HOW MANY RATIONAL POINTS DOES A RANDOM CURVE HAVE?

WEI HO

Abstract. A large part of modern arithmetic geometry is dedicated to or motivated by the study of rational points on varieties. For an elliptic curve over $\mathbb{Q}$, the set of rational points forms a finitely generated abelian group. The ranks of these groups, when ranging over all elliptic curves, are conjectured to be evenly distributed between rank 0 and rank 1, with higher ranks being negligible. We will describe these conjectures and discuss some results on bounds for average rank, highlighting the recent work of Bhargava and Shankar.

Contents

1. Rational points on varieties

Finding solutions to polynomial equations is one of the oldest problems in mathematics. Over the last few centuries, mathematicians have formalized the questions and established rigorous language to discuss this simple idea in different variations. While we have made tremendous strides in understanding the structure of these solutions in the last few decades, there remain many fundamental open questions, which lie at the forefront of modern arithmetic geometry.

2010 *Mathematics Subject Classification.* Primary 11G05, 14H52. Secondary 11G30, 14H25.
Partially supported by NSF grant DMS-0902853.

For the simplest case, let $f(x_1, \ldots, x_n)$ be a polynomial with coefficients in $\mathbb{Q}$. We may ask for rational solutions to $f = 0$, i.e., numbers $a_1, \ldots, a_n \in \mathbb{Q}$ such that $f(a_1, \ldots, a_n) = 0$. To phrase the question more geometrically, let X be the **variety** associated to f, which may be viewed geometrically as the *zero locus* of f, or solutions to $f = 0$, in $\mathbb{C}^n$. Then X will be $(n-1)$-dimensional, if f is not identically zero. Our problem may be restated as finding **rational points** on X, the set of which is denoted $X(\mathbb{Q})$. More specifically, we may ask questions such as the following:

- Does there exist a single rational point on X?
- Can we describe all rational points on X?
- If there are only finitely many rational points, can we enumerate them?
- If $X(\mathbb{Q})$ is an infinite set, what structure does $X(\mathbb{Q})$ have?

If we instead use any finite number of polynomials $f_1, \ldots, f_m \in \mathbb{Q}[x_1, \ldots, x_n]$, we define the analogous variety X to be the common zero locus of all of the polynomials f_i in $\mathbb{C}^n$. For general choices of f_i, the dimension of X will be $n-m$, if nonnegative, and 0 otherwise; the rule of thumb is that each polynomial condition imposed should reduce the dimension of X by 1.

Remark 1.1. While we restrict our attention to varieties defined over $\mathbb{Q}$, i.e., defined by polynomials with coefficients in $\mathbb{Q}$, many of the results that we will discuss have analogues over other number fields.

Even when X is 1-dimensional, mathematicians have not yet fully understood how many rational points are on X! In this note, we focus on this case, where X is a **curve**.

1.1. **Genus and the trichotomy.** The arithmetic and the geometry of algebraic curves rely heavily on an invariant called the **genus**. The genus of a curve may be defined in many ways, but the most intuitive definition is topological. A smooth curve X as defined above may be thought of as a **Riemann surface**[1] with finitely many punctures; after taking an appropriate compactification by filling these punctures, the resulting compact Riemann surface has a topological genus, which is essentially the number of "holes" or "handles." For example, a complex curve that is homeomorphic to a sphere (after compactification) has genus 0, while a genus 1 curve looks like the surface of a donut.

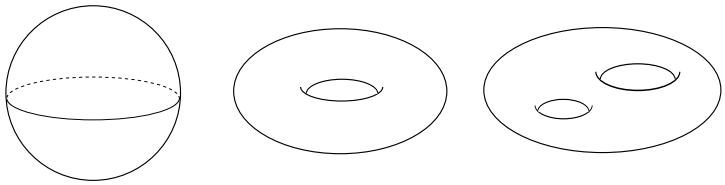


FIGURE 1. From left to right: curves of genus 0, 1, and 2.

For simplicity, we assume in the sequel that our curves X have been compactified[2], have no singularities[3], and are connected.

[1]The complex points of a smooth curve form a *two*-dimensional real manifold.

[2]In other words, we always implicitly work with projective curves.

[3]Intuitively, a *singularity* on a curve is a point where it is not smooth, like a node or a cusp. More precisely, it is a point with more than one tangent direction along the curve.

Geometric and arithmetic properties of curves are heavily influenced by their genera. For example, we have the following trichotomy:

	genus 0	genus 1	genus ≥ 2
canonical bundle	anti-ample	trivial	ample
curvature	positive	zero	negative
Kodaira dimension	$\kappa = -\infty$	$\kappa = 0$	$\kappa = 1$
automorphism group	3-dimensional	1-dimensional	finite
rational points	Hasse principle	finitely generated	finitely many

We now discuss the last row in more detail.

Curves of genus 0. For genus 0 curves, there are either no rational points at all or infinitely many, and it is fairly easy to determine which case applies to any given curve by the **Hasse principle**.

In particular, a genus 0 curve X has a rational point if and only if it has a point everywhere *locally*, which means that the equations defining X have a solution over the real numbers $\mathbb{R}$ and the p-adic numbers $\mathbb{Q}_p$ for all primes p. The non-existence of a $\mathbb{Q}_p$-point is always due to an *obstruction* modulo a power of the prime p.

