The second moment of the number of integral points on elliptic curves is bounded

Levent Alpöge and Wei Ho

Abstract. Let K be a number field and S a finite set of places of K containing all archimedean places. In this paper, we show that the second moment of the number of S-integral points on elliptic curves over K is bounded. In particular, we prove that, for any positive real number $\vartheta \leq \log_2 5 = 2.3219\ldots$, the ϑ-th moment of the number of S-integral points is bounded for the family of all integral short Weierstrass curves ordered by naive height, or for any positive density subfamily thereof. For certain other families of elliptic curves over $\mathbb{Q}$, such as those with a marked point or two marked points, we prove that the average of the number of integral points is bounded.

The main new ingredient in our proof, which is of independent interest, is an upper bound on the number of S-integral points on an affine integral Weierstrass model of an elliptic curve over K depending only on the rank of the curve, the class group and degree of K, and the number of primes of K whose *square* divides the discriminant of the curve. For example, the number of integral points on an affine integral Weierstrass model of an elliptic curve E over $\mathbb{Q}$ is bounded by $2^{\mathrm{rank}(E)}O(1)^s$, where s is the number of prime squares dividing the discriminant of E. We obtain this bound by generalizing a bijection first observed by Mordell between integral points on these curves and certain types of binary quartic forms.

The theorems on moments then follow from this new upper bound and results on bounds on the average sizes of Selmer groups in the families. In order to prove the bounds for the ϑ-th moment for the edge case when $\vartheta = \log_2 5$ (and the analogous cases for the other families), we introduce a method to count orbits of coregular representations with weights.

1. Introduction

In this paper, we prove several theorems about the number of integral points on elliptic curves over a number field K. We bound the number of integral points using only the rank of the elliptic curve, the higher order divisors of its discriminant, and the degree and class number of K. For any finite set S of places of K containing all archimedean places, we also obtain an analogous bound for S-integral points. Using this bound, we show that the second moment of the number of (S-)integral points on elliptic curves over K is bounded. To our knowledge, this is the first instance of an unconditional bound on higher moments for arithmetic data on elliptic curves.

We first give an explicit upper bound on the number of integral points on affine integral Weierstrass models of elliptic curves over $\mathbb{Q}$, depending only on the rank and the number of square divisors of the discriminant of the curve:

Theorem 1.1. *Let $A, B \in \mathbb{Z}$ be such that $\Delta_{A,B} := -16(4A^3 + 27B^2) \neq 0$. Let $\mathcal{E}_{A,B}$ be the affine integral model $y^2 = x^3 + Ax + B$ of the associated elliptic curve $E_{A,B}$ over $\mathbb{Q}$. Then*

$$|\mathcal{E}_{A,B}(\mathbb{Z})| \ll 2^{\operatorname{rank} E_{A,B}(\mathbb{Q})} \prod_{p^2 \mid \Delta_{A,B}} \min\left(4\left\lfloor \frac{v_p(\Delta_{A,B})}{2} \right\rfloor + 1, 7^{2^7}\right).$$

Here v_p denotes the p-adic valuation for a prime p, and $\mathcal{E}_{A,B}(R)$ denotes the set of solutions $\{(x,y) \in R^2 : y^2 = x^3 + Ax + B\}$ for any ring R. By $f \ll g$ we mean that there is a positive absolute constant $c > 0$ such that $|f| \leq c|g|$.

Remark 1.2. If $v_p(\Delta_{A,B}) = 2$ or 3, the factor in Theorem 1.1 for p may be improved to 4 (rather than 5). This results from a slightly more careful analysis of the p-adic argument at the end of Bombieri–Schmidt's paper [BS87]; see Remark 3.6.

Our general bound for S-integral points for an elliptic curve over a number field K is as follows:

Theorem 1.3. *Fix $C = 7^{2^7}$. Let K be a number field, and let $\mathcal{O}_K$ denote its ring of integers. Let $A, B \in \mathcal{O}_K$ 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 $\mathfrak{p}^2 \mid \Delta_{A,B}$, and let $\mathcal{O}_{K,S}$ denote the ring of S-integers in K, and let $\mathrm{Cl}(R)$ denote the class group of the ring R.*

Let $\mathcal{E}_{A,B} : y^2 = x^3 + Ax + B$ be the affine Weierstrass model of the associated elliptic curve $E_{A,B}$ over K. Then we have the bound

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

Remark 1.4. There are several weaker but simpler variants of the bound (1). First, since $\mathrm{Cl}(\mathcal{O}_{K,S})$ is a quotient of $\mathrm{Cl}(\mathcal{O}_K)$, one may replace $\mathrm{Cl}(\mathcal{O}_{K,S})[2]$ with $\mathrm{Cl}(\mathcal{O}_K)[2]$. Also, clearly Theorem 1.3 implies the weaker bound

$$|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})| \leq 2^{\operatorname{rank} E_{A,B}(K)} C^{2|S|+1} h_K,$$

where h_K denotes the class number of K.

Moreover, taking S to be as small as possible, namely the union of the infinite places and $\omega_{\geq 2}(\Delta_{A,B}) := \{\mathfrak{p} : v_{\mathfrak{p}}(\Delta_{A,B}) \geq 2\}$, we obtain the following bound on integral ($\mathcal{O}_K$) points:

$$|\mathcal{E}_{A,B}(\mathcal{O}_K)| \leq 2^{\operatorname{rank} E_{A,B}(K)} C^{2[K:\mathbb{Q}] + 2\omega_{\geq 2}(\Delta_{A,B}) + 1} \left|\mathrm{Cl}(\mathcal{O}_K)[2]\right|.$$

Remark 1.5. Note that the right-hand side of (1) must depend on S: if $\operatorname{rank}(E_{A,B}(K)) > 0$, the left-hand side may be made arbitrarily large by expanding S.

Mordell was the first to prove the finiteness of the number of integral points on an elliptic curve (essentially by the invariant-theoretic method we employ in this paper), a theorem generalized by Siegel to all curves of genus $g \geq 1$. Previous upper bounds on the number of integral points on elliptic curves have similar shapes but are not suitable for our applications on moments; our bound is significantly stronger on average. For example, Helfgott and Venkatesh [HV06] show that, for any integral model $\mathcal{E}_{A,B}$ of the associated curve $E_{A,B}$ over $\mathbb{Q}$,

$$|\mathcal{E}_{A,B}(\mathbb{Z})| \ll 1.33^{\operatorname{rank} E_{A,B}(\mathbb{Q})} O(1)^{\omega(\Delta_{A,B})} (\log |\Delta_{A,B}|)^2$$

where $\omega(n)$ denotes the number of distinct prime factors of n.[1] For quasi-minimal short Weierstrass models $\mathcal{E}_{A,B}$ of elliptic curves $E_{A,B}$ over K (i.e., such that $\mathrm{Norm}_{K/\mathbb{Q}}(\Delta_{A,B})$ is minimized subject to $A, B \in \mathcal{O}_K$), Silverman [Sil87] shows that

$$|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})| \ll O_K(1)^{(1+\operatorname{rank} E_{A,B}(K))(1+\omega_{\mathrm{ss}}(\Delta_{A,B}))+|S|}$$

where $\omega_{\mathrm{ss}}(\Delta_{A,B})$ denotes the number of primes of semistable bad reduction and the $O(1)$ is already on the order of 10^{10} for $K = \mathbb{Q}$. In fact, stronger bounds[2] for S-integral points on quasi-minimal short Weierstrass models $\mathcal{E}_{A,B}$ over number fields were proven by Hindry–Silverman [HS88]:

$$|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})| \ll O_K(1)^{(1+\operatorname{rank} E_{A,B}(K))(1+\sigma_{A,B})+|S|}, \tag{2}$$

and for curves with $h(\Delta_{A,B}) \geq \frac{1}{2} h(j(E_{A,B}))$, by David [Dav92]:

$$|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})| \ll O_K(\sigma_{A,B}^6 (\log(1 + \sigma_{A,B}))^3)^{1+\operatorname{rank} E_{A,B}(K)+|S|}.$$

[1] Even if this type of bound were generalized to other number fields K, it would not be strong enough for our application on moments because of the $O(1)^{\omega(\Delta_{A,B})}(\log|\Delta_{A,B}|)^2$ factor.

[2] Since one has control on the average size of n-Selmer groups only for small n, and thus control on the average size of $n^{\operatorname{rank} E_{A,B}}$ only for small n, these bounds are also unsuitable for our application to moments.

Here $\sigma_{A,B} := \frac{\log \mathrm{Norm}_{K/\mathbb{Q}}(\Delta_{A,B})}{\log \mathrm{Norm}_{K/\mathbb{Q}}(N_{A,B})}$ is the Szpiro ratio of $E_{A,B}$ over K and $N_{A,B}$ denotes the conductor of $E_{A,B}$. Since the ABC conjecture implies that the Szpiro ratio is at most $6 + o(1)$, the Hindry–Silverman bound (2) implies that, conditional on ABC and uniform boundedness of ranks for elliptic curves over K, the number of S-integral points is uniformly bounded for quasi-minimal elliptic curves over a fixed K. (In fact, all one needs is Lang's conjecture that $\hat{h}(P) \gg h(E_{A,B})$ for non-torsion $P \in E_{A,B}(K)$.[3])

We use Theorem 1.3 to prove that the second moment of the number of S-integral points on elliptic curves over K is bounded.

In particular, we consider the family $\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ of all integral Weierstrass models

$$\mathcal{E}_{A,B} : y^2 = x^3 + Ax + B$$

of elliptic curves over K, where $A, B \in \mathcal{O}_K$ with $\Delta_{A,B} \neq 0$, and order this family by *height*

$$H(\mathcal{E}_{A,B}) = H(A,B) := \prod_{v \mid \infty} \max(|A|_v^{1/4}, |B|_v^{1/6}). \tag{3}$$

We note that the same boundedness-of-moments results also hold when ordering elliptic curves over K instead by the usual height on the weighted projective line $\mathbb{P}(4,6)$, which gives the height $\widetilde{H}(\mathcal{E}_{A,B}) := \mathrm{Norm}(I(A,B)) \prod_{v\mid\infty} \max(|A|_v^{1/4}, |B|_v^{1/6})$ where $I(A,B) := \{a \in K : a^4 A, a^6 B \in \mathcal{O}_K\} \subseteq K$. See Remark 4.2.

Not only do we prove that the second moment of the number of integral points in this family is bounded, we obtain the following slightly stronger result.

Theorem 1.6. *Fix $C = 7^{2^7}$. Let K be a number field, and let S be a finite set of places of K containing all infinite places. Let $\mathscr{F}$ be a subset of $\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ of positive lower density (ordering by height), and let $0 < \vartheta \leq \log_2 5 = 2.3219\ldots$ a positive real number. We have*

$$\mathrm{Avg}_{\mathscr{F}}(|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}) \ll_{\mathscr{F}} \left(C^{2|S|}\, |\mathrm{Cl}(\mathcal{O}_{K,S})[2]|\right)^{\vartheta} \tag{4}$$

where the average is taken over all $\mathcal{E}_{A,B} \in \mathscr{F}$ ordered by height.

More precisely, let

$$\mathscr{F}^{\leq T} := \{(A,B) \in \mathscr{F} : \Delta_{A,B} \neq 0, H(A,B) \leq T\}$$

be the set of all $(A,B) \in \mathscr{F}$ with naive height up to T. Then there exists a constant $c_{\mathscr{F}}$, depending only on $\mathscr{F}$, such that

$$\limsup_{T\to\infty} \frac{\sum_{(A,B)\in\mathscr{F}^{\leq T}} |\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}}{|\mathscr{F}^{\leq T}|} \leq c_{\mathscr{F}} \left(C^{2|S|} |\mathrm{Cl}(\mathcal{O}_{K,S})[2]|\right)^{\vartheta}.$$

In fact, we may take $c_{\mathscr{F}}$ to be

$$O\left(\limsup_{T\to\infty} \frac{|\mathscr{F}_{\mathrm{univ}}^{\leq T}(\mathcal{O}_K)|}{|\mathscr{F}^{\leq T}|}\right),$$

and we may even replace the limit by $\limsup_{T\to\infty} \frac{|\mathscr{G}^{\leq T}|}{|\mathscr{F}^{\leq T}|}$ where $\mathscr{F} \subseteq \mathscr{G}$ and $\mathscr{G}$ is a "large family" in the sense of [BSW22] (and thus, for a large family $\mathscr{F}$, we make take $c_{\mathscr{F}} = c_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)}$).

[3]Alternatively, Abramovich [Abr97] has shown that the Lang–Vojta conjecture for varieties of log general type implies uniform boundedness of the number of S-integral points on a stably minimal model of an elliptic curve.

Remark 1.7. More generally, from Theorem 1.3, we have that the same moments of S-integral points on elliptic curves are bounded even when the set S is allowed to vary with (A, B), as long as the number of primes in the set $S(A, B)$ for each (A, B) does not grow too quickly (indeed, so long as $\operatorname{Avg} t^{|S(A,B)|} < \infty$ for a sufficiently large constant $t \ll_\vartheta 1$).

One expects that elliptic curves should have no "unexpected points" on average, i.e., that all these moments should be 0 (note that the point at infinity is not an integral point). In [Alp14], it is proved that for $0 < \vartheta < \log_3 5 = 1.4649\ldots$, the average in (4) for integral points on elliptic curves over $K = \mathbb{Q}$ is bounded (and thus by taking $\vartheta = 1$, that the average number of integral points is bounded, a result also proved by D. Kim [Kim18]).

