR718935

[48] Laurent Fargues, *Motives and automorphic forms: the (potentially) abelian case*, 2006. Notes from the 2006 IHES-Orsay Summer School. `https://webusers.imj-prg.fr/~laurent.fargues/Motifs_abeliens.pdf` (accessed February 2020).

[49] Victor Flynn, *My Genus 2 Site*, available at `https://people.maths.ox.ac.uk/flynn/genus2/`. Accessed: 2018-01-22.

[50] Éric Gaudron and Gaël Rémond, *Polarisations et isogénies*, Duke Math. J. **163** (2014), no. 11, 2057–2108. MR3263028

[51] ______, *Théorème des périodes et degrés minimaux d'isogénies*, Comment. Math. Helv. **89** (2014), no. 2, 343–403. MR3225452

[52] J. W. L. Glaisher, *Tables of $1 \pm 2^{-n} + 3^{-n} \pm 4^{-n} + \&c.$ and $1 + 3^{-n} + 5^{-n} + 7^{-n} + \&c.$ to 32 places of decimals*, Quarterly Journal of Pure and Applied Mathematics **45** (1914), 141–158, available at `https://babel.hathitrust.org/cgi/pt?id=uc1.$b417565&view=1up&seq=11`.

[53] David Grant, *Formal groups in genus two*, J. Reine Angew. Math. **411** (1990), 96–121. MR1072975

[54] D. R. Heath-Brown, *A new form of the circle method, and its application to quadratic forms*, J. Reine Angew. Math. **481** (1996), 149–206. MR1421949

[55] H. A. Helfgott and A. Venkatesh, *Integral points on elliptic curves and 3-torsion in class groups*, J. Amer. Math. Soc. **19** (2006), no. 3, 527–550. MR2220098

[56] Guy Henniart, *Représentations l-adiques abéliennes*, Seminar on Number Theory, Paris 1980-81 (Paris, 1980/1981), 1982, pp. 107–126. MR693314

[57] Haruzo Hida, *On abelian varieties with complex multiplication as factors of the Jacobians of Shimura curves*, Amer. J. Math. **103** (1981), no. 4, 727–776. MR623136

[58] Jun-ichi Igusa, *Modular forms and projective invariants*, Amer. J. Math. **89** (1967), 817–855. MR0229643

[59] H. Jacquet and R. P. Langlands, *Automorphic forms on* $\mathrm{GL}(2)$, Lecture Notes in Mathematics, Vol. 114, Springer-Verlag, Berlin-New York, 1970. MR0401654

[60] G. A. Kabatjanskiĭ and V. I. Levenšteĭn, *Bounds for packings on the sphere and in space*, Problemy Peredači Informacii **14** (1978), no. 1, 3–25. MR0514023

[61] Chandrashekhar Khare and Jean-Pierre Wintenberger, *Serre's modularity conjecture. I*, Invent. Math. **178** (2009), no. 3, 485–504. MR2551763

[62] ———, *Serre's modularity conjecture. II*, Invent. Math. **178** (2009), no. 3, 505–586. MR2551764

[63] Helmut Klingen, *Introductory lectures on Siegel modular forms*, Cambridge Studies in Advanced Mathematics, vol. 20, Cambridge University Press, Cambridge, 1990. MR1046630

[64] Robert P. Langlands, *Base change for* $\mathrm{GL}(2)$, Annals of Mathematics Studies, vol. 96, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1980. MR574808

[65] Erez Lapid and Jonathan Rogawski, *On twists of cuspidal representations of* $\mathrm{GL}(2)$, Forum Math. **10** (1998), no. 2, 175–197. MR1611951

[66] Aaron Levin, *Siegel's theorem and the Shafarevich conjecture*, J. Théor. Nombres Bordeaux **24** (2012), no. 3, 705–727. MR3010636

[67] D. W. Masser and G. Wüstholz, *Endomorphism estimates for abelian varieties*, Math. Z. **215** (1994), no. 4, 641–653. MR1269495

[68] David Masser and Gisbert Wüstholz, *Isogeny estimates for abelian varieties, and finiteness theorems*, Ann. of Math. (2) **137** (1993), no. 3, 459–472. MR1217345

[69] J. S. Milne, *The action of an automorphism of* **C** *on a Shimura variety and its special points*, Arithmetic and geometry, Vol. I, 1983, pp. 239–265. MR717596

[70] L. J. Mordell, *The diophantine equation* $y^2 - k = x^3$, Quarterly Journal of Pure and Applied Mathematics **45** (1914), 170–186, available at `https://babel.hathitrust.org/cgi/pt?id=uc1.$b417565&view=1up&seq=11`.