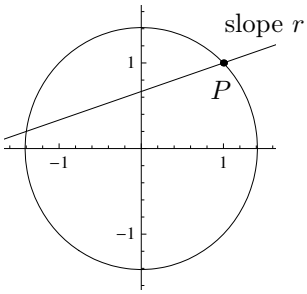
Example 1.2. Let X be the curve given by the vanishing of the polynomial $f = x^2 + y^2 - 3$. If there exists a rational solution to $f = 0$, by clearing denominators, there are relatively prime integers r, s, and t such that $r^2 + s^2 = 3t^2$. Because squares of integers are congruent to 0 or 1 modulo 4, reducing the equation modulo 4 shows that r^2 and s^2 are both congruent to 0 modulo 4. This in turn implies that all three integers are even, which is a contradiction. Therefore, the equation $f = 0$ has an obstruction modulo 4, implying that X has no point over $\mathbb{Q}_2$ and thus no rational point.

In fact, checking for local obstructions may be completed in a finite number of steps. Any curve of genus 0 over $\mathbb{Q}$ is isomorphic to a (compactified) plane conic defined by the vanishing of a polynomial of the form

$$ax^2 + by^2 + c, \tag{1}$$

where a, b, and c are squarefree, pairwise relatively prime integers. A theorem of Legendre implies that (1) has a rational solution if and only if a, b, and c do not all have the same sign, and $-ab$ is a square modulo c, $-bc$ is a square modulo a, and $-ac$ is a square modulo b.

If there is a single rational point P on a conic, then all other rational points come from intersecting the conic with a line of rational slope through P.



Example 1.3. If X is given by $x^2 + y^2 - 2 = 0$, then by inspection, the point $(x, y) = (1, 1)$ lies on X. All other points are parametrized by drawing lines of rational (or infinite) slope r through $(1, 1)$, and a simple computation shows that $X(\mathbb{Q})$ is the union of the point $(1, -1)$ and the points

$$\left(\frac{r^2 - 2r - 1}{r^2 + 1}, \frac{-r^2 - 2r + 1}{r^2 + 1}\right)$$

for all $r \in \mathbb{Q}$.

Curves of genus at least 2. Mordell's 1922 conjecture [Mor22] predicted that there could not be very many rational points on a curve of genus ≥ 2; it was proved by Faltings [Fal83] (as a corollary to an even more powerful theorem):

Theorem 1.4 (Faltings 1983)**.** *Let X be a curve of genus at least 2 over $\mathbb{Q}$. Then the set $X(\mathbb{Q})$ of rational points is finite.*

The original proof uses deep ideas from p-adic Hodge theory, Arakelov theory, and moduli theory, and later proofs and improvements by Vojta [Voj91], Faltings [Fal91], Bombieri [Bom90], and others use Diophantine approximation methods. None of the proofs, however, are *effective* in the sense of giving a list of the points in $X(\mathbb{Q})$. In practice, a combination of techniques — including Chabauty's method, Brauer-Manin obstructions, and descent — often are enough to produce the points in $X(\mathbb{Q})$. In §4.3, we will outline recent progress on bounding the number of rational points on curves of genus ≥ 2.

Curves of genus 1. The case of genus 1 curves is the richest arithmetically, the most complicated, and the most mysterious to this day. A genus one curve defined over $\mathbb{Q}$ may have no rational points at all, finitely many, or infinitely many — and it is generally difficult to determine which! Techniques for other genera, like the Hasse principle, no longer apply, e.g., there are plenty of genus one curves which have points everywhere locally but no global rational point.

Genus one curves over $\mathbb{Q}$ with a given rational point are known as **elliptic curves**. Section 1.2 will describe the structure of rational points on elliptic curves in more detail.

1.2. **Rational points on elliptic curves.** An elliptic curve over $\mathbb{Q}$ is isomorphic to the projective closure of the zero locus of a Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{2}$$

with all $a_i \in \mathbb{Q}$. When defining a nonsingular curve, the equation (2) may be transformed (over $\mathbb{Q}$) into **short Weierstrass form**

$$y^2 = x^3 + Ax + B \tag{3}$$

for $A, B \in \mathbb{Q}$ with nonzero discriminant $\Delta = -16(4A^3 + 27B^2)$; the nonvanishing of the discriminant ensures that the curve is nonsingular. There is a marked rational point "at infinity," which is denoted O. We say that the elliptic curve given by (3) has **height**

$$\mathrm{ht}(E) := \max(4|A|^3, 27B^2).$$

The coefficients of 4 and 27 are for convenience; only the exponents matter for most purposes.

Many such equations define isomorphic elliptic curves. In particular, for $t \in \mathbb{Q}^\times$, scaling x and y by t^{-2} and t^{-3}, respectively, in equation (3) gives the new equation

$$y^2 = x^3 + t^4A + t^6B.$$

In other words, the group $\mathbb{Q}^\times$ acts on the space of all equations of the form (3) with nonzero discriminant. To choose one representative equation from each $\mathbb{Q}^\times$-orbit, we define **minimal Weierstrass** equations to be those of the form (3), with $A, B \in \mathbb{Z}$ and the condition that there is no prime p such that p^4 divides A and p^6 divides B. Each elliptic curve over $\mathbb{Q}$ has a unique minimal Weierstrass model, and we will call its discriminant the **minimal** discriminant of the elliptic curve.

FIGURE 2. The real points of the elliptic curves $y^2 = x^3 - x + 1$ (left) and $y^2 = x^3 - x$ (right).