Remark 1.8. A related but different question is to show that most elliptic curves have very few integral points; perhaps the strongest known result in this direction is that 80% of curves in $\mathscr{F}_{\text{univ}}$ have at most 2 integral points (by combining the fact that 100% of rank 1 curves in $\mathscr{F}_{\text{univ}}$ have at most 2 points [Alp14, Lemma 20] with Bhargava–Shankar's result [BS13b] that at least 80% of curves in $\mathscr{F}_{\text{univ}}$ have rank 0 or 1). Note that these bounds do not imply that the average number of integral points is bounded, since it is a priori possible that there is some small exceptional subset in which the curves have an enormous number of points.

Remark 1.9. Theorem 1.6 gives a bound on $\operatorname{Avg}(|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^\vartheta)$ for $\vartheta \leq \log_2 5$ by using Bhargava–Shankar's bound on the average size of 5-Selmer groups over this family [BS13b]. If a bound on the average size of n-Selmer groups over this family were known, the same argument would yield a similar bound for all $\vartheta \leq \log_2(n)$.

Remark 1.10. The equality case $\vartheta = \log_2 5$ of Theorem 1.6 is treated by modifying the proof of Bhargava–Shankar's average 5-Selmer bound to use *weighted* 5-Selmer elements instead; see Section 5. To the best of our knowledge, this is the first time that these geometry-of-numbers techniques in arithmetic statistics have been used to count weighted orbits of a representation, and we believe that this generalization may be useful in other applications.

For two other families of elliptic curves over $\mathbb{Q}$, i.e., the families

$$\begin{aligned} &\mathscr{F}_1 := \{y^2 + d_3 y = x^3 + d_2 x^2 + d_4 x \mid d_2, d_3, d_4 \in \mathbb{Z},\ \Delta \neq 0\} \\ \text{and} \qquad &\mathscr{F}_2 := \{y^2 + d_1 xy + d_3 y = (x - d_2)(x - d_2')(x - d_2'') \mid \\ &\qquad\qquad d_1, d_2, d_2', d_2'', d_3 \in \mathbb{Z},\ d_2 + d_2' + d_2'' = 0,\ \Delta \neq 0\}, \end{aligned}$$

of elliptic curves in Weierstrass form with one and two marked points, respectively, ordered by an analogous notion of height, we also find that the average of the number of integral points is bounded.

Theorem 1.11. *Let $\mathscr{F}$ be a subfamily of elliptic curves in $\mathscr{F}_1$ (respectively, $\mathscr{F}_2$) of positive lower density. For any positive real number $\vartheta \leq \log_2 3 = 1.5850\ldots$ (resp., $\vartheta \leq 1$), we have*

$$\operatorname{Avg}_{\mathcal{E} \in \mathscr{F}}(|\mathcal{E}(\mathbb{Z})|^\vartheta) \ll_{\vartheta, \mathscr{F}} 1$$

where the average is taken over all $\mathcal{E} \in \mathscr{F}$ ordered by height.

Method of proof. Theorem 1.1 follows from studying a bijection first observed by Mordell between integral points on an integral Weierstrass model $\mathcal{E}_{A,B}$ of an elliptic curve and binary quartics of the form $X^4 + 6cX^2Y^2 + 8dXY^3 + eY^4$ with $c, d, e \in \mathbb{Z}$ and invariants $I = -48A$ and $J = -1728B$. The natural map taking the integral point to an element of the 2-Selmer group of the elliptic curve translates precisely to taking the corresponding binary quartic to its $\mathrm{PGL}_2(\mathbb{Q})$-equivalence class.

By working explicitly (and using results of Bombieri–Schmidt [BS87] and Evertse [Eve97] on Thue equations), we bound the size of a fibre of this map by

$$\ll \prod_{p^2 \mid \Delta_{A,B}} \min\left\{4\left\lfloor \frac{v_p(\Delta_{A,B})}{2} \right\rfloor + 1,\, 7^{2^7}\right\}.$$

The image lies in $E_{A,B}(\mathbb{Q})/2E_{A,B}(\mathbb{Q})$, whose size is at most $\leq 4 \cdot 2^{\operatorname{rank} E_{A,B}(\mathbb{Q})}$, giving the theorem. Theorem 1.3 follows the same basic strategy, after generalizing the algebraic constructions to S-integral points over K and using the fact that Evertse in fact treats solutions to Thue-Mahler equations over $\mathcal{O}_{K,S}$.

Theorem 1.6 for $\vartheta < \log_2 5$ follows fairly straightforwardly from Theorem 1.3, Hölder's inequality, standard analytic techniques, and knowledge of bounds on the average sizes of 5-Selmer groups in this family (these bounds over $\mathbb{Q}$ come from Bhargava–Shankar's work [BS13b] and have been extended to number fields K by Bhargava–Shankar–Wang [BSW22]). As mentioned in Remark 1.10, the equality case $\vartheta = \log_2 5$ is treated by modifying the geometry-of-numbers proof for the average 5-Selmer bound to count *weighted* 5-Selmer elements.

For the families $\mathscr{F}_1$ and $\mathscr{F}_2$, the results on moments, for $\vartheta \leq \log_2 3$ or ≤ 1, respectively, follow from the same techniques as for Theorem 1.6 and bounds on the average 3-Selmer (resp., 2-Selmer) group size from [BH22]. Note that the equality case is of most interest for the family $\mathscr{F}_2$, where it gives an upper bound for the average number of integral points on curves in $\mathscr{F}_2$. We expect that identical results on bounding moments of integral points for these types of families over any number field hold, by using [BSW22] to generalize the average Selmer theorems in [BH22] to number fields.

Acknowledgments. We thank Shabnam Akhtari, Noah Albrecht, Pieter Belmans, Manjul Bhargava, Zev Klagsbrun, Erik Metz, Arul Shankar, and Joe Silverman for helpful comments and conversations. Several questions by Joe Silverman inspired us to add more general results over arbitrary number fields and for S-integral points. LA was supported by NSF grant DMS-2002109 and the Society of Fellows, and WH was supported by NSF grants DMS-1701437 and DMS-1844763, the Sloan Foundation, and the Minerva Research Foundation.

Notation. For the remainder of the paper, let K be a number field and let $\mathcal{O}_K$ be its ring of integers. Any ideal denoted $\mathfrak{p}$ is prime. For a finite set of places S of K, let $\mathcal{O}_{K,S}$ be the ring of S-integers. In the sequel, as is standard, an element of K is called *integral* if it is in $\mathcal{O}_K$, and *S-integral* if it is in $\mathcal{O}_{K,S}$. Affine Weierstrass models over $\mathcal{O}_K$ of elliptic curves are denoted $\mathcal{E}$ or $\mathcal{E}_{A,B}$, and the associated elliptic curve over K is denoted E. We will sometimes, by a small abuse of notation, refer to $\mathcal{E}$ as an elliptic curve.

2. Binary quartic forms and integral points on elliptic curves

2.1. Preliminaries on binary quartic forms.

Given a binary quartic form

$$f(X,Y) = aX^4 + bX^3Y + cX^2Y^2 + dXY^3 + eY^4 \tag{5}$$

with coefficients in K, the group $\mathrm{SL}_2(K)$ naturally acts by linear substitutions of the variables, i.e., for $g \in \mathrm{SL}_2(K)$, one has

$$g \cdot f(X,Y) = f((X,Y)\cdot g). \tag{6}$$

There exist degree 2 and 3 polynomial invariants I and J that generate the $\mathrm{SL}_2(K)$-invariant ring as a polynomial ring. The standard normalizations of I and J are as follows:

$$I = 12ae - 3bd + c^2,$$
$$J = 72ace - 27ad^2 - 27b^2e + 9bcd - 2c^3.$$

The discriminant $\Delta(f) = \frac{1}{27}(4I^3 - J^2)$ of f is a polynomial invariant with $\mathbb{Z}$-coefficients. It is well known that if $\Delta(f)$ is nonzero, then the double cover $Z^2 = f(X,Y)$ of $\mathbb{P}^1$ is a genus one curve with Jacobian isomorphic to the elliptic curve given by

$$y^2 = x^3 - \frac{I}{3}x - \frac{J}{27}.$$

Conversely, a smooth genus one curve over K with a rational degree 2 divisor or line bundle (thereby giving a degree 2 map to $\mathbb{P}^1$) has a model of the form $Z^2 = f(X,Y)$ for a binary quartic form f over K.

Let S be a finite set of primes of K. We say that a binary quartic form (5) is *S-integral* if $a,b,c,d,e \in \mathcal{O}_{K,S}$ and *S-integer-matrix* if additionally 4 divides b and d and 6 divides c. Both conditions are preserved by the action of $\mathrm{SL}_2(\mathcal{O}_{K,S})$. For an S-integer-matrix binary quartic form f, there are polynomial invariants $I'(f), J'(f)$ with $\mathbb{Z}$-coefficients such that $12I' = I$ and $432J' = J$, so that the elliptic curve associated to f is isomorphic to the curve given by

$$y^2 = x^3 - 4I'x - 16J'. \tag{7}$$

In the sequel, we will mostly work with binary quartics of a special type, so we name them as follows:

Definition 2.1. We say a binary quartic form (5) is *demonic* if it is monic with no X^3Y-coefficient, i.e., if $a = 1$, $b = 0$, and $c,d,e \in K$.

2.2. **Mordell's construction.** In [Mor69, Chapter 25], Mordell shows that, given an integral point on an affine Weierstrass model of an elliptic curve $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$, there exists an integer-matrix binary quartic form $f(X,Y)$ and $p,q \in \mathbb{Z}$ such that $f(p,q) = 1$ and $I'(f) = -4A$ and $J'(f) = -4B$; however, his construction is not explicit. Conversely, given an integer-matrix binary quartic form $f(X,Y)$ such that I' and J' are multiples of 4 and $p,q \in \mathbb{Z}$ such that $f(p,q) = 1$, one may explicitly produce (using covariants of f) an integral point on the elliptic curve (7). In the next two subsections, we give a geometric explanation of Mordell's construction (generalized to S-integral points on elliptic curves over number fields K), which yields an explicit construction of a monic S-integer-matrix binary quartic form associated to an S-integral point on an elliptic curve.

Let S be a finite set of places of K. Let E be an elliptic curve over K with affine S-integral Weierstrass model

$$\mathcal{E}_{A,B} : y^2 = x^3 + Ax + B \tag{8}$$

with $A, B \in \mathcal{O}_{K,S}$. Let O denote the point at infinity. Given a point $P = (x_0, y_0) \in \mathcal{E}_{A,B}(K)$, the degree 2 divisor $O + P$ induces a map from E to $\mathbb{P}^1$ as a double cover ramified at a degree four subscheme of $\mathbb{P}^1$. In other words, we obtain a binary quartic form over K, which is easily computed [CFS10, BH22]:

$$f(X,Y) = X^4 - 6x_0X^2Y^2 + 8y_0XY^3 + (-4A - 3x_0^2)Y^4. \tag{9}$$

It is easy to check that $I'(f) = -4A$ and $J'(f) = -4B$.

Conversely, given a (demonic) binary quartic of the form $f(X,Y) = X^4 + 6cX^2Y^2 + 4dXY^3 + eY^4$ with $a,b,c,d,e \in \mathcal{O}_{K,S}$, we may easily solve for the coefficients of the elliptic curve and the integral point by equating the coefficients with (9). We obtain the elliptic curve

$$E : y^2 = x^3 - \frac{3c^2 + e}{4}x + \frac{c^3 + d^2 - ce}{4}$$

which contains the point $P = (x_0, y_0) = (-c, d/2)$. Assuming that $I'(f) = 3c^2 + e$ and $J'(f) = -c^3 - d^2 + ce$ are both divisible by 4, we immediately have that d must be divisible by 2, in which case the elliptic curve E and the point P both have S-integral coefficients.

It is clear that these constructions are inverse to one another. We thus obtain the explicit maps for the bijection in the following theorem, a generalization of Mordell's construction:

Theorem 2.2. *The following two sets are in bijection:*

(i) S-integral affine Weierstrass models $y^2 = x^3 + Ax + B$ of elliptic curves with S-integral points (x_0, y_0), namely with $A, B, x_0, y_0 \in \mathcal{O}_{K,S}$,

(ii) binary quartics $X^4 + 6cX^2Y^2 + 8dXY^3 + eY^4$ with $c, d, e \in \mathcal{O}_{K,S}$ and $c^2 - e$ divisible by 4.

Note that the binary quartics in Theorem 2.2 are S-integer-matrix and demonic. In the proof of Theorem 1.3, we will only need that the set of *integral* Weierstrass models with S-integral points, which is a subset of the set (i), injects into the set (ii) of these demonic S-integer-matrix binary quartics.

2.3. **Binary quartics with representations of** 1. We now relate the sets in Theorem 2.2 with binary quartic forms with representations of 1, which is Mordell's original correspondence [Mor69, Chapter 25] when $K = \mathbb{Q}$ and S contains only the infinite place. This subsection is not needed for proving the main theorems in this paper, but we include it to give a more modern interpretation of Mordell's work.

We show that S-integer-matrix binary quartic forms $f(X, Y)$ with an S-integral representation of 1 (i.e., with $p, q \in \mathcal{O}_{K,S}$ with $f(p, q) = 1$) may be transformed, under the standard action of $\mathrm{SL}_2(\mathcal{O}_{K,S})$, to demonic S-integer-matrix binary quartics. An element $g \in \mathrm{SL}_2(\mathcal{O}_{K,S})$ acts on $f(X, Y)$ by linear transformations as in (6) and on (p, q) satisfying $f(p, q) = 1$ by $(p, q) \cdot g$.