[71] Laurent Moret-Bailly, *Points entiers des variétés arithmétiques*, Séminaire de Théorie des Nombres, Paris 1985–86, 1987, pp. 147–154. MR1017909

[72] ———, *Groupes de Picard et problèmes de Skolem. I, II*, Ann. Sci. École Norm. Sup. (4) **22** (1989), no. 2, 161–179, 181–194. MR1005158

[73] David Mumford, *A remark on Mordell's conjecture*, Amer. J. Math. **87** (1965), 1007–1016. MR186624

[74] ———, *Tata lectures on theta. II*, Progress in Mathematics, vol. 43, Birkhäuser Boston, Inc., Boston, MA, 1984. Jacobian theta functions and differential equations, With the collaboration of C. Musili, M. Nori, E. Previato, M. Stillman and H. Umemura. MR742776

[75] M. Ram Murty and Hector Pasten, *Modular forms and effective Diophantine approximation*, J. Number Theory **133** (2013), no. 11, 3739–3754. MR3084298

[76] Michael Musty, Sam Schiavone, Jeroen Sijsling, and John Voight, *A database of Belyi maps*, Proceedings of the Thirteenth Algorithmic Number Theory Symposium, 2019, pp. 375–392. MR3952023

[77] Stefan Patrikis, José Felipe Voloch, and Yuri G. Zarhin, *Anabelian geometry and descent obstructions on moduli spaces*, Algebra Number Theory **10** (2016), no. 6, 1191–1219. MR3544295

[78] Fabien Pazuki, *Theta height and Faltings height*, Bull. Soc. Math. France **140** (2012), no. 1, 19–49. MR2903770

[79] ———, *Minoration de la hauteur de Néron-Tate sur les surfaces abéliennes*, Manuscripta Math. **142** (2013), no. 1-2, 61–99. MR3081000

[80] Bjorn Poonen, *Unramified covers of Galois covers of low genus curves*, Math. Res. Lett. **12** (2005), no. 4, 475–481. MR2155225

[81] Michel Raynaud, *Hauteurs et isogénies*, 1985, pp. 199–234. Seminar on arithmetic bundles: the Mordell conjecture (Paris, 1983/84). MR801923

[82] K. A. Ribet, *Endomorphism algebras of abelian varieties attached to newforms of weight* 2, Seminar on Number Theory, Paris 1979–80, 1981, pp. 263–276. MR633903

[83] Kenneth A. Ribet, *Abelian varieties over* **Q** *and modular forms*, Algebra and topology 1992 (Taejŏn), 1992, pp. 53–79. MR1212980

[84] B. Riemann, *Ueber das Verschwinden der ϑ-Functionen*, J. Reine Angew. Math. **65** (1866), 161–172. MR1579313

[85] Robert S. Rumely, *Arithmetic over the ring of all algebraic integers*, J. Reine Angew. Math. **368** (1986), 127–133. MR850618

[86] ______, *Capacity theory on algebraic curves*, Lecture Notes in Mathematics, vol. 1378, Springer-Verlag, Berlin, 1989. MR1009368

[87] Samuel Ruth, *A bound on the average rank of j-invariant zero elliptic curves*, ProQuest LLC, Ann Arbor, MI, 2013. Thesis (Ph.D.)–Princeton University. MR3211431

[88] Edward F. Schaefer, *Class groups and Selmer groups*, J. Number Theory **56** (1996), no. 1, 79–114. MR1370197

[89] Edward Frank Schaefer, *Class groups and Selmer groups*, ProQuest LLC, Ann Arbor, MI, 1992. Thesis (Ph.D.)–University of California, Berkeley. MR2688490