The solutions of a Weierstrass equation lying in any field have a rich structure. The complex points of an elliptic curve make up a one-holed torus, as discussed earlier. The real points are smooth curves in $\mathbb{R}^2$ with one or two components; see Figure 2.

Group Law. A beautiful and incredibly useful fact is that the set of solutions with values in any given field forms a group! An even more powerful statement for rational points is given by a theorem of Mordell [Mor22]:

Theorem 1.5 (Mordell 1922)**.** *The set $E(\mathbb{Q})$ of rational points of an elliptic curve E defined over $\mathbb{Q}$ forms a finitely generated abelian group, i.e.,*

$$E(\mathbb{Q}) = \mathbb{Z}^r \oplus E(\mathbb{Q})_{\text{tors}} \tag{4}$$

for some nonnegative integer r and finite abelian group $E(\mathbb{Q})_{\text{tors}}$.

The group structure on the points of an elliptic curve uses the point O at infinity as the identity element, and it is most easily described geometrically. For the graph of an elliptic curve in short Weierstrass form, as seen in Figure 3, the line L through any two points P_1 and P_2 will intersect a third point P_3 by Bezout's theorem. The vertical line through P_3 intersects another point on the elliptic curve, which is the composition $P_1 + P_2$ of P_1 and P_2.

In other words, the three (not necessarily distinct) intersection points P_1, P_2, and P_3 of any line L with the elliptic curve sum to the identity in the group law. The identity point O may be one of these points, e.g., a vertical line intersects O, a point P, and its negative. Moreover, if P_1 and P_2 are rational points, then the line L has rational slope, so $P_1 + P_2$ is also a rational point.

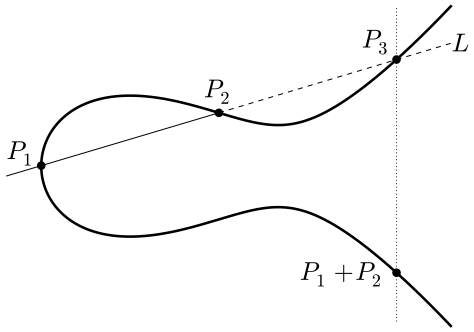


FIGURE 3. The group law on an elliptic curve.

For an elliptic curve E over $\mathbb{Q}$, the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ of $E(\mathbb{Q})$ is fairly well understood, by a deep theorem of Mazur [Maz77]:

Theorem 1.6 (Mazur 1977)**.** *For an elliptic curve E defined over $\mathbb{Q}$, the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ is one of the following groups:*

$$\begin{aligned} &\mathbb{Z}/d\mathbb{Z} && \text{for } 1 \le d \le 10 \text{ or } d = 12 \\ &\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2d\mathbb{Z} && \text{for } 1 \le d \le 4. \end{aligned}$$

By Theorem 1.6 and Hilbert's irreducibility theorem, almost all[4] elliptic curves have a trivial torsion subgroup. Moreover, there are explicit methods to quickly compute the torsion subgroup for any given elliptic curve. For example, the rational 2-torsion points of an elliptic curve in short Weierstrass form are determined by factoring the right hand side cubic polynomial of (3) over $\mathbb{Q}$.

Example 1.7. The elliptic curve $y^2 = x(x-1)(x-2)$ has rational 2-torsion subgroup $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, consisting of the points $(0,0)$, $(1,0)$, $(2,0)$, and O (as the identity). The elliptic curve $y^2 = x(x^2+x+1)$ has only the rational 2-torsion points $(0,0)$ and O.

In contrast, the **rank** of $E(\mathbb{Q})$, denoted r in (4), is much more mysterious! It is not even known if the rank can be arbitrarily large; the current record — due to Elkies — is an elliptic curve of rank at least 28. It is quite difficult in general to rigorously prove that a given elliptic curve has a certain rank, though L-function computations often give conjectural answers, as we describe very briefly below.

Analytic rank. The **L-function** $L(E,s)$ of an elliptic curve E over $\mathbb{Q}$ is a Dirichlet series given by an Euler product formula involving the number N_p of $\mathbb{F}_p$-points on the reduction of E over $\mathbb{F}_p$, for all primes p:

$$L(E,s) := \prod_{\text{good } p} \frac{1}{1 - a_p p^{-s} + p^{1-2s}} \prod_{\text{bad } p} \frac{1}{1 - a_p p^{-s}} = \sum_{n \ge 1} a_n n^{-s}$$

where $a_p = 1 + p - N_p$ for "good" primes p and $a_p = -1$, 0, or 1 for a finite set of "bad" primes p. By the work of Wiles and others [Wil95, TW95, BCDT01], the L-function extends to an entire function on the complex plane, and $\Lambda(E,s) := \text{cond}(E)^{s/2}(2\pi)^{-s}\Gamma(s)L(E,s)$ satisfies a functional equation

$$\Lambda(E,s) = u_E \Lambda(E, 2-s),$$

where $\text{cond}(E)$ is an invariant of E called the **conductor**. The **root number** u_E is either $+1$ or -1.

The order of vanishing of this L-function $L(E,s)$ at $s = 1$ is the **analytic rank** of E. The Birch and Swinnerton-Dyer (BSD) Conjecture [BSD63, BSD65] claims that the analytic rank of E is equal to the rank of E defined earlier. A more refined version in fact gives a formula for the coefficient of the leading term in the Taylor expansion at $s = 1$ in terms of arithmetic invariants of E, including the Tate–Shafarevich group $Ш(E)$, which will be discussed in §2.3.