Lemma 2.3. *There is a bijection between demonic S-integer-matrix binary quartics*

$$X^4 + 6cX^2Y^2 + 4dXY^3 + eY^4 \tag{10}$$

with $c, d, e \in \mathcal{O}_{K,S}$ and $\mathrm{SL}_2(\mathcal{O}_{K,S})$-equivalence classes of triples (f, p, q), where f is an S-integer-matrix binary quartic form and $p, q \in \mathcal{O}_{K,S}$ with $f(p, q) = 1$.

Furthermore, restricting to demonic S-integer-matrix binary quartics where d is divisible by 2 and $c^2 - e$ is divisible by 4 gives a bijection with triples (f, p, q) where $I'(f)$ and $J'(f)$ are divisible by 4.

Proof. Given an S-integer-matrix binary quartic form $f(X, Y)$ and $p, q \in \mathcal{O}_{K,S}$ with $f(p, q) = 1$, because p and q must generate the unit ideal, there exist $\alpha, \beta \in \mathcal{O}_{K,S}$ with $\alpha p + \beta q = 1$. Since the action of $\left(\begin{smallmatrix} \alpha & \beta \\ -q & p \end{smallmatrix}\right)$ takes (p, q) to $(1, 0)$, there exists an $\mathrm{SL}_2(\mathcal{O}_{K,S})$-transformation taking f to a monic S-integer-matrix binary quartic form. Then "completing the quartic" (which is possible because of the coefficients of 4 and 6 for an S-integer-matrix form) shows that there exists a $\mathrm{SL}_2(\mathcal{O}_{K,S})$-transformation of f giving a binary quartic of the form (10).

Given two binary quartics f and f' of the form (10), each with the representation $(p, q) = (1, 0)$ of 1, it is straightforward to check explicitly that there is no nontrivial element of $\mathrm{SL}_2(\mathcal{O}_{K,S})$ taking $(f, 1, 0)$ to $(f', 1, 0)$.

The last statement follows trivially since for the binary quartic (10), we compute $I' = 3c^2 + e$ and $J' = -c^3 - d^2 + ce$. $\square$

Combining Lemma 2.3 with Theorem 2.2, we have the following:

Corollary 2.4. *The following sets are in bijection:*

(i) S-integral affine Weierstrass models $y^2 = x^3 + Ax + B$ of elliptic curves with integral points (x_0, y_0), namely with $A, B, x_0, y_0 \in \mathcal{O}_{K,S}$,

(ii) binary quartics $X^4 + 6cX^2Y^2 + 8dXY^3 + eY^4$ with $c, d, e \in \mathcal{O}_{K,S}$ and $c^2 - e$ divisible by 4,

(iii) $\mathrm{SL}_2(\mathcal{O}_{K,S})$-equivalence classes of triples (f, p, q), where $f(X, Y)$ is an S-integer-matrix binary quartic form with $4 \mid I'(f)$ and $4 \mid J'(f)$ and $p, q \in \mathcal{O}_{K,S}$ with $f(p, q) = 1$.

3. Counting integral points on elliptic curves

3.1. **Integral points and Selmer elements.** Let E be an elliptic curve over K with an integral affine Weierstrass model $\mathcal{E}_{A,B}$ of the form (8), and let S be a finite set of primes of K. We consider the sequence of maps

$$\Psi\colon \mathcal{E}_{A,B}(\mathcal{O}_{K,S}) \hookrightarrow E(K) \to E(K)/2E(K) \overset{\xi}{\hookrightarrow} \mathrm{Sel}_2(E/K) \tag{11}$$

where $\mathcal{E}_{A,B}(\mathcal{O}_{K,S})$ denotes the S-integral points on $E_{A,B}$ and $\mathrm{Sel}_2(E/K)$ is the 2-Selmer group of E over K.

It is well known that elements of $\mathrm{Sel}_2(E/K)$ may be represented as binary quartic forms $f(X,Y)$ over K such that the Jacobian of the associated genus one curve $C(f) : Z^2 = f(X,Y)$ is isomorphic to E and C is locally soluble. More precisely, elements of $\mathrm{Sel}_2(E/K)$ are in bijection with $\mathrm{PGL}_2(K)$-equivalence classes of such binary quartic forms (see, e.g., [BSD63, BS15a, BH16]). The $\mathrm{PGL}_2(K)$-action on binary quartic forms is induced from the following twisted action of $\mathrm{GL}_2(K)$ on binary quartics: for $g \in \mathrm{GL}_2(K)$ and a binary quartic $f(X,Y)$, we have $(g \cdot f)(X,Y) = (\det g)^{-2} f((X,Y)\cdot g)$. The ring of $\mathrm{PGL}_2(K)$-invariants is still the polynomial ring generated by I and J.

The map $\xi : E(K)/2E(K) \hookrightarrow \mathrm{Sel}_2(E/K)$ sends a rational point $P \in E(K)$ to the rational binary quartic form arising from the degree 2 map $E \to \mathbb{P}^1$ given by the divisor $O + P$ (as described in §2.2). The composition Ψ of the maps in (11) is thus given by one direction of the bijection in Theorem 2.2, from an S-integral point $P = (x_0, y_0) \in \mathcal{E}_{A,B}(\mathcal{O}_{K,S})$ to the $\mathrm{PGL}_2(K)$-equivalence class of the corresponding S-integer-matrix demonic binary quartic form $f_P(X,Y) := X^4 - 6x_0X^2Y^2 + 8y_0XY^3 + (-4A - 3x_0^2)Y^4$. Note that the genus one curve $C(f_P)$ associated to such a form (in fact, any monic binary quartic form) is automatically globally soluble over K; indeed, $f_P(1,0) = 1$ gives a rational solution. This is not surprising since, by construction, the image of P in $\mathrm{Sel}_2(E/K)$ lies in the subset of globally soluble forms, namely the image of $E(K)/2E(K)$.

Writing $E(K) \cong \mathbb{Z}^{\mathrm{rk}\, E(K)} \oplus E(K)_{\mathrm{tors}}$, we see that $|E(K)/2E(K)| \leq 4 \cdot 2^{\mathrm{rank}(E(K))}$. Hence the image of ξ, and thus the image of the composition map Ψ, is of size at most

$$\leq 4 \cdot 2^{\mathrm{rank}(E(K))}.$$

Therefore, to prove Theorem 1.3, it suffices to show that the size of each fibre of the map Ψ is bounded as follows:

Proposition 3.1. *Let K be a number field and S a finite set of places of K containing all infinite places of K. Let $f(X,Y) = X^4 + a_2X^2Y^2 + a_3XY^3 + a_4Y^4 \in \mathcal{O}_{K,S}[X,Y]$ be a demonic S-integer-matrix binary quartic form such that the discriminant $\Delta(f)$ is squarefree in $\mathcal{O}_{K,S}$. Then*

$$\#\{\gamma \in \mathrm{PGL}_2(K) \text{ such that } \gamma \cdot f \text{ is demonic and } S\text{-integer-matrix}\} \leq C^{2|S|+1}|\mathrm{Cl}(\mathcal{O}_{K,S})[2]|,$$

where $C = 7^{2^7}$.

Note that the condition that $\Delta(f)$ is squarefree can be arranged by enlarging S; this condition is chosen to coincide with the hypothesis in Theorem 1.3 that S contains all $\mathfrak{p}$ for which $\mathfrak{p}^2 \mid \Delta(f)$.

3.2. **The fiber bound.** To prove Proposition 3.1, we first establish properties of any $\gamma \in \mathrm{PGL}_2(K)$ that sends a demonic binary quartic form f to another demonic form. We then show that each such γ gives rise to a solution of a Thue-Mahler equation, and invoke the work of Evertse [Eve84] that the number of such solutions is bounded.

Lemma 3.2. *Let K be a number field and S a finite set of places of K containing all infinite places of K. Let $f(X,Y) = X^4 + a_2X^2Y^2 + a_3XY^3 + a_4Y^4 \in \mathcal{O}_{K,S}[X,Y]$ be a demonic S-integral binary quartic form. For any $\gamma \in \mathrm{PGL}_2(K)$ such that $\gamma \cdot f$ is demonic and S-integer-matrix, write $\gamma =: \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in \mathcal{O}_{K,S}$. Then we have*

(i) *$(a,b) = (a,b,c,d)$ as ideals of $\mathcal{O}_{K,S}$, and*
(ii) *$f(a,b) = (\det \gamma)^2$ divides $\Delta(f) \cdot (a,b)^4$ in $\mathcal{O}_{K,S}$.*

Lemma 3.2 follows from localizing and proving the following simpler version for principal ideal domains:

Lemma 3.3. *Let R be a principal ideal domain with field of fractions k. 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 form. Let $\gamma \in \mathrm{PGL}_2(k)$ such that $\gamma \cdot f$ is demonic, R-integer-matrix, and also in $R[X,Y]$. Write $\gamma =: \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in R$ and $(a,b,c,d) = (1)$ as ideals of R. Then we have*

(i) $(a,b) = (1)$ *as ideals of R, and*
(ii) $f(a,b) = (\det\gamma)^2$ *divides the discriminant $\Delta(f)$ of f in R.*

Here $f \in R[X,Y]$ is called *R-integer-matrix* if 4 divides the coefficients of X^3Y and XY^3 in f and 6 divides the coefficient of X^2Y^2 in f as elements of R.

Proof. Write $(a,b) =: (g)$, so there exist $\alpha, \beta \in R$ with $a = g\alpha$ and $b = g\beta$. Since $(\alpha,\beta) = (1)$, there exist $\widetilde{\alpha}, \widetilde{\beta} \in R$ 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 $\eta := c\widetilde{\alpha} - d\widetilde{\beta} \in R$. Define

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

implying 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 integer-matrix is preserved by the action of $\mathrm{SL}_2(R)$. Write

$$\widetilde{f}(X,Y) =: \widetilde{a}_0X^4 + \widetilde{a}_1X^3Y + \widetilde{a}_2X^2Y^2 + \widetilde{a}_3XY^3 + \widetilde{a}_4Y^4 \in R[X,Y].$$

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

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

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 (i).

Now since $g \in R^\times$ it follows that $\det\gamma$ divides $\widetilde{a}_1 = -4\eta\det\gamma$, 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 $\widetilde{a}_1^2$), proving (ii). □

Proof of Lemma 3.2. Given the binary quartic form $f(X,Y)$ and $\gamma \in \mathrm{PGL}_2(K)$ as in the lemma, it suffices to show the claim upon localizing. For any prime $\mathfrak{p} \notin S$ of K, we may almost directly apply Lemma 3.3 with $R := \mathcal{O}_{K,S\cup\{\mathfrak{p}\}}$ and $k := \operatorname{Frac} R$; the only difficulty is that $(a,b,c,d)_{\mathfrak{p}} := (a,b,c,d)\cdot R$ is not necessarily the unit ideal in R. However, since R is a discrete valuation ring, $(a,b,c,d)_{\mathfrak{p}}$ is a nonzero principal ideal (δ), so we need only divide a,b,c,d through by δ to apply Lemma 3.3. Then since f has degree 4, we obtain $f(a,b) = (\det\gamma)^2\delta^4$. The desired result for K follows. □

Remark 3.4. If $\Delta(f)$ is squarefree in $\mathcal{O}_{K,S}$, in the setup of Lemma 3.2, we in fact obtain $(f(a,b)) = (a,b)^4$ as ideals. Indeed, since f is a homogeneous quartic, we have $(a,b)^4 \mid (f(a,b))$, and since $(\det\gamma)^2 = (f(a,b)) \mid (\Delta(f))(a,b)^4$ and $(\Delta(f))$ is squarefree by hypothesis, we have $(f(a,b)) \mid (a,b)^4$. By unique factorization of ideals, we thus see that

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

which implies that the ideal (a,b) represents a 2-torsion class in $\mathrm{Cl}(\mathcal{O}_{K,S})$. This is the source of the factor of $|\mathrm{Cl}(\mathcal{O}_{K,S})[2]|$ in the bound of Proposition 3.1.

Proof of Proposition 3.1. The key idea is to relate the fibers of Ψ to S-integral solutions to Thue-Mahler equations; the number of such solutions is bounded by work of Bombieri, Bombieri–Schmidt, Evertse, and many others.

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

Lemma 3.2 shows that for any $\gamma \in \mathrm{PGL}_2(K)$ (represented by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a,b,c,d \in \mathcal{O}_{K,S}$) such that $\gamma \cdot f$ is demonic and S-integer-matrix, we have that $(a,b) = (a,b,c,d)$ as ideals of $\mathcal{O}_{K,S}$ and $f(a,b) = (\det\gamma)^2$ is a square dividing $\Delta(f) \cdot (a,b,c,d)^4$ in $\mathcal{O}_{K,S}$. Further, by Remark 3.4, since $\Delta(f)$ is squarefree in $\mathcal{O}_{K,S}$, we have that $(a,b,c,d)^2 = (a,b)^2 = (\det\gamma)$.