[90] Norbert Schappacher, *Periods of Hecke characters*, Lecture Notes in Mathematics, vol. 1301, Springer-Verlag, Berlin, 1988. MR935127

[91] Jean-Pierre Serre, *Abelian l-adic representations and elliptic curves*, Research Notes in Mathematics, vol. 7, A K Peters, Ltd., Wellesley, MA, 1998. With the collaboration of Willem Kuyk and John Labute, Revised reprint of the 1968 original. MR1484415

[92] Arul Shankar and Xiaoheng Wang, *Rational points on hyperelliptic curves having a marked non-Weierstrass point*, Compos. Math. **154** (2018), no. 1, 188–222. MR3719247

[93] Goro Shimura, *On analytic families of polarized abelian varieties and automorphic functions*, Ann. of Math. (2) **78** (1963), 149–192. MR156001

[94] ______, *On the real points of an arithmetic quotient of a bounded symmetric domain*, Math. Ann. **215** (1975), 135–164. MR572971

[95] Andrew Snowden, *On two dimensional weight two odd representations of totally real fields* (2009), available at `https://arxiv.org/abs/0905.4266`.

[96] Michael Stoll, *On the height constant for curves of genus two*, Acta Arith. **90** (1999), no. 2, 183–201. MR1709054

[97] ______, *On the height constant for curves of genus two. II*, Acta Arith. **104** (2002), no. 2, 165–182. MR1914251

[98] Marco Streng, *Computing Igusa class polynomials*, Math. Comp. **83** (2014), no. 285, 275–309. MR3120590

[99] P. Tannery, *Diophanti alexandrini opera omnia cum graecis commentariis*, Bibliotheca scriptorum Graecorum et Romanorum Teubneriana, In aedibus B.G. Teubneri, 1893.

[100] John Tate, *Duality theorems in Galois cohomology over number fields*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), 1963, pp. 288–295. MR0175892

[101] Richard Taylor, *Representations of Galois groups associated to modular forms*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 1995, pp. 435–442. MR1403943

[102] Ila Kapur Varma, *On local-global compatibility for cuspidal regular algebraic automorphic representations of* GL_n, ProQuest LLC, Ann Arbor, MI, 2015. Thesis (Ph.D.)–Princeton University. MR3450147

[103] Paul Vojta, *Siegel's theorem in the compact case*, Ann. of Math. (2) **133** (1991), no. 3, 509–548. MR1109352

[104] Rafael von Känel, *Integral points on moduli schemes of elliptic curves*, Trans. London Math. Soc. **1** (2014), no. 1, 85–115. MR3296485

[105] Michel Waldschmidt, *Transcendance et exponentielles en plusieurs variables*, Invent. Math. **63** (1981), no. 1, 97–127. MR608530

[106] André Weil, *Oeuvres scientifiques/Collected papers. I. 1926–1951*, Springer Collected Works in Mathematics, Springer, Heidelberg, 2014. Reprint of the 2009 [MR2883738] and 1979 [MR0537937] editions. MR3328832

[107] Jürgen Wolfart, *Triangle groups and Jacobians of CM type* (2000), available at `https://www.math.uni-frankfurt.de/~wolfart/Artikel/jac.pdf`.

[108] X, *The integer solutions of the equation* $y^2 = ax^n + bx^{n-1} + \cdots + k$, Journal of the London Mathematical Society **s1-1** (1926), no. 2, 66–68, available at `https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/jlms/s1-1.2.66`. (Extract from a letter from C. L. Siegel to L. J. Mordell.)

[109] Hiroyuki Yoshida, *ON THE REPRESENTATIONS OF THE GALOIS GROUPS OBTAINED FROM HILBERT MODULAR-FORMS*, ProQuest LLC, Ann Arbor, MI, 1973. Thesis (Ph.D.)–Princeton University. MR2623917

[110] Kentaro Yoshitomi, *On height functions on Jacobian surfaces*, Manuscripta Math. **96** (1998), no. 1, 37–66. MR1624348