Thus, assuming the BSD conjecture, one may study the rank of an elliptic curve by understanding its analytic rank, e.g., by computing the apparent order of vanishing of $L(E,s)$ at $s = 1$.

[4]Here, "almost all" means that the density of curves, as defined in §2.1, with trivial torsion subgroup is 1, when ordered by height.

2. Ranks of elliptic curves

Most of the remainder of this note is devoted to conjectures and results on the *distribution* of ranks for elliptic curves. In other words, if we choose a "random" elliptic curve, what do we expect its rank to be?

2.1. **Densities and averages.** In order to formulate our question rigorously, we need to specify what is meant by "random" in this setting. We first need an ordering for the (infinite) set of elliptic curves, usually by some sort of invariant; possibilities include ordering elliptic curves up to isomorphism by conductor or by minimal discriminant, or ordering short Weierstrass equations by height or discriminant.

In all these cases, there are only finitely many objects (elliptic curves up to isomorphism or short Weierstrass equations with integral coefficients, for example) with the absolute value of that invariant bounded by any positive number X. When the invariant is the discriminant or the conductor, this finiteness is due to Siegel's classical theorem on the finiteness of S-integral points on elliptic curves [Sie66]. We may thus define quotients like the following:

$$\mathrm{P}(\text{invariant}, X, \mathrm{rk} = i) := \frac{\#\{\text{elliptic curves } E \text{ with invariant } \leq X \text{ and rank } i\}}{\#\{\text{elliptic curves } E \text{ with invariant } \leq X\}}$$

where the invariant could be the conductor, absolute value of the discriminant, or the height, for example. Then we might ask whether this quantity converges as X tends to infinity, and if so, we may consider the limit

$$\mathrm{P}(\text{invariant}, \mathrm{rk} = i) := \lim_{X\to\infty} \mathrm{P}(\text{invariant}, X, \mathrm{rk}(E) = i)$$

as the **density** of elliptic curves, ordered by that invariant, with rank i (or equivalently, the probability an elliptic curve has rank i). We may define a lower density or upper density by instead using liminf or limsup, respectively.

We then define the **average rank** for elliptic curves, ordered by an invariant, as

$$\lim_{X\to\infty} \sum_{i\geq 0} i \cdot \mathrm{P}(\text{invariant}, X, \mathrm{rk} = i) = \lim_{X\to\infty} \frac{\sum_{\text{invariant}(E)\leq X} \mathrm{rk}(E)}{\sum_{\text{invariant}(E)\leq X} 1}$$

if this limit exists. Again, a lower average or an upper average is defined using liminf or limsup, respectively; we will sometimes call these the limsup and the liminf of the average rank.

We may also define averages or higher moments for distributions of other quantities associated to elliptic curves in an analogous way.

2.2. **The Minimalist Conjecture.** The basic conjecture for the distribution of ranks of elliptic curves is based on the philosophy that elliptic curves should not have any more points than they must.

The widely believed Parity Conjecture, which is a consequence of a weak form of the BSD conjecture, asserts that the parity of the rank of an elliptic curve is equal to the parity of the analytic rank, which is even exactly when $u_E = +1$. It is also strongly expected that the root numbers u_E have probability $1/2$ of being $+1$, and probability $1/2$ of being -1. We are therefore led to the following:

Minimalist Conjecture. *The densities of elliptic curves having rank* 0 *and having rank* 1 *are both exactly* $1/2$.

Although no ordering is specified in the statement above, it is conjectured for any reasonable ordering, such as the examples given in §2.1. Note that the Minimalist Conjecture implies that the density of rank i elliptic curves, for $i \geq 2$, is 0.

The first version of the conjecture was stated in the 1979 work of Goldfeld [Gol79] for quadratic twist families of elliptic curves. Given an elliptic curve E in short Weierstrass form (3), define its quadratic twist E_d by a nonzero squarefree integer d as the elliptic curve $y^2 = x^3 + d^2Ax + d^3B$. Then Goldfeld's conjecture [Gol79] asserts that for a fixed elliptic curve E, the average rank of the elliptic curves E_d is $1/2$, when ordered by $|d|$, that is,

$$\lim_{X\to\infty} \frac{\sum_{|d|\leq X} \mathrm{rk}(E)}{\sum_{|d|\leq X} 1} = \frac{1}{2},$$

where the sums are over nonzero squarefree integers d. Much work has been done towards this conjecture for quadratic twist families; see Silverberg's survey [Sil07].

The Minimalist Conjecture for all elliptic curves and for quadratic twist families is also supported by the philosophy of Katz–Sarnak [KS99] and later random matrix theory computations and heuristics of Keating–Snaith [KS00], Conrey–Keating–Rubinstein–Snaith [CKRS02], Watkins [Wat08], and others. See [BMSW07, Poo12] for excellent surveys of many aspects of this conjecture.

At various points since Goldfeld's work, the conjecture has been disbelieved, mostly because computations have not seemed to support it. For example, in [KS99], the data of Kramarz–Zagier [ZK87] (extended later by Watkins [Wat07]) for a special family of elliptic curves is noted to have a large number of higher rank elliptic curves, with the suggestion that computational capabilities were not yet powerful enough to reflect the true distribution.