Therefore, by the definition of P, there is a prime ideal $\mathfrak{p} \in P$ and an $\alpha \in K^\times$ for which $\mathfrak{p} = \alpha \cdot (a,b,c,d)$. Since $\mathfrak{p} \subseteq \mathcal{O}_K \subseteq \mathcal{O}_{K,S}$ it follows that $\alpha a, \alpha b, \alpha c, \alpha d \in \mathcal{O}_{K,S}$. Scaling each of a,b,c,d by α (which 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 $\mathcal{O}_{K,S}$.

Thus, given $\gamma \in \mathrm{PGL}_2(K)$ for which $\gamma \cdot f$ is both demonic and S-integer-matrix, we get a pair $(a,b) \in \mathcal{O}_{K,S}^2$, well defined up to the action of $\mathcal{O}_{K,S}^\times$ (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

$$\begin{aligned} \Phi\colon \{\gamma \in \mathrm{PGL}_2(K) : \gamma \cdot f \text{ is demonic and } S\text{-integer-matrix}\} & \\ \to \bigcup_{\mathfrak{p}\in P} \{(a,b) \in \mathcal{O}_{K,S}^2 \mid (a,b) = \mathfrak{p}, (f(a,b)) = \mathfrak{p}^4\}/\mathcal{O}_{K,S}^\times, & \end{aligned} \tag{12}$$

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 $(a,b) \in \mathcal{O}_{K,S}^2/\mathcal{O}_{K,S}^\times$, 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. We now bound the size of the codomain. Write

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

Thus the codomain is $\bigcup_{\mathfrak{p}\in P} M_{\mathfrak{p}}/\mathcal{O}_{K,S}^\times$, and we will compute an upper bound for each term $M_{\mathfrak{p}}/\mathcal{O}_{K,S}^\times$.

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

Next we enlarge $M_\mathfrak{p}$ as follows. Since $(f(a,b)) = \mathfrak{p}^4$ implies that $f(a,b) \in \mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^\times$, we observe that

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

and hence

$$M_\mathfrak{p}/\mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^\times \subseteq \{(a,b) \in \mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^2 : f(a,b) \in \mathcal{O}_{K,S\cup\{p\}}^\times\}/\mathcal{O}_{K,S\cup\{p\}}^\times.$$

Now we use the following theorem of Evertse [Eve84, Theorem 3]: given a number field K with a finite set of places Ξ containing all infinite places and a homogeneous polynomial $F \in \mathcal{O}_{K,\Xi}[x,y]$ with at least three distinct roots in $\mathbb{P}^1(\overline{\mathbb{Q}})$, one has

$$\left|\{(a,b) \in \mathcal{O}_{K,\Xi}^2 : F(a,b) \in \mathcal{O}_{K,\Xi}^\times\}/\mathcal{O}_{K,\Xi}^\times\right| \leq 7^{(\deg F)^3([K:\mathbb{Q}]+2|\Xi|)}.$$

Applying this theorem here with $F = f$ and $\Xi = S \cup \{\mathfrak{p}\}$, and using $[K:\mathbb{Q}] \leq 2|S|$, we find

$$\begin{aligned}\left|M_\mathfrak{p}/\mathcal{O}_{K,S}^\times\right| &= \left|M_\mathfrak{p}/\mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^\times\right| \\ &\leq \left|\{(a,b) \in \mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^2 : f(a,b) \in \mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^\times\}/\mathcal{O}_{K,S\cup\{\mathfrak{p}\}}^\times\right| \\ &\leq C^{2|S|+1},\end{aligned}$$

where $C = 7^{2^7}$.

To bound the codomain of Φ, we sum this over $\mathfrak{p} \in P$ and use $|P| = |\mathrm{Cl}(\mathcal{O}_{K,S})[2]|$ to obtain

$$|\{\gamma \in \mathrm{PGL}_2(K) : \gamma \cdot f \text{ is demonic and } S\text{-integer-matrix}\}| \leq C^{2|S|+1}|\mathrm{Cl}(\mathcal{O}_{K,S})[2]|. \qquad \square$$

When $K = \mathbb{Q}$, the fiber bound may be improved as follows by combining arguments of Bombieri–Schmidt [BS87] with Evertse's bounds [Eve84] used in the proof of Proposition 3.1:

Proposition 3.5. *Let $f(X,Y) = X^4 + a_2X^2Y^2 + a_3XY^3 + a_4Y^4 \in \mathbb{Z}[X,Y]$ be a demonic binary quartic form. The number of elements $\gamma \in \mathrm{PGL}_2(\mathbb{Q})$ such that $\gamma \cdot f$ is demonic is*

$$\ll \prod_{p^2|\Delta(f)} \min\left\{4\left\lfloor\frac{v_p(\Delta(f))}{2}\right\rfloor + 1,\, 7^{2^7}\right\}.$$

Proof. By the same argument as in the proof of Proposition 3.1, here using just Lemma 3.3 with $R = \mathbb{Z}$, we reduce to bounding

$$\sum_{\delta^2|\Delta(f)} \#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f(a,b) = \delta^2\}. \tag{13}$$

We divide the set of primes $p^2 \mid \Delta(f)$ into two sets to obtain a hybrid bound. Let T be the set[4] of primes such that $C = 7^{2^7} \leq 4\left\lfloor\frac{v_p(\Delta(f))}{2}\right\rfloor + 1$, and let

$$D := \prod_{\substack{p^2|\Delta(f)\\ p\notin T}} p^{v_p(\Delta(f))}.$$

Given δ such that $\delta^2 \mid \Delta(f)$, set $\nu := \gcd(\delta, D)$ and $\mu := \frac{\delta}{\nu}$.

The argument of Bombieri–Schmidt in [BS87, Section VI], specifically Lemma 7 and the second-to-last paragraph, produces $\leq 4^{\omega(\nu)}$ many quartic forms $f_{\nu,i}$, depending only on f and ν, such that

[4]Taking T to be the empty set in this argument gives a weaker but simpler upper bound for (13), and thus for Proposition 3.1, of $\prod_{p^2|\Delta(f)}\left(4\left\lfloor\frac{v_p(\Delta(f))}{2}\right\rfloor + 1\right)$.

the number of relatively prime solutions to $f(a,b) = \delta^2 = \nu^2\mu^2$ is bounded above by the sum of the numbers of relatively prime solutions of $f_{\nu,i}(a,b) = \mu^2$. Rewriting (13) in terms of μ and ν gives

$$\begin{aligned}
&\sum_{\delta^2 \mid \Delta(f)} \#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f(a,b) = \delta^2\} \\
&\qquad = \sum_{\nu^2 \mid D} \sum_{\mu^2 \mid \frac{\Delta(f)}{D}} \#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f(a,b) = \mu^2\nu^2\} \\
&\qquad \leq \sum_{\nu^2 \mid D} \sum_{i=1}^{4^{\omega(\nu)}} \sum_{\mu^2 \mid \frac{\Delta(f)}{D}} \#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f_{\nu,i}(a,b) = \mu^2\}. \qquad (14)
\end{aligned}$$

To bound the number of solutions to $f_{\nu,i}(a,b) = \mu^2$, we use Evertse's theorem with $K = \mathbb{Q}$, $\Xi = T \cup \{\infty\}$, and $F = f_{\nu,i}$. The innermost sum of (14) is at most

$$\#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f_{\nu,i}(a,b) \in \mathbb{Z}[T^{-1}]^\times = \mathcal{O}_{\mathbb{Q},T}^\times\} \leq C^{|T|+3/2}.$$

Hence the size (13) of the codomain of Φ is

$$\begin{aligned}
\sum_{\delta^2 \mid \Delta(f)} \#\{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1, f(a,b) = \delta^2\} &\ll \sum_{\nu^2 \mid D} \sum_{i=1}^{4^{\omega(\nu)}} C^{|T|} \\
&= C^{|T|} \sum_{\nu^2 \mid D} 4^{\omega(\nu)} \\
&= C^{|T|} \prod_{p^2 \mid D} \left(4 \left\lfloor \frac{v_p(\Delta(f))}{2} \right\rfloor + 1\right). \qquad \square
\end{aligned}$$

Remark 3.6. When $2 \leq v_p(\Delta(f)) < 4$, evidently $p \notin S$, and either $p \nmid \nu$ (in which case there is no factor corresponding to p) or $v_p(\nu) = 2$. By simply enumerating cases of f over $\mathbb{Q}_p$ one finds that the number of disks required for [BS87, Lemma 7] is in fact at most 3, because at least two roots lie in the same residue disk modulo p. This translates into a factor of 3 corresponding to p, rather than 4, and thus gives the claim of Remark 1.2 (after applying this improved bound in the following proof of Theorem 1.1).

3.3. **Bounds on the number of integral points on an elliptic curve.** Combining Proposition 3.1 with the bound on the image of Ψ gives Theorem 1.3 immediately.

Proof of Theorem 1.3. We showed that the map $\Psi\colon \mathcal{E}_{A,B}(\mathcal{O}_{K,S}) \to E(K)/2E(K) \subseteq \mathrm{Sel}_2(E/K)$, taking an integral point of $\mathcal{E}_{A,B}$ to the $\mathrm{PGL}_2(K)$-equivalence class of its corresponding binary quartic form f (by Theorem 2.2), has image of size $\ll 2^{\mathrm{rk}\, E(K)}$. Recall that the binary quartic f has invariants $I = -48A$ and $J = -1728B$, so $\Delta(f) = 2^8\Delta_{A,B}$. Applying Proposition 3.1 bounds the size of a fibre of the map Ψ, and combining the two estimates gives the theorem. $\square$

Theorem 1.1 follows from Proposition 3.5 in exactly the same way.

Remark 3.7. There have been significant improvements to bounds on the number of solutions to Thue-Mahler equations since [Eve84], but the more recent work (e.g., [BS87, Eve97, AO10, Akh12]) usually requires a hypothesis of irreducibility of the Thue-Mahler form that does not always apply for the binary quartics in Propositions 3.1 and 3.5. A long case-by-case analysis would make it possible to invoke these bounds here, but they would only improve the constant C in Theorems 1.3 and 1.6 and not change the shape of the bounds.

However, it is possible to use these better Thue-Mahler bounds *on average* to improve the bounds on the moments in Section 4, since most of the reducible binary quartics are in a negligible set. For

$K = \mathbb{Q}$, we can use even more: once $\Delta(f) \gg 1$ and f is irreducible, Akhtari [Akh12] gives a much better bound of 26 for the number of solutions to $f(x,y) = 1$. We do not work out the optimal constants in Section 4, since our goal is simply to show that the ϑ-th moments are bounded.

4. Bounding moments of the number of integral points on elliptic curves

Theorem 1.6 follows from "averaging" the bound in Theorem 1.1 and analytic techniques. The additional crucial input is Bhargava–Shankar's result that the average size of the 5-Selmer group of elliptic curves in $\mathscr{F}_{\text{univ}}$, ordered by height, is bounded (proved over $\mathbb{Q}$ in [BS13b, Theorem 31] and over global fields in [BSW22]):

Theorem 4.1 ([BSW22]). *Let K be a number field. Then the average size of the 5-Selmer group of elliptic curves in $\mathscr{F}_{\text{univ}}(\mathcal{O}_K)$, ordered by height, is*

$$\operatorname{Avg}_{\mathcal{E}_{A,B} \in \mathscr{F}_{\text{univ}}(\mathcal{O}_K)} |\operatorname{Sel}_5(\mathcal{E}_{A,B})| = 6.$$

Remark 4.2. As mentioned in the introduction, the family $\mathscr{F}_{\text{univ}}(\mathcal{O}_K)$ is ordered by the height defined by (3), in both Theorem 4.1 and in our averaging results. A slightly different height $\widetilde{H}$, which coincides with the usual height on weighted projective space $\mathbb{P}(4,6)$, may also be used, as [BSW22] also proves that the average 5-Selmer size is 6 for short Weierstrass elliptic curves E over K ordered by $\widetilde{H}$. Since $\widetilde{H}$ is invariant under scaling, i.e., all integral models of a given elliptic curve E over K have the same value of $\widetilde{H}$, we must fix a quasi-minimal integral model for each curve in order to count integral points. Our results and proofs below may be modified appropriately to show that the ϑ-th moment of the number of S-integral points on (quasi-minimal integral models of) elliptic curves over K, ordered by $\widetilde{H}$, is bounded, for $0 < \vartheta \le \log_2 5$.

4.1. Moments of the number of integral points in families of elliptic curves.

We now prove the following slightly weaker version of Theorem 1.6; the remaining case where $\vartheta = \log_2 5$ will be handled in Section 5.

Theorem 4.3. *Let $\mathscr{F} \subseteq \mathscr{F}_{\text{univ}}(\mathcal{O}_K)$ be a subset of positive lower density (ordering by height). Let S be a finite set of places of K. Let $0 < \vartheta < \log_2 5 = 2.3219\ldots$. Then*

$$\operatorname{Avg}_{\mathscr{F}}(|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}) \ll_{\vartheta,\mathscr{F}} \left(C^{2|S|}\,|\operatorname{Cl}(\mathcal{O}_{K,S})[2]|\right)^{\vartheta}$$

where the average is taken over all elliptic curves $\mathcal{E}_{A,B} \in \mathscr{F}$ ordered by height and $C = 7^{2^7}$.