The more general computations for the family of all elliptic curves ordered by conductor, by Brumer–McGuinness [BM90], Stein–Watkins [SW02], Cremona [Cre06], and Bektemirov–Mazur–Stein–Watkins [BMSW07], also display a surprisingly large percentage of higher rank elliptic curves. As a result, their data imply asymptotics for average ranks that appear significantly higher than $1/2$; see Figure 4. It has

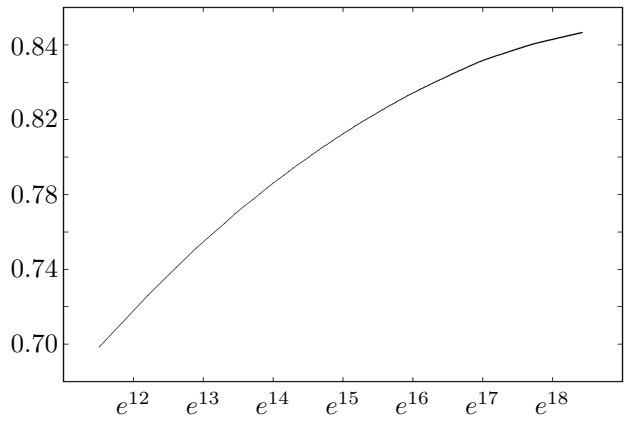


FIGURE 4. Average rank of elliptic curves in the Stein-Watkins database, up to conductor 10^8. Data and graph by Bektemirov–Mazur–Stein–Watkins [BMSW07].

been suggested that when restricted to these computational ranges, the data shows the strong effects of secondary terms for the asymptotic number of curves of given rank up to conductor X.

Theoretical work on this conjecture is perhaps more optimistic. Brumer [Bru92] showed that, assuming the Generalized Riemann Hypothesis (GRH), the limsup of the average analytic rank of all elliptic curves ordered by height is bounded above by 2.3. Thus, also assuming the BSD conjecture, Brumer's result implies that the limsup of the average rank is bounded above by 2.3. This bound was improved by Heath-Brown [HB04] to 2, and then by Young [You06] to 25/14, still assuming BSD and GRH. Young's bound was the first theoretical result implying that a positive proportion of elliptic curves have rank 0 or 1, assuming BSD and GRH.

Finally, the recent work of Bhargava–Shankar [BS10a] gives an unconditional upper bound:

Theorem 2.1 (Bhargava–Shankar 2010)**.** *The limsup of the average rank of elliptic curves over $\mathbb{Q}$, ordered by height, is bounded above by* 3/2*.*

They consider elliptic curves in short Weierstrass form with integral coefficients, and the theorem holds for both all such Weierstrass equations or only minimal ones. Much of the remainder of this note will focus on this theorem, as well as generalizations and corollaries; see also Poonen's Bourbaki exposé [Poo12] for an excellent detailed exposition of [BS10a].

2.3. **Selmer groups.** Studying the Selmer group for an elliptic curve is one of the only currently known methods to establish an upper bound on its rank. This method is used both for computations for individual curves (e.g., Cremona's `mwrank` program [Cre12]) and results for all curves together (as in Theorem 2.1).

The utility of Selmer groups comes from two facts: first, they are finite and often computable, and second, the size of the Selmer group of an elliptic curve gives an upper bound for its rank. More precisely, for a prime p, the p-Selmer group $\mathrm{Sel}_p(E)$ of an elliptic curve E is an elementary abelian p-group, i.e., isomorphic to the product of a nonnegative number of $\mathbb{Z}/p\mathbb{Z}$'s, and its p-rank $\mathrm{rk}_p(\mathrm{Sel}_p(E))$ is the number of factors of $\mathbb{Z}/p\mathbb{Z}$. Then

$$\mathrm{rk}_p(\mathrm{Sel}_p(E)) \geq \mathrm{rk}(E). \tag{5}$$

Bhargava–Shankar [BS10a] prove Theorem 2.1 by combining (5) for $p = 2$ and the following stronger result:

Theorem 2.2 (Bhargava–Shankar 2010)**.** *The average size of the* 2*-Selmer group for elliptic curves over $\mathbb{Q}$, ordered by height, is* 3*.*

Definitions. This subsection may be safely skipped at a first reading; it is more important to understand how to actually access elements of the p-Selmer group, as described in the next subsection.

We define the p-Selmer group for an elliptic curve E over $\mathbb{Q}$ and a prime p. One of the motivating ideas behind the definition is that local computations are often much more feasible than global ones; we also saw this idea in action in §1.1 during the discussion on genus zero curves.

Because the points of E over any field form a group, there is a multiplication-by-p map $E(\bar{\mathbb{Q}}) \xrightarrow{p} E(\bar{\mathbb{Q}})$, which is surjective and whose kernel is the p-torsion subgroup

$E(\bar{\mathbb{Q}})[p]$ of $E(\bar{\mathbb{Q}})$. The Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts on each of these groups, so the short exact sequence

$$0 \to E(\bar{\mathbb{Q}})[p] \to E(\bar{\mathbb{Q}}) \to E(\bar{\mathbb{Q}}) \to 0$$

of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-modules induces a long exact sequence of Galois cohomology, from which we extract the first row of the commutative diagram (6) below.

$$\begin{array}{ccccccccc} (6) \qquad 0 & \longrightarrow & E(\mathbb{Q})/pE(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q}, E[p]) & \xrightarrow{\alpha} & H^1(\mathbb{Q}, E)[p] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow{\scriptstyle \prod \mathrm{Res}_\nu} & \overset{\beta}{\searrow} & \downarrow{\scriptstyle \prod \mathrm{Res}_\nu} & & \\ & 0 \to & \prod_\nu E(\mathbb{Q}_\nu)/pE(\mathbb{Q}_\nu) & \to & \prod_\nu H^1(\mathbb{Q}_\nu, E[p]) & \to & \prod_\nu H^1(\mathbb{Q}_\nu, E)[p] & \to & 0 \end{array}$$

The analogous procedure over each local completion $\mathbb{Q}_\nu$ for primes ν (including $\mathbb{Q}_\nu := \mathbb{R}$ for $\nu = \infty$) gives the second row of (6). The leftmost vertical map is given by the inclusions of $E(\mathbb{Q})$ into each $E(\mathbb{Q}_\nu)$, and the latter two vertical maps are products over all primes ν of the usual restriction maps $\mathrm{Res}_\nu : H^1(\mathbb{Q}, A) \to H^1(\mathbb{Q}_\nu, A)$ for $A = E[p]$ and $A = E$. Then we make the following definitions:

(i) The **p-Selmer group** $\mathrm{Sel}_p(E)$ of E is the kernel of the map β in (6).
(ii) The **Tate–Shafarevich group** $Ш(E)$ is the kernel of the map

$$\prod_\nu \mathrm{Res}_\nu : H^1(\mathbb{Q}, E) \to \prod_\nu H^1(\mathbb{Q}_\nu, E).$$

Applying the Snake Lemma to a variant of (6) gives the key exact sequence

$$0 \to E(\mathbb{Q})/pE(\mathbb{Q}) \to \mathrm{Sel}_p(E) \to Ш(E)[p] \to 0, \tag{7}$$

which leads to the inequality (5).

Visualizing elements of Selmer groups. Elements of the Tate–Shafarevich group $Ш(E)$ and the p-Selmer group $\mathrm{Sel}_p(E)$ of an elliptic curve E over $\mathbb{Q}$ may be concretely realized using genus one curves and their Jacobians.

The **Jacobian** $\mathrm{Jac}(C)$ of a genus one curve C is the connected component of its automorphism group[5]. It is also a curve of genus one, and if C has a rational point, then the Jacobian of C is in fact isomorphic to C over $\mathbb{Q}$. As any automorphism group of course has a group structure, including an identity element, the Jacobian of C is an elliptic curve! We call a genus one curve C a **torsor**, or principal homogeneous space, for $\mathrm{Jac}(C)$. In other words, a torsor for an elliptic curve E is a genus one curve with an action of E that is simply transitive over $\bar{\mathbb{Q}}$.

The elements of $H^1(\mathbb{Q}, E)$ may be thought of as isomorphism classes of torsors for E, that is, genus one curves C over $\mathbb{Q}$ whose Jacobians are isomorphic to E. A trivial torsor is isomorphic to E itself, meaning that it has a rational point, and similarly, a torsor C that maps to 0 under Res_ν has a point in $\mathbb{Q}_\nu$.

Therefore, the elements of the Tate–Shafarevich group $Ш(E)$ are exactly those torsors, up to isomorphism, that have points over every local completion $\mathbb{Q}_\nu$, also known as **locally soluble**. The nonzero elements of $Ш(E)$ are locally soluble torsors without a global rational point, implying that they fail the Hasse principle.

[5]This definition only works for curves of genus one. More generally, the Jacobian is defined to be the dual of the moduli space of degree 0 line bundles.

Elements of the p-Selmer group may be represented as locally soluble torsors C for E, along with a degree p line bundle on C (or equivalently, a rational degree p divisor[6] on C). This degree p line bundle on C is equivalent to remembering an algebraic map from C to $(p-1)$-dimensional projective space. As we will describe in more detail in later sections, this description of p-Selmer elements may be made yet more explicit for small values of p.

Here is a table summarizing these interpretations of the groups, for an elliptic curve E over $\mathbb{Q}$:

Group	Elements (up to isomorphism)
$H^1(\mathbb{Q}, E)$	torsors C for E
$Ш(E)$	locally soluble torsors C for E
$\mathrm{Sel}_p(E)$	pairs (C, L): locally soluble torsors C for E with degree p line bundles L on C
$E(\mathbb{Q})/pE(\mathbb{Q})$	pairs (C, L): trivial(ized) torsors C for E with degree p line bundles L on C

Heuristics and other work. In the last several years, several heuristics have been developed for the distributions for the Tate–Shafarevich group and p-Selmer groups for elliptic curves over $\mathbb{Q}$ (and over other number fields).

Delaunay's heuristics [Del01, Del07] for the distribution of Tate–Shafarevich groups generalizes the ideas behind the Cohen–Lenstra–Martinet heuristics for class groups of number fields [CL84, CM87]. The main idea is that Tate–Shafarevich groups appear as random finite abelian groups with nondegenerate alternating bilinear pairings, weighted by the inverse of the size of the automorphism group.

The work of Poonen–Rains [PR12] models p-Selmer groups as intersections of random maximal isotropic subspaces in an infinite-dimensional quadratic space over $\mathbb{F}_p$. They obtain conjectural distributions for p-Selmer groups of elliptic curves (and abelian varieties). All of the currently known theoretical results on average sizes of Selmer groups, such as Theorem 2.2, agree with the Poonen–Rains heuristics.