Proof. We may immediately reduce to the case of $\mathscr{F} = \mathscr{F}_{\text{univ}}(\mathcal{O}_K)$ because

$$\sum_{\mathcal{E}_{A,B} \in \mathscr{F}^{\le T}} |\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta} \le \sum_{\mathcal{E}_{A,B} \in \mathscr{F}_{\text{univ}}^{\le T}(\mathcal{O}_K)} |\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}$$

and, by the positive lower density hypothesis,

$$\sum_{\mathcal{E}_{A,B} \in \mathscr{F}^{\le T}} 1 \gg_{\mathscr{F}} \sum_{\mathcal{E}_{A,B} \in \mathscr{F}_{\text{univ}}^{\le T}(\mathcal{O}_K)} 1.$$

Thus, we may now assume that $\mathscr{F} = \mathscr{F}_{\text{univ}}(\mathcal{O}_K)$.

Upon applying Theorem 1.3 we find that

$$\begin{aligned}&\operatorname{Avg}_{\mathscr{F}_{\text{univ}}(\mathcal{O}_K)}\left(|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}\right)\\ &\qquad \ll C^{(2|S|+1)\vartheta}|\operatorname{Cl}(\mathcal{O}_{K,S})[2]|^{\vartheta}\operatorname{Avg}_{\mathscr{F}_{\text{univ}}(\mathcal{O}_K)}\left((2^{\vartheta})^{\operatorname{rank}E(K)}O(1)^{\omega_{\ge 2}(\Delta_{A,B})}\right),\end{aligned}$$

where $C = 7^{2^7}$ as before. Thus it suffices to show that

$$\operatorname{Avg}_{\mathscr{F}_{\text{univ}}(\mathcal{O}_K)}\left((2^{\vartheta})^{\operatorname{rank}E(K)}O(1)^{\omega_{\ge 2}(\Delta_{A,B})}\right) \ll O(1)^{(\log_2 5 - \vartheta)^{-1}}.$$

Set $\varepsilon := \frac{\log_2 5}{\vartheta} - 1 > 0$. By Hölder's inequality with dual exponent pair $(1+\varepsilon, 1+\varepsilon^{-1})$, we obtain

$$\begin{aligned}
&\mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)} \left((2^{\vartheta})^{\operatorname{rank} E(K)} O(1)^{\omega_{\geq 2}(\Delta_{A,B})} \right) \\
&\qquad \ll \mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)} \left((2^{\vartheta(1+\varepsilon)})^{\operatorname{rank} E(K)} \right)^{\frac{1}{1+\varepsilon}} \mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)} \left(O(1)^{\omega_{\geq 2}(\Delta_{A,B})(1+\varepsilon^{-1})} \right)^{\frac{\varepsilon}{1+\varepsilon}} \qquad (15) \\
&\qquad \ll \mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)} \left(5^{\operatorname{rank} E(K)} \right) \mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)} \left(\left(O(1)^{(\log_2 5 - \vartheta)^{-1}} \right)^{\omega_{\geq 2}(\Delta_{A,B})} \right)^{\varepsilon}. \qquad (16)
\end{aligned}$$

We apply Theorem 4.1 to bound the first average in (16) by $O(1)$. To bound the second average, we would like to apply Lemma 4.5 below with the discriminant polynomial $\Delta(A,B)$ on elliptic curves $\mathcal{E}_{A,B} \in \mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ with $H(A,B) \leq X$, $\eta = 1$, and $\lambda = X$. (Note that the weights $\vec{\alpha}$ are irrelevant in this case.) To apply Lemma 4.5, we only need to verify its condition (b), namely that for a squarefree ideal $\mathfrak{d} \subseteq \mathcal{O}_K$ with $\mathrm{Nm}\,\mathfrak{d} \leq X^{\delta}$,

$$\operatorname*{Prob}_{\substack{\mathcal{E}_{A,B} \in \mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K) \\ H(A,B) \leq X}} \left(\mathfrak{d}^2 \mid \Delta_{A,B} \right) \ll O(1)^{\#\{\mathfrak{p} \mid \mathfrak{d}\}} (\mathrm{Nm}\,\mathfrak{d})^{-2},$$

which we now check.

First, the number of solutions $(x,y) \in \mathcal{O}_K/\mathfrak{d}^2$ of $-16(4x^3 + 27y^2) \equiv 0 \pmod{\mathfrak{d}^2}$ is

$$\ll \mathrm{Nm}(\mathfrak{d})^2 O(1)^{\#\{\mathfrak{p} \mid \mathfrak{d}\}}. \qquad (17)$$

Indeed, by the Chinese remainder theorem, it suffices to check (17) when $\mathfrak{d}$ is prime, and then upon fixing $x \in (\mathcal{O}_K/\mathfrak{d})^{\times}$, we find a quadratic in y, whose number of solutions is bounded by $O(1)^{\#\{\mathfrak{p} \mid \mathfrak{d}\}}$ by Hensel lifting.

Also, the set $\{(A,B) \in \mathcal{O}_K^2 : H(A,B) \leq X\}$ covers each congruence class modulo $\mathfrak{d}^2$ roughly evenly, since given $(x,y) \in \mathcal{O}_K/\mathfrak{d}^2$, we see that

$$\begin{aligned}
&\# \left\{ (A,B) \in \mathcal{O}_K^2 : H(A,B) \leq X \text{ and } (A,B) \equiv (x,y) \pmod{\mathfrak{d}^2} \right\} \\
&\qquad \ll \mathrm{Nm}(\mathfrak{d})^{-4} \# \left\{ (A,B) \in \mathcal{O}_K^2 : H(A,B) \leq X \right\}. \qquad (18)
\end{aligned}$$

Combining these two observations yields

$$\begin{aligned}
&\# \left\{ (A,B) \in \mathcal{O}_K^2 : H(A,B) \leq X, \Delta_{A,B} \equiv 0 \pmod{\mathfrak{d}^2} \right\} \qquad (19) \\
&\qquad \ll \mathrm{Nm}(\mathfrak{d})^{-2} O(1)^{\#\{\mathfrak{p} \mid \mathfrak{d}\}} \# \left\{ (A,B) \in \mathcal{O}_K^2 : H(A,B) \leq X \right\}. \qquad \square
\end{aligned}$$

We now prove the lemmas that will give an upper bound for the second average in (16), namely the average of $O(1)^{\omega_{\geq 2}(\Delta)}$ as Δ varies in a reasonable way. The first lemma shows that a small-norm divisor of the discriminant ideal can be used as a proxy for the discriminant itself when averaging.

Lemma 4.4. *Let δ, c, X be positive real numbers. Let $\mathfrak{z} \subseteq \mathcal{O}_K$ be an ideal with $\mathrm{Nm}(\mathfrak{z}) \leq X^c$. Then there exists a squarefree ideal $\mathfrak{d}$ such that $\mathfrak{d}^2 \mid \mathfrak{z}$ with*

(i) *$\mathrm{Nm}(\mathfrak{d}) \leq X^{\delta}$, and*
(ii) *$\#\{\mathfrak{p}^2 \mid \mathfrak{z}\} \leq (2 + \delta^{-1}c)\#\{\mathfrak{p}^2 \mid \mathfrak{d}\}$.*

Proof. Note that $\#\{\mathfrak{p}^2 \mid \mathfrak{z} : \mathrm{Nm}(\mathfrak{p}) > X^{\delta/2}\} \ll \delta^{-1}c$. Let

$$\mathfrak{n} = \prod_{\substack{\mathfrak{p}^2 \mid \mathfrak{z} \\ \mathrm{Nm}\,\mathfrak{p} \leq X^{\delta/2}}} \mathfrak{p},$$

so $\mathrm{Nm}(\mathfrak{n}) \leq X^{c/2}$ since $\mathfrak{n}^2 \mid \mathfrak{z}$. Thus if $\mathrm{Nm}(\mathfrak{n}) \leq X^{\delta}$, then taking $\mathfrak{d} = \mathfrak{n}$ suffices.

We may now assume that $\mathrm{Nm}(\mathfrak{n}) > X^\delta$. First, for every divisor $\mathfrak{n}'$ of $\mathfrak{n}$, we claim there exists an ideal $\mathfrak{d}' \mid \mathfrak{n}'$ for which

$$\min(X^{\delta/2}, \mathrm{Nm}(\mathfrak{n}')) \leq \mathrm{Nm}(\mathfrak{d}') \leq X^\delta. \tag{20}$$

This follows from the fact that $\mathrm{Nm}(\mathfrak{p}) \leq X^{\delta/2}$ for each $\mathfrak{p} \mid \mathfrak{n}$ by construction. We simply start with $\mathfrak{n}$ and remove one prime $\mathfrak{p}$ at a time until (20) is satisfied. We call this the "deletion argument".

An analogous argument shows that for every divisor $\mathfrak{n}'$ of $\mathfrak{n}$ for which $\mathrm{Nm}(\mathfrak{n}') \leq X^{\delta/2}$, there is an ideal $\mathfrak{d}'$ such that $\mathfrak{n}' \mid \mathfrak{d}' \mid \mathfrak{n}$ and $X^{\delta/2} \leq \mathrm{Nm}(\mathfrak{d}') \leq X^\delta$. Here, we instead add one prime $\mathfrak{p}$ at a time (which we can do because $\mathrm{Nm}\,\mathfrak{n} > X^\delta$) until the norm condition is satisfied; call this the "insertion argument".

Let $\mathfrak{d}$ be a divisor of $\mathfrak{n}$ for which $X^{\delta/2} \leq \mathrm{Nm}(\mathfrak{d}) \leq X^\delta$ (and thus satisfying (i)), for which $\#\{\mathfrak{p} \mid \mathfrak{d}\}$ is maximized subject to this condition, namely

$$\mathfrak{d} := \operatorname{argmax}\left\{\#\{\mathfrak{p} \mid \mathfrak{d}'\} : X^{\delta/2} \leq \mathrm{Nm}(\mathfrak{d}') \leq X^\delta\right\}.$$

Note that we are maximizing over a nonempty finite set (by the deletion argument).

Let $\mathfrak{a}$ be the ideal such that $\mathfrak{n} = \mathfrak{a}\mathfrak{d}$. Because $\mathrm{Nm}(\mathfrak{n}) > X^\delta$ by assumption, we have $\mathrm{Nm}(\mathfrak{a}) > 1$. By repeatedly applying the deletion argument to $\mathfrak{a}$ as needed, we may decompose

$$\mathfrak{n} = \mathfrak{a}_1 \cdots \mathfrak{a}_N \mathfrak{d} \tag{21}$$

with $X^{\delta/2} \leq \mathrm{Nm}(\mathfrak{a}_i) \leq X^\delta$ for all $i < N$ and $\mathrm{Nm}(\mathfrak{a}_N) < X^{\delta/2}$; note that $N \geq 1$ since $\mathfrak{a} \neq (1)$. Since

$$X^{\delta(N-1)/2} \leq \mathrm{Nm}(\mathfrak{n}) \leq X^{c/2},$$

we conclude that $1 \leq N \leq 1 + \delta^{-1}c$. Moreover, our choice of $\mathfrak{d}$ gives

$$\#\{\mathfrak{p} \mid \mathfrak{d}\} \geq \#\{\mathfrak{p} \mid \mathfrak{a}_i\}$$

for *all* $1 \leq i \leq N$: this follows for $i < N$ from the maximality imposed in the definition of $\mathfrak{d}$, and for $i = N$ from this maximality and the insertion argument. Hence the decomposition (21) gives

$$\#\{\mathfrak{p} \mid \mathfrak{n}\} \leq (N+1)\#\{\mathfrak{p} \mid \mathfrak{d}\},$$

so (ii) is satisfied as well, which proves the lemma. ☐

We now establish a general lemma that is used both in Theorem 4.3 and in Section 5. Given a polynomial $\Delta(v_1, \ldots, v_n)$ whose behavior on a bounded set of $\vec{v}$ is somewhat controlled, we obtain an upper bound for the average size of a constant raised to the number of primes whose squares divide $\Delta(\vec{v})$, as $\vec{v}$ ranges over that set.

Lemma 4.5. *Let K be a number field. Let λ, t, η, C_1, C_2, and C_3 be positive real numbers, and suppose $\lambda \geq 1$. Let N be a positive integer, and fix $\vec{\alpha} \in (\mathbb{R}^+)^N$ giving an action of $\mathbb{R}^+$ on $\vec{u} \in (K \otimes_{\mathbb{Q}} \mathbb{R})^N$ by $\lambda \cdot \vec{u} := (\lambda^{\alpha_i} u_i)_i$. Fix a bounded open subset $\mathcal{S} \subseteq (K \otimes_{\mathbb{Q}} \mathbb{R})^N$.*

Suppose $\Delta \in \mathcal{O}_K[X_1, \ldots, X_N]$ satisfies

(a) *$\Delta(\lambda \cdot \vec{u}) \leq C_1 \lambda^{C_2}$ for all $\vec{u} \in \mathcal{S}$, and*
(b) *for all $\delta > 0$ with $\delta \ll_{\eta,\vec{\alpha}} 1$ and squarefree ideals $\mathfrak{a} \subseteq \mathcal{O}_K$ with $\mathrm{Nm}(\mathfrak{a}) \leq \lambda^\delta$,*

$$\operatorname*{Prob}_{\vec{v} \in \mathcal{O}_K^N \cap \lambda \cdot \mathcal{S}} \left(\mathfrak{a}^2 \mid \Delta(\vec{v})\right) \leq C_3^{\#\{\mathfrak{p} \mid \mathfrak{a}\}} (\mathrm{Nm}(\mathfrak{a}))^{-1-\eta}. \tag{22}$$

Then

$$\operatorname*{Avg}_{\vec{v} \in \mathcal{O}_K^N \cap \lambda \cdot \mathcal{S}} \left(t^{\#\{\mathfrak{p}^2 \mid \Delta(\vec{v})\}}\right) \leq t^{O(1)} \zeta_K(1+\eta)^{C_3 t^{O(1)}}$$

where both $O(1)$'s depend on η, $\vec{\alpha}$, C_1, and C_2.

Proof. Fix $\delta \asymp_{\eta,\vec{\alpha}} 1$. For each $\vec{v} \in \mathcal{O}_K^N \cap \lambda \cdot \mathcal{S}$, we apply Lemma 4.4 with $\mathfrak{z} = \Delta(\vec{v})$ to find a squarefree ideal $\mathfrak{d}(\vec{v})$ such that $\mathfrak{d}(\vec{v})^2 \mid (\Delta(\vec{v}))$ with

(i) $\mathrm{Nm}\,\mathfrak{d}(\vec{v}) \leq \lambda^\delta$, and

(ii) $\#\{\mathfrak{p}^2 \mid \Delta(\vec{v})\} \ll_{C_1,C_2} \delta^{-1}(\#\{\mathfrak{p} \mid \mathfrak{d}(\vec{v})\} + 1)$.

In the following computation, every instance of $O(\cdot)$ is dependent on C_1 and C_2, which we omit for notational convenience:

$$\begin{aligned}
\sum_{\vec{v}\in\mathcal{O}_K^N\cap\lambda\cdot\mathcal{S}} t^{\#\{\mathfrak{p}^2|\Delta(\vec{v})\}} &\leq t^{O(\delta^{-1})} \sum_{\substack{\mathfrak{d}\text{ squarefree}\\ \mathrm{Nm}\,\mathfrak{d}\leq\lambda^\delta}} t^{O(\delta^{-1}\#\{\mathfrak{p}|\mathfrak{d}\})} \sum_{\substack{\vec{v}\in\mathcal{O}_K^N\cap\lambda\cdot\mathcal{S}\\ \mathfrak{d}^2|\Delta(\vec{v})}} 1 && \text{by (ii)}\\
&\leq t^{O(\delta^{-1})}\left(\sum_{\mathrm{Nm}\,\mathfrak{d}\leq\lambda^\delta} \frac{(C_3 t^{O(\delta^{-1})})^{\#\{\mathfrak{p}|\mathfrak{d}\}}}{(\mathrm{Nm}\,\mathfrak{d})^{1+\eta}}\right)\left(\sum_{\vec{v}\in\mathcal{O}_K^N\cap\lambda\cdot\mathcal{S}} 1\right) && \text{by (b)}\\
&\leq t^{O(\delta^{-1})}\left(\sum_{\mathfrak{d}\subseteq\mathcal{O}_K} \frac{(C_3 t^{O(\delta^{-1})})^{\#\{\mathfrak{p}|\mathfrak{d}\}}}{(\mathrm{Nm}\,\mathfrak{d})^{1+\eta}}\right)\left(\sum_{\vec{v}\in\mathcal{O}_K^N\cap\lambda\cdot\mathcal{S}} 1\right)\\
&\leq t^{O(\delta^{-1})}\zeta_K(1+\eta)^{C_3 t^{O(\delta^{-1})}} \sum_{\vec{v}\in\mathcal{O}_K^N\cap\lambda\cdot\mathcal{S}} 1,
\end{aligned}$$

where the last inequality uses the fact that $(1+a)^b \geq 1+ab$ for $a,b>0$. □

4.2. **Families of elliptic curves with marked points.** The arguments in Section 4.1 may be modified appropriately to give averages or moments on the number of integral points on elliptic curves over $\mathbb{Q}$ in some other families for which we have finite upper bounds on the average d-Selmer group size for some $d \geq 2$. These families include the families

$$\mathscr{F}_1 = \{y^2 + d_3 y = x^3 + d_2 x^2 + d_4 x \mid d_2, d_3, d_4 \in \mathbb{Z},\ \Delta \neq 0\} \qquad \text{and}$$

$$\mathscr{F}_2 = \{y^2 + d_1 xy + d_3 y = (x-d_2)(x-d_2')(x-d_2'') \mid d_1, d_2, d_2', d_2'', d_3 \in \mathbb{Z},\ d_2 + d_2' + d_2'' = 0,\ \Delta \neq 0\}$$

of elliptic curves over $\mathbb{Q}$. The family $\mathscr{F}_1$ has a marked point at $(0,0)$, and the family $\mathscr{F}_2$ has two (usually independent) marked points. The height $H(\mathcal{E})$ of a curve $\mathcal{E}$ in these families is again a measure of the size of the coefficients, defined as $\max_i \left\{|d_i|^{\frac{12}{i}}\right\}$ (for $\mathscr{F}_2$, we include $|d_2'|^6$ and $|d_2''|^6$ in that maximum). By [BH22, Theorem 1.1], the average size of the 3-Selmer group (resp., 2-Selmer group) in $\mathscr{F}_1$ (resp., $\mathscr{F}_2$), ordered by height, is bounded. We claim that the average number of integral points on the curves in these families is bounded, and in fact, a stronger statement holds:

Theorem 4.6. *For any positive lower density family $\mathscr{F}$ in $\mathscr{F}_1$ (respectively, $\mathscr{F}_2$) and any positive real number $\vartheta \leq \log_2 3 = 1.5850\ldots$ (resp., $\vartheta \leq 1$), we have*

$$\mathrm{Avg}_{\mathscr{F}}(|\mathcal{E}(\mathbb{Z})|^\vartheta) \ll_{\vartheta,\mathscr{F}} 1$$

where the average is taken over all $\mathcal{E} \in \mathscr{F}$ ordered by height.

Proof. We first explain the case of $\mathscr{F}_1$. For $\vartheta < \log_2 3$, the proof follows the same outline as that of Theorem 4.3. As before, we reduce to the case of $\mathscr{F} = \mathscr{F}_1$. Let $\mathscr{F}^{\leq T} := \{\mathcal{E} \in \mathscr{F} : \Delta_{\mathcal{E}} \neq 0 \text{ and } H(\mathcal{E}) \leq T\}$ represent the curves in $\mathscr{F}$ of height at most T. Given an integral model $\mathcal{E} \in \mathscr{F}$, let E over $\mathbb{Q}$ be the corresponding elliptic curve. The bound of Theorem 1.1 and Hölder's inequality give an inequality analogous to (15):

$$\sum_{\mathcal{E}\in\mathscr{F}^{\leq T}} |\mathcal{E}(\mathbb{Z})|^\vartheta \ll \left(\sum_{\mathcal{E}\in\mathscr{F}^{\leq T}} (2^{\vartheta(1+\varepsilon)})^{\mathrm{rk}\,E(\mathbb{Q})}\right)^{\frac{1}{1+\varepsilon}} \left(\sum_{\mathcal{E}\in\mathscr{F}^{\leq T}} O(1)^{\omega_{\geq 2}(\Delta_{\mathcal{E}})(1+\varepsilon^{-1})}\right)^{\frac{\varepsilon}{1+\varepsilon}} \tag{23}$$

The first term is bounded as before, by choosing $0 < \varepsilon < \frac{\log_2 3}{s} - 1$ so that

$$\left(2^{\vartheta(1+\varepsilon)}\right)^{\mathrm{rk}\,E(\mathbb{Q})} \leq 3^{\mathrm{rk}\,E(\mathbb{Q})} \leq |E(\mathbb{Q})/3E(\mathbb{Q})| \leq |\mathrm{Sel}_3(E)|\,.$$

The bounds on the average 3-Selmer size from [BH22] imply that

$$\sum_{\mathcal{E}\in\mathscr{F}^{\leq T}} \left(2^{\vartheta(1+\varepsilon)}\right)^{\operatorname{rk} E(\mathbb{Q})} \ll \left|\mathscr{F}^{\leq T}\right|.$$

To bound the second term in (23), we again note that, by Lemma 4.5, it suffices to check that

$$\operatorname*{Prob}_{\mathcal{E}\in\mathscr{F}^{\leq T}} \left(m^2 \mid \Delta_{\mathcal{E}}\right) \ll O(1)^{\omega(m)} m^{-2} \tag{24}$$

when $m < T^{\delta}$ is squarefree.

But for $\mathscr{F} = \mathscr{F}_1$ and $m < T^{\delta}$, each fiber of the natural reduction map $\mathscr{F}^{\leq T} \to (\mathbb{Z}/m^2\mathbb{Z})^3$ sending $E \in \mathscr{F}_1$ to (d_2, d_3, d_4) modulo m^2 is of size

$$\ll \left(\frac{T^{1/6}}{m^2}+1\right)\left(\frac{T^{1/4}}{m^2}+1\right)\left(\frac{T^{1/3}}{m^2}+1\right) \ll \left|\mathscr{F}^{\leq T}\right| m^{-6}.$$

Also, there are $\ll m^4 O(1)^{\omega(m)}$ solutions (d_2, d_3, d_4) modulo m^2 to the discriminant vanishing modulo m^2. Indeed, by the Chinese remainder theorem, it suffices to verify this for $m = p$ prime. In that case, by Hensel's lemma, for each of the $\ll p^2$ many mod p solutions with a nonvanishing differential, there are p^2 lifts to a mod p^2 solution, whereas there are $\ll p$ many mod p solutions with a vanishing differential (and trivially $\leq p^3$ lifts of each to a mod p^2 solution). We thus obtain the bound (24).

The argument for $\mathscr{F} = \mathscr{F}_2$ is entirely analogous, using the upper bound on the average size of the 2-Selmer group for curves in $\mathscr{F}_2$ from [BH22]. The equality cases $(\mathscr{F}, \vartheta) = (\mathscr{F}_1, \log_2 3)$ or $(\mathscr{F}_2, 1)$ are proven in Theorem 5.4 below. □

Remark 4.7. While there are other families of elliptic curves with known average d-Selmer upper bounds, in many cases (such as the family of Mordell curves and families with marked 2- or 3-torsion points), the discriminant often—or even always!—has square factors. We thus cannot use these same averaging techniques to prove analogous moment bounds for these families.

5. Arithmetic statistics with weights

We introduce a method to leverage the "usual" geometry-of-numbers techniques for proving bounds on average in arithmetic statistics to obtain bounds for *weighted* averages, provided the relevant weight function is sufficiently well behaved. Rather than formulate a general theorem, we apply the method to bound the weighted average of the number of d-Selmer elements of elliptic curves $\mathcal{E} \in \mathscr{F}$ ordered by height, with each d-Selmer element weighted by $O(1)^{\omega_{\geq 2}(\Delta_{\mathcal{E}})}$, when $(\mathscr{F}, d) \in \{(\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K), 5), (\mathscr{F}_1, 3), (\mathscr{F}_2, 2)\}$.

These weighted averages then give the equality cases of $\vartheta = \log_2 d$ for Theorems 1.6 and 1.11. As previously noted, the equality case of most interest here is likely that of $\mathscr{F}_2$, since it gives a bound for the average number of integral points on curves in $\mathscr{F}_2$. However, we note that in fact these equality cases recover all of Theorems 1.6 and 1.11, since they imply that the ϑth moment for all $\vartheta < \log_2 d$ will also be bounded.

Let us first briefly sketch the key ideas in the unweighted counting method that we will be modifying. In some recent papers such as [BS15a, BS15b, BS13a, BS13b, BH22], the main goal is to count (average) the d-Selmer elements ($d = 2, 3, 4$, or 5) of elliptic curves in either $\mathscr{F}_{\mathrm{univ}}$ or other families; these Selmer elements are parametrized by orbits—with both global and local conditions—of a group $G(\mathbb{Q})$ acting on associated representations $V(\mathbb{Q})$. It is possible to count the required rational orbits by counting integral orbits, each of which corresponds to a lattice point in a fundamental domain $\mathcal{F}$. Davenport's Lemma roughly says that the number of such lattice points is the volume of the domain, or at least the well-behaved part of the domain; other methods are used to count the lattice points in the remaining cusps. One thereby obtains an asymptotic count of the number of lattice points in the fundamental domain, and after applying sieves to impose the necessary local conditions, these counts give the desired Selmer averages.

In this section, we upper bound the average number of Selmer elements with weights of $O(1)^{\omega_{\geq 2}(\Delta_{\mathcal{E}})}$. We interpret this weighted average as a weighted average of the relevant vectors in the lattice $V(\mathbb{Z})$ (in the fundamental domain $\mathcal{F}$), and the weights (in fact, the discriminants) depend only on the $G(\mathbb{Q})$-invariants of the vector. This weighted average is an integral over the fundamental domain, and we split this integral into two pieces, depending on the size of the so-called torus parameter in defining $\mathcal{F}$. For large torus parameters, we use a pointwise bound on the weight function to reduce to integrating the unweighted volumes over this region that is "polynomially high in the cusp." For small torus parameters, we may use Davenport's Lemma to obtain enough equidistribution over congruence classes to apply Lemma 4.5, which then bounds the integrand by the unweighted count.

5.1. **Weighted averages of 5-Selmer elements of elliptic curves.**

Theorem 5.1. *Let $\mathscr{F} \subseteq \mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ be a subset of positive lower density, ordered by height. Let S be a finite set of places of K containing all infinite places. Let $\vartheta = \log_2 5 = 2.3219\ldots$. Then*

$$\mathrm{Avg}_{\mathscr{F}}(|\mathcal{E}_{A,B}(\mathcal{O}_{K,S})|^{\vartheta}) \ll_{\mathscr{F}} \zeta_K(2)^{O(1)} O(1)^{|S|} |\mathrm{Cl}(\mathcal{O}_{K,S})|^{\vartheta},$$

where the average is taken over all $\mathcal{E}_{A,B} \in \mathscr{F}$ ordered by height.