Even more recently, Bhargava, Kane, Lenstra, Poonen, and Rains [BKL$^+$] have extended these heuristics to model both Selmer groups and Tate-Shafarevich groups simultaneously, by studying the distribution of the exact sequence that is the direct limit over n of (7) with p replaced by p^n.

There has also been recent progress in studying distributions of 2-Selmer groups for quadratic twist families, including work of Heath-Brown, Swinnerton-Dyer, Kane, Yu, Mazur, Rubin, and Klagsbrun, among many others [HB93, HB94, SD08, Kan12, Yu06, Yu05, MR10, KMR11].

Finally, a common theme in arithmetic geometry is to replace a number field by a function field, since geometry often helps in the latter case. Here, if $\mathbb{Q}$ is replaced by the function field $\mathbb{F}_q(t)$, de Jong [dJ02] gives an upper bound for the average size of 3-Selmer groups of elliptic curves over $\mathbb{F}_q(t)$. His methods for parametrizing elements of the Selmer group are similar to those described in §3.1 for Theorem 2.2 (and analogous results for Theorem 4.2).

[6]A rational degree p divisor on C is equivalent to a formal sum of p points of $C(\bar{\mathbb{Q}})$ which are together defined over $\mathbb{Q}$.

3. The average size of 2-Selmer groups

We now explain the main ideas behind Theorem 2.2 and the following stronger statement from [BS10a]:

Theorem 3.1 (Bhargava–Shankar 2010)**.** *Let $\mathscr{F}$ be any family of elliptic curves $E : y^2 = x^3 + Ax + B$ defined by finitely many congruence conditions on the integral coefficients A and B. Then the average size of $\mathrm{Sel}_2(E)$ for elliptic curves E in $\mathscr{F}$, ordered by height, is* 3.

The key observations are that binary quartic forms are closely related to elements of 2-Selmer groups of elliptic curves, and that it is possible to "count" integral binary quartic forms using techniques from the geometry of numbers.

More precisely, we will see in §3.1 that binary quartic forms with rational coefficients, up to standard transformations, with certain local properties correspond exactly to 2-Selmer elements of elliptic curves. The classical invariant theory of binary quartic forms plays a crucial role in this relationship; in particular, it gives the vertical map in diagram (8) below.

(8)
$$\begin{array}{ccc} \left\{\begin{array}{l}\text{2-Selmer elements}\\ \text{of elliptic curves } E\end{array}\right\} & \xrightarrow[\text{conditions}]{\text{local}} & \left\{\begin{array}{l}\text{binary quartic forms}\\ \text{up to equivalence}\end{array}\right\} \\ & \overset{\text{fiber over } E = \mathrm{Sel}_2(E)}{\rightsquigarrow} & \Big\downarrow \text{invariant theory} \\ & & \{\text{elliptic curves } E\} \end{array}$$

In §3.2, we explain how suitably enhanced techniques from the geometry of numbers are used to count the number of binary quartic forms with bounded height[7]. Incorporating the local conditions by using sieve methods (see §3.3) produces a count of 2-Selmer elements for elliptic curves up to a given height. Because the fiber of the squiggly arrow in diagram (8) above an elliptic curve E is exactly $\mathrm{Sel}_2(E)$, dividing this count by the number of elliptic curves up to the same height, and then taking the limit of that quotient as the height tends to infinity, gives the average we seek.

This method is not known to work if the elliptic curves are ordered by discriminant or by conductor, instead of by height; the asymptotic number of elliptic curves with discriminant or conductor less than X, as X tends to infinity, is not even known.

3.1. Binary quartic forms and elliptic curves.

In the classical work [BSD63] of Birch and Swinnerton-Dyer that inspired the BSD conjecture, they study and use the relationship between binary quartic forms and 2-Selmer elements of elliptic curves.

A **binary quartic form** over $\mathbb{Q}$ is a homogeneous polynomial of degree 4 in two variables with rational coefficients, e.g.,

$$f(x_1, x_2) := ax_1^4 + bx_1^3x_2 + cx_1^2x_2^2 + dx_1x_2^3 + ex_2^4 \tag{9}$$

[7]The height of a binary quartic form is the same height, up to a constant, as for its associated elliptic curve.

with $a, b, c, d, e \in \mathbb{Q}$. The set of all binary quartic forms over $\mathbb{Q}$ is a 5-dimensional $\mathbb{Q}$-vector space V, with coordinates given by the coefficients a, b, c, d, and e. The group $\mathrm{GL}_2(\mathbb{Q})$ acts on the elements of V via

$$g \cdot f(x_1, x_2) = (\det g)^{-2} f((x_1, x_2) \cdot g) \tag{10}$$

for all $g \in \mathrm{GL}_2(\mathbb{Q})$; since scalar matrices act trivially, this action induces an action of $\mathrm{PGL}_2(\mathbb{Q})$. We call two binary quartic forms f and f' **equivalent** if there exists $g \in \mathrm{PGL}_2(\mathbb{Q})$ and $\lambda \in \mathrm{GL}_1(\mathbb{Q}) = \mathbb{Q}^\times$ such that $f' = \lambda^2(g \cdot f)$. In other words, the space V is a certain representation of the group $\mathrm{PGL}_2(\mathbb{Q}) \times \mathrm{GL}_1(\mathbb{Q})$, and two binary quartic forms are equivalent if they are in the same orbit of the group.