Theorem 5.1 follows immediately from the following weighted average bound, in exactly the same way as the beginning of the proof of Theorem 4.3.

Theorem 5.2.

$$\mathrm{Avg}_{\mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)}\left(O(1)^{\omega_{\geq 2}(\Delta_{A,B})}|\mathrm{Sel}_5(E/K)|\right) \ll \zeta_K(2)^{O(1)}$$

where the average is taken over all $\mathcal{E}_{A,B} \in \mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ ordered by height.

Proof. We take $K = \mathbb{Q}$ for convenience; the argument is exactly the same for general number fields K (since Lemma 4.5 is phrased in that generality), but when $K = \mathbb{Q}$, we may give precise citations to the intermediate results in [BS13b] rather than [BSW22]. So for the rest of the proof, let $K = \mathbb{Q}$. Let $\mathscr{F}_{\mathrm{univ}}^{\leq X}(\mathcal{O}_K)$ denote the elliptic curves $\mathcal{E}_{A,B} \in \mathscr{F}_{\mathrm{univ}}(\mathcal{O}_K)$ with $H(A,B) \leq X$, ordered by height.

To show the boundedness of this average, note that since $5^{\mathrm{rank}\, E(K)} \leq 1 + (|\mathrm{Sel}_5(E/K)| - 1)$ and we have shown in the course of the proof of Theorem 4.3 that

$$\mathrm{Avg}_{\mathcal{E}_{A,B} \in \mathscr{F}_{\mathrm{univ}}^{\leq X}(\mathcal{O}_K)}\left(O(1)^{\omega_{\geq 2}(\Delta_{A,B})}\right) \ll \zeta_K(2)^{O(1)},$$

it suffices to show that

$$\mathrm{Avg}_{\mathcal{E} \in \mathscr{F}_{\mathrm{univ}}^{\leq X}(\mathcal{O}_K)}\left(\sum_{1 \neq \alpha \in \mathrm{Sel}_5(E/K)} O(1)^{\omega_{\geq 2}(\Delta(\alpha))}\right) \tag{25}$$

is similarly bounded, i.e., where we only consider non-identity 5-Selmer elements.

Let $V(K) = K^5 \otimes \wedge^2(K^5)$ and $G(K)$ be the subquotient of $\mathrm{GL}_5(K)^2$ defined by $\{(g_1, g_2) \in \mathrm{GL}_5(K) \times \mathrm{GL}_5(K) : \det(g_1)^2 \det g_2 = 1\}/\{(\lambda I_5, \lambda^{-2} I_5) : \lambda \in K^\times\}$. Then elements of the 5-Selmer group of an elliptic curve $\mathcal{E}_{A,B}$ are in a bijective correspondence with $G(K)$-orbits on the set of locally soluble $\vec{v} \in V(K)$ with invariants equal to $-3A$ and $-27B$ [FS16], and the usual discriminant $\Delta(\mathcal{E}_{A,B})$ equals the discriminant $\Delta(\vec{v})$ of $\vec{v}$ under this correspondence. By relating rational and integral orbits, counting nontrivial 5-Selmer elements can be translated into counting so-called *strongly irreducible* locally soluble $G(\mathbb{Z})$-orbits of $V(\mathbb{Z})$, as in [BS13b, BSW22].

For the rest of this proof, we adopt notation from [BS13b] as needed. For example, let $\mathcal{F}$ be the fundamental domain for the left action of $G(\mathbb{Z})$ on $G(\mathbb{R})$ written as $\{nak : n \in N'(a), a \in A', k \in K\}$, where $K = \mathrm{SO}_5(\mathbb{R})^2$, $A' = \{a(s_1, \ldots, s_8) : s_i > c\}$ is a certain subset of pairs of diagonal matrices ($c > 0$ is an absolute constant), and N' is a certain subset of pairs of lower triangular matrices, as defined in [BS13b, p. 8]. Let $G_0 \subset G(\mathbb{R})$ be a compact semialgebraic left K-invariant set that is the closure of an open nonempty set. Let $dh = dn\, d^*a\, dk$ be the Haar measure on $G(\mathbb{R})$, suitably normalized, and set $C_{G_0} = 5\int_{h\in G_0} dh$. Let V^+ and V^- denote the subsets with positive and negative discriminant, respectively, and let $(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}$ denote the strongly irreducible elements in $V(\mathbb{Z})^{\pm}$. For a subset $\mathcal{S}$ of $V(\mathbb{Z})$, let $N(\mathcal{S};X)$ denote the number of $\vec{v} \in \mathcal{S}$ of height up to X. Let $R^{\pm}$ be fundamental sets for the action of $G(\mathbb{R})$ on $V^{\pm}(\mathbb{R})$ as in [BS13b, (4) on p. 8]. Finally, let $B^{\pm}(n,a;X)$ be the multiset $naG_0 \cdot R^{\pm}(X)$.

Now since nontrivial 5-Selmer elements correspond to locally soluble strongly irreducible $G(\mathbb{Q})$-orbits, whence fewer in number than strongly irreducible $G(\mathbb{Z})$-orbits, we may upper bound (25) by

$$\widetilde{N}(V(\mathbb{Z})^{\pm};Y) := \frac{1}{C_{G_0}} \int_{na\in\mathcal{F}} \sum_{\vec{v}\in B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dn\, d^*a$$

with $Y \asymp X^{12}$, as in [BS13b, (8) on p. 9] (but replacing the weight function 1 by $O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}$ and noting that their height $H(\vec{v}) \gg H(A,B)^{12}$). We now split the integral by the size of the torus parameter $\vec{s}$ that determines $a = a(\vec{s}) \in A'$:

$$\begin{aligned}\widetilde{N}(V(\mathbb{Z})^{\pm};Y) &= \frac{1}{C_{G_0}} \left(\int_{na\in\mathcal{F}:||\vec{s}||_\infty \leq Y^\eta} + \int_{na\in\mathcal{F}:||\vec{s}||_\infty > Y^\eta} \right) \sum_{\vec{v}\in B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dn\, d^*a \\ &\leq \frac{1}{C_{G_0}} \int_{na\in\mathcal{F}:||\vec{s}||_\infty \leq Y^\eta} \sum_{\vec{v}\in B^{\pm}(n,a;Y)\cap V(\mathbb{Z})^{\pm}} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dn\, d^*a \\ &\qquad + O\left(Y^{o(1)} \int_{na\in\mathcal{F}:||\vec{s}||_\infty > Y^\eta} |B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}|\, dn\, d^*a \right), \end{aligned} \tag{26}$$

where $\eta \in \mathbb{R}^+$ with $\eta \asymp 1$ a small constant.

To bound the second summand in (26), [BS13b, Proposition 18] gives

$$\begin{aligned}\int_{na\in\mathcal{F}:||\vec{s}||_\infty > Y^\eta} \left|\{\vec{v}\in B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}} : a_{12}(\vec{v}) = 0\}\right|\, dn\, d^*a &\ll N(V(\mathbb{Z})^{\mathrm{irr}}(0);Y) \\ &\ll Y^{5/6-\Omega(1)}\end{aligned}$$

where $V(\mathbb{Z})^{\mathrm{irr}}(0)$ is the set of strongly irreducible $\vec{v}\in V(\mathbb{Z})$ where a specific entry a_{12} of $\vec{v}$ vanishes. But from the proof of [BS13b, Proposition 18], we have

$$\left|\{\vec{v}\in B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}} : a_{12}(\vec{v}) \neq 0\}\right| = 0$$

if $Y^{1/60}w(a_{12}) \ll 1$, where $w(a_{12}) = s_1^{-3}s_2^{-6}s_3^{-4}s_4^{-2}s_5^{-4}s_6^{-3}s_7^{-2}s_8^{-1}$ as defined in the proof of [BS13b, Proposition 18]. So

$$\begin{aligned}&\int_{na\in\mathcal{F}:||\vec{s}||_\infty > Y^\eta} |B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}|\, dn\, d^*a \\ &\qquad \ll Y^{5/6-\Omega(1)} + \int_{na\in\mathcal{F}:||\vec{s}||_\infty > Y^\eta \text{ and } Y^{1/60}w(a_{12})\gg 1} |B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}|\, dn\, d^*a.\end{aligned} \tag{27}$$

Now just as in the deduction of (15) from (14) in [BS13b, p. 16], Davenport's Lemma (e.g., [BS13b, Proposition 17]) implies that when $Y^{1/60}w(a_{12}) \gg 1$, we have

$$\left|B^{\pm}(n,a;Y)\cap(V(\mathbb{Z})^{\pm})^{\mathrm{irr}}\right| \leq \left|B^{\pm}(n,a;Y)\cap V(\mathbb{Z})^{\pm}\right| \ll \mathrm{Vol}(B^{\pm}(n,a;Y)).$$

But then

$$\begin{aligned}\int_{na\in\mathcal{F}:||\vec{s}||_\infty>Y^\eta \text{ and } Y^{1/60}w(a_{12})\gg 1} &\left|B^\pm(n,a;Y)\cap (V(\mathbb{Z})^\pm)^{\mathrm{irr}}\right| \, dn\, d^*a \\ &\ll \int_{na\in\mathcal{F}:||\vec{s}||_\infty>Y^\eta \text{ and } Y^{1/60}w(a_{12})\gg 1} \mathrm{Vol}(B^\pm(n,a;Y))\, dn\, d^*a \\ &\ll \mathrm{Vol}(B^\pm(1,1;Y))\int_{na\in\mathcal{F}:||\vec{s}||_\infty>Y^\eta \text{ and } Y^{1/60}w(a_{12})\gg 1} dn\, d^*a \\ &\ll Y^{5/6-\Omega(1)}. \end{aligned} \tag{28}$$

Replacing the second summand in (26) with the bounds from (27) and (28) yields

$$\widetilde{N}(V(\mathbb{Z})^\pm;Y) \le \frac{1}{C_{G_0}}\int_{na\in\mathcal{F}:||\vec{s}||_\infty\le Y^\eta} \sum_{\vec{v}\in B^\pm(n,a;Y)\cap V(\mathbb{Z})^\pm} O(1)^{\omega_{\ge 2}(\Delta(\vec{v}))}\, dn\, d^*a + O\left(Y^{5/6-\Omega(1)}\right). \tag{29}$$

In order to bound the first summand, we wish to apply Lemma 4.5. To check the nontrivial hypothesis (b) in Lemma 4.5, we repeat the arguments at the end of the proof of Theorem 4.6. A similar Hensel lifting argument proves the analogue of (17). We obtain equidistribution in congruence classes as in (18) by applying Davenport's Lemma (e.g., [BS13b, Proposition 17]) and using $||\vec{s}||_\infty \le Y^\eta$ to conclude that for all $m \le Y^\eta$ and $\vec{v}_0 \in V(\mathbb{Z}/m\mathbb{Z})$,

$$\left|\{\vec{v}\in B^\pm(n,a;Y)\cap V(\mathbb{Z}) : \vec{v}\equiv \vec{v}_0 \pmod m\}\right| \ll m^{-\dim V}\left|B^\pm(n,a;Y)\cap V(\mathbb{Z})\right|,$$

since both sides are $\asymp \mathrm{Vol}\left(\frac{B^\pm(n,a;Y)-\vec{v}_0}{m}\right) = m^{-\dim V}\mathrm{Vol}\left(B^\pm(n,a;Y)\right)$.

Thus, we may now apply Lemma 4.5, which shows that when $||\vec{s}||_\infty \le Y^\eta$,

$$\sum_{\vec{v}\in B^\pm(n,a;Y)\cap V(\mathbb{Z})^\pm} O(1)^{\omega_{\ge 2}(\Delta(\vec{v}))} \ll \sum_{\vec{v}\in B^\pm(n,a;Y)\cap V(\mathbb{Z})^\pm} 1. \tag{30}$$

Integrating (30) over $\mathcal{F}$ shows that the first summand of (29) is bounded by $N(V(\mathbb{Z})^\pm;Y)$ (the unweighted count).[5] We thus have

$$\widetilde{N}(V(\mathbb{Z})^\pm;Y) \ll N(V(\mathbb{Z})^\pm;Y) + O(Y^{5/6-\Omega(1)}).$$

But by [BS13b, Theorem 12] and [BS13b, Remark 13], we have

$$N(V(\mathbb{Z})^\pm;Y) \ll Y^{5/6} \asymp X^{10} \asymp \sum_{\substack{(A,B)\in\mathscr{F}_{\mathrm{univ}}(\mathbb{Z})\\ H(A,B)\le X}} 1,$$

whence

$$\mathrm{Avg}_{\mathcal{E}\in\mathscr{F}^{\le X}_{\mathrm{univ}}(\mathbb{Z})}\left(\sum_{1\ne\alpha\in\mathrm{Sel}_5(E/\mathbb{Q})} O(1)^{\omega_{\ge 2}(\Delta(\alpha))}\right) \ll 1$$

as desired. □

5.2. **Weighted averages of 2- and 3-Selmer elements of elliptic curves in families with marked points.** We now prove the equality cases of Theorem 4.6 for $\mathscr{F}_1$ and $\mathscr{F}_2$, using nearly identical arguments as in Section 5.1.

[5] This application of Lemma 4.5 is where the $\zeta_K(2)$ term appears, but here we are taking $K = \mathbb{Q}$, so we may just use 1 on the right side of (30).

Theorem 5.3. *Let $(\mathscr{G}, \vartheta) \in \{(\mathscr{F}_1, \log_2 3), (\mathscr{F}_2, 1)\}$. Let $\mathscr{F} \subseteq \mathscr{G}$ be a subfamily of positive lower density ordered by height. Let S be a finite set of places of $\mathbb{Q}$. Then:*

$$\mathrm{Avg}_{\mathscr{F}}(|\mathcal{E}(\mathbb{Z}[S^{-1}])|^{\vartheta}) \ll_{\mathscr{F}} 1,$$

where the average is taken over all $\mathcal{E} \in \mathscr{F}$ ordered by height.

Just as in Section 5.1, after repeating the arguments in the beginning of the proof of Theorem 4.3, Theorem 5.3 follows from a weighted average bound:

Theorem 5.4. *Let $(\mathscr{G}, d) \in \{(\mathscr{F}_1, 3), (\mathscr{F}_2, 2)\}$. Then*

$$\mathrm{Avg}_{\mathscr{G}}\left(O(1)^{\omega_{\geq 2}(\Delta_{\mathcal{E}})}|\mathrm{Sel}_d(E)|\right) \ll 1,$$

where the average is taken over all $\mathcal{E} \in \mathscr{G}$ ordered by height.

Proof. This proof is completely analogous to the proof of Theorem 5.2, except with references to [BH22] instead of [BS13b]. For the remainder of the proof, we use the notation of [BH22]. For example, when $\mathscr{G} = \mathscr{F}_1$, let the group G be the quotient of SL_3^3 by the stabilizer μ_3^2 and V be the representation $3 \otimes 3 \otimes 3$ of G; when $V = \mathscr{F}_2$, let G be the quotient of SL_4^2 by the analogous μ_2^3 and V be the representation $2 \otimes 2 \otimes 2 \otimes 2$ of G (see Cases 4 and 7 in [BH22, Theorem 3.1]). Let n be the dimension of V and k be the degree of the discriminant, so $(n, k) = (27, 36)$ and $(16, 24)$ for $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively. For elliptic curves E whose affine Weierstrass models are in $\mathscr{F}_1$ and $\mathscr{F}_2$, the notation $S'(E)$ denotes a subgroup of the d-Selmer group $\mathrm{Sel}_d(E)$ that is generated by the marked points; for 100% of the curves in $\mathscr{F}_1$ (resp., $\mathscr{F}_2$), this group has order 3 (resp., 4).

The subset of $V(\mathbb{R})$ with nonzero discriminant is split into N connected components denoted $V^{(i)}$ for $1 \leq i \leq N$. Each contains a fundamental set $R^{(i)}$ defined in [BH22, §5.1]. Let $\mathcal{F}$ be a fundamental domain for the left action of $G(\mathbb{Z})$ on $G(\mathbb{R})$ that is Haar-measurable and contained in a standard Siegel set, written as a subset of a product of the same sort of fundamental domain $\mathcal{F}_j$ for SL_j ($j = 3$ or 2 for $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively). We may explicitly specify

$$\mathcal{F}_2 = \{\nu\alpha\kappa : \nu(x) \in N'(\alpha), \alpha(s) \in A', \kappa \in K\}$$
$$\mathcal{F}_3 = \{\nu\alpha\kappa : \nu(x, x', x'') \in N'(\alpha), \alpha(t, u) \in A', \kappa \in K\}$$

where A' is a certain subset of the appropriate tori (indexed by s or (t, u), resp.), $N'(\alpha)$ is a subset of lower triangular matrices, and $K = \mathrm{SO}_i(\mathbb{R})$, as in [BH22, §5.2]. As before, let $G_0 \subset G(\mathbb{R})$ be a compact semialgebraic left K-invariant set that is the closure of an open nonempty set. Let $E^{(i)}(\nu, \alpha, Y)$ denote the multiset $\nu\alpha G_0 \cdot R^{(i)} \cap \{\vec{v} \in V^{(i)} : H(\vec{v}) < X\}$.

We now begin the proof. Because $\mathrm{Avg}_{\mathcal{E} \in \mathscr{G}: H(\mathcal{E}) \leq X}\left(O(1)^{\omega_{\geq 2}(\Delta_{\mathcal{E}})}\right) \ll 1$ (as shown in the proof of Theorem 4.6 using Lemma 4.5), we reduce to showing

$$\mathrm{Avg}_{\mathcal{E} \in \mathscr{G}: H(\mathcal{E}) \leq X}\left(\sum_{\vec{v} \in \mathrm{Sel}_d(E) - S'(E)} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\right) \ll 1. \tag{31}$$

We are removing the orbits corresponding to the subgroup $S'(E)$ in $\mathrm{Sel}_d(E)$ because we only wish to modify the count for *irreducible* orbits. We bound the number of irreducible $G(\mathbb{Q})$-orbits by the number of irreducible $G(\mathbb{Z})$-orbits (here, we drop the constants $C_{G_0}^{(i)} \asymp 1$):

$$\begin{aligned}
&\mathrm{Avg}_{\mathcal{E} \in \mathscr{G}: H(\mathcal{E}) \leq X}\left(\sum_{\vec{v} \in \mathrm{Sel}_d(E) - S'(E)} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\right) \\
&\qquad \ll \sum_{i=1}^{N} \int_{\nu\alpha \in N(\alpha)A'} \sum_{\vec{v} \in E^{(i)}(\nu, \alpha, Y) \cap V(\mathbb{Z})^{\mathrm{irr}}} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dg
\end{aligned} \tag{32}$$

with $Y \asymp X$, as in the derivation of [BH22, (22) in §7.1] (but replacing the weight function 1 by $O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}$ and noting that we are using the same height). Let $\widetilde{N}(V(\mathbb{Z})^{(i)};Y)$ denote the ith summand in (32). Writing $\vec{\sigma} := \vec{s}$ or $(\vec{t},\vec{u})$ (where, e.g., s_j corresponds to the torus parameter s in the jth factor of $\mathcal{F}_2$ in $\mathcal{F} = \mathcal{F}_2^4$), we split the integral to obtain

$$\begin{aligned}\widetilde{N}(V(\mathbb{Z})^{(i)};Y) &= \left(\int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty\leq Y^\eta} + \int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty > Y^\eta}\right) \sum_{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dg \\ &\leq \int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty\leq Y^\eta} \sum_{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dg \\ &\quad + O\left(Y^{o(1)} \int_{na\in N(\alpha)A':||\vec{\sigma}||_\infty > Y^\eta} \left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| dg\right),\end{aligned} \tag{33}$$

where $\eta \in \mathbb{R}^+$ with $\eta \asymp 1$ is a small constant.

We use [BH22, Proposition 7.3] to see that

$$\int_{\nu\alpha\in\mathcal{F}:||\vec{\sigma}||_\infty > Y^\eta} \left|\{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}} : b_{\min}(\vec{v}) = 0\}\right| dg \ll Y^{n/k-\Omega(1)}$$

where $b_{\min}$ is a specific (the 111 or 1111) entry of $\vec{v}\in V$. Again, as seen in (31) of [BH22],

$$\left|\{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}} : b_{\min}(\vec{v}) \neq 0\}\right| = 0$$

when $Y^{1/k}w(b_{\min}) \ll 1$, where $w(b_{\min}) = \prod_{j=1}^3 t_j^{-2}u_j^{-1}$ or $\prod_{j=1}^4 s_j^{-1}$ for $\mathscr{G} = \mathscr{F}_1$ or $\mathscr{F}_2$, respectively. Thus, we may write

$$\begin{aligned}&\int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty > Y^\eta} \left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| dg \\ &\qquad \ll Y^{n/k-\Omega(1)} + \int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty > Y^\eta, Y^{1/k}w(b_{\min})\gg 1} \left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| dg.\end{aligned}$$

By Davenport's Lemma (e.g., [BH22, Lemma 7.2]), when $Y^{1/k}w(b_{\min}) \gg 1$,

$$\left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| \leq \left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| \ll \mathrm{Vol}(E^{(i)}(\nu,\alpha,Y)),$$

so a computation like (28) gives

$$\int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty > Y^\eta, Y^{1/k}w(b_{\min})\gg 1} \left|E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})^{\mathrm{irr}}\right| dg \ll Y^{n/k-\Omega(\eta)}.$$

Combining this with (33) yields

$$\widetilde{N}(V(\mathbb{Z})^{(i)};Y) \leq \int_{\nu\alpha\in N(\alpha)A':||\vec{\sigma}||_\infty\leq Y^\eta} \sum_{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))}\, dg + O\left(Y^{n/k-\Omega(1)}\right).$$

Finally, we conclude by applying Lemma 4.5, again verifying the hypotheses by repeating the Hensel lifting argument proving the analogue of (17) and modifying the equidistribution argument proving (18) by applying Davenport's Lemma and using the fact that $||\vec{\sigma}||_\infty \leq Y^\eta$. By Lemma 4.5, when $||\vec{\sigma}||_\infty \leq Y^\eta$,

$$\sum_{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})} O(1)^{\omega_{\geq 2}(\Delta(\vec{v}))} \ll \sum_{\vec{v}\in E^{(i)}(\nu,\alpha,Y)\cap V(\mathbb{Z})} 1,$$

which gives

$$\widetilde{N}(V(\mathbb{Z})^{(i)};Y) \ll N(V(\mathbb{Z})^{(i)};Y) + O(Y^{n/k-\Omega(1)}),$$

where $N(V(\mathbb{Z})^{(i)};Y)$ denotes the number of integral points in $V^{(i)}$ of height less than Y. But by [BH22, Theorem 7.1] and [BH22, Theorem 9.1],

$$N(V(\mathbb{Z})^{(i)};Y) \ll Y^{n/k} \asymp X^{n/k} \asymp \sum_{\mathcal{E}\in\mathscr{G}:H(\mathcal{E})\leq X} 1,$$

so we have bounded each summand of (32) and thus obtain the desired bound (31). □

References

[Abr97] Dan Abramovich, *Uniformity of stably integral points on elliptic curves*, Invent. Math. **127** (1997), no. 2, 307–317. MR 1427620

[Akh12] Shabnam Akhtari, *Upper bounds for the number of solutions to quartic Thue equations*, Int. J. Number Theory **8** (2012), no. 2, 335–360. MR 2890483

[Alp14] Levent Alpöge, *The average number of integral points on elliptic curves is bounded*, 2014, https://arxiv.org/abs/1412.1047.

[AO10] Shabnam Akhtari and Ryotaro Okazaki, *Quartic Thue equations*, J. Number Theory **130** (2010), no. 1, 40–60. MR 2569841

[BH16] Manjul Bhargava and Wei Ho, *Coregular spaces and genus one curves*, Cambridge J. Math. **4** (2016), no. 1, 1–119.

[BH22] ———, *On average sizes of Selmer groups and ranks in families of elliptic curves having marked points*, 2022, https://arxiv.org/abs/2207.03309.

[BS87] E. Bombieri and W. M. Schmidt, *On Thue's equation*, Invent. Math. **88** (1987), no. 1, 69–81.

[BS13a] Manjul Bhargava and Arul Shankar, *The average number of elements in the 4-Selmer groups of elliptic curves is 7*, 2013, http://arxiv.org/abs/1312.7333.

[BS13b] ———, *The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1*, 2013, http://arxiv.org/abs/1312.7859.

[BS15a] ———, *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.

[BS15b] ———, *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.

[BSD63] B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves. I*, J. Reine Angew. Math. **212** (1963), 7–25.

[BSW22] Manjul Bhargava, Arul Shankar, and Xiaoheng Wang, *Geometry-of-numbers methods over global fields II: Coregular vector spaces*, 2022, preprint.

[CFS10] John E. Cremona, Tom A. Fisher, and Michael Stoll, *Minimisation and reduction of 2-, 3- and 4-coverings of elliptic curves*, Algebra Number Theory **4** (2010), no. 6, 763–820.

[Dav92] Sinnou David, *Autour d'une conjecture de S. Lang*, Approximations diophantiennes et nombres transcendants (Luminy, 1990), de Gruyter, Berlin, 1992, pp. 65–98. MR 1176524

[Eve84] J.-H. Evertse, *On equations in S-units and the Thue-Mahler equation*, Invent. Math. **75** (1984), no. 3, 561–584. MR 735341

[Eve97] Jan-Hendrik Evertse, *The number of solutions of the Thue-Mahler equation*, J. Reine Angew. Math. **482** (1997), 121–149. MR 1427659

[FS16] Tom Fisher and Mohammad Sadek, *On genus one curves of degree 5 with square-free discriminant*, J. Ramanujan Math. Soc. **31** (2016), no. 4, 359–383.

[HS88] M. Hindry and J. H. Silverman, *The canonical height and integral points on elliptic curves*, Invent. Math. **93** (1988), no. 2, 419–450.

[HV06] 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.

[Kim18] Dohyeong Kim, *Descent for the punctured universal elliptic curve, and the average number of integral points on elliptic curves*, Acta Arith. **183** (2018), no. 3, 201–222. MR 3813218

[Mor69] L. J. Mordell, *Diophantine equations*, Pure and Applied Mathematics, Vol. 30, Academic Press, London-New York, 1969. MR 0249355

[Sil87] Joseph H. Silverman, *A quantitative version of Siegel's theorem: integral points on elliptic curves and Catalan curves*, J. Reine Angew. Math. **378** (1987), 60–100. MR 895285