Under the action (10) of $\mathrm{GL}_2(\mathbb{Q})$, or equivalently, under the induced action of $\mathrm{PGL}_2(\mathbb{Q})$, the invariants of a binary quartic form (9) form a polynomial ring generated by two invariants:

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

The **discriminant** $\Delta(f) := 4I(f)^3 - J(f)^2$ is nonzero exactly if f has four distinct roots over $\bar{\mathbb{Q}}$. The **height** of f is $\mathrm{ht}(f) := \max(|I(f)^3|, J(f)^2/4)$.

Genus one curves from binary quartic forms. Given a binary quartic form $f(x_1, x_2)$ with nonzero discriminant, one may construct a genus one curve $C(f)$ explicitly as the smooth compactification of the affine curve

$$y^2 = f(x_1, x_2).$$

This genus one curve is the double cover of the projective line ramified at exactly the roots of f (which may not be individually defined over $\mathbb{Q}$). It therefore comes equipped with a degree 2 line bundle $L(f)$, namely the pullback of the line bundle $\mathcal{O}(1)$ from $\mathbb{P}^1$; equivalently, a rational degree 2 divisor on $C(f)$ is given by the formal sum of the two points in the preimage of any rational point on $\mathbb{P}^1$ under this double cover.

If f' is an equivalent binary quartic form, then $C(f')$ and $C(f)$ are isomorphic, and the line bundles for each also correspond to one another under this isomorphism. In fact, binary quartic forms with nonzero discriminant up to equivalence are exactly in one-to-one correspondence with isomorphism classes of genus one curves with degree 2 line bundles!

Moreover, the Jacobian $E(f)$ of $C(f)$ depends only on the two invariants $I(f)$ and $J(f)$; it may be written in short Weierstrass form as

$$E(f) : y^2 = x^3 - \frac{I(f)}{3}x - \frac{J(f)}{27}. \tag{11}$$

Therefore, from our visualization of 2-Selmer elements described in §2.3, we see that for a binary quartic form f, if $C(f)$ is locally soluble, then the pair $(C(f), L(f))$ corresponds to an element of $\mathrm{Sel}_2(E(f))$. More precisely, let $V(\mathbb{Q})^{\mathrm{ls}}$ be the subset of locally soluble binary quartic forms $f(x_1, x_2)$ over $\mathbb{Q}$ with $\Delta(f) \neq 0$, i.e., those for which $y^2 = f(x_1, x_2)$ has a $\mathbb{Q}_\nu$-solution for all primes ν (including $\mathbb{Q}_\infty = \mathbb{R}$). Note that $V(\mathbb{Q})^{\mathrm{ls}}$ is preserved under the action of $\mathrm{GL}_2(\mathbb{Q}) \times \mathbb{Q}^\times$.

The equivalence classes of $V(\mathbb{Q})^{\mathrm{ls}}$ are in correspondence with 2-Selmer elements of elliptic curves; that is, we have the bijection

$$\mathrm{PGL}_2(\mathbb{Q}) \times \mathbb{Q}^\times \backslash V(\mathbb{Q})^{\mathrm{ls}} \overset{1-1}{\longleftrightarrow} \left\{ (E, \zeta) : \begin{matrix} E \text{ elliptic curve} \\ \zeta \in \mathrm{Sel}_2(E) \end{matrix} \right\} / \cong .$$

For any specific elliptic curve $E_{AB} : y^2 = x^3 + Ax + B$, we may specialize to the correspondence

$$\mathrm{PGL}_2(\mathbb{Q}) \backslash V_{AB}(\mathbb{Q})^{\mathrm{ls}} \overset{1-1}{\longleftrightarrow} \mathrm{Sel}_2(E_{AB}),$$

where $V_{AB}(\mathbb{Q})^{\mathrm{ls}}$ consists of binary quartic forms f with invariants $I(f) = -3A$ and $J(f) = -27B$.

Finding binary quartic forms with specified invariants is the best known way to explicitly compute the 2-Selmer group (and often, the rank) for a given elliptic curve; see, e.g., Cremona's `mwrank` program [Cre12].

Example 3.2. The only rational binary quartic forms, up to the action of $\mathrm{PGL}_2(\mathbb{Q})$, with invariants $I = 48$ and $J = -432$ are $f_0 = x_1^4 - 6x_1^2x_2^2 + 4x_1x_2^3 + x_2^4$ and $f_1 = x_1^4 + 4x_1x_2^3 + 4x_2^4$. They each have Jacobian isomorphic to the elliptic curve E given by $y^2 = x^3 - 16x + 16$. Thus,

$$\mathrm{Sel}_2(E) \cong \mathbb{Z}/2\mathbb{Z},$$

with f_0 representing the identity element. In this case, because E has at least one rational point $(x, y) = (0, 4)$ and $E(\mathbb{Q})_{\mathrm{tors}}$ is trivial, the sequence (7) implies that $\mathrm{rk}(E) = 1$ and $Ш(E)[2] = 0$.

In order to find the average size of the 2-Selmer group, the goal is therefore to count the number of equivalence classes in $V(\mathbb{Q})^{\mathrm{ls}}$ up to bounded height. The first step is to simply count the number of $\mathrm{PGL}_2(\mathbb{Z})$-equivalence classes of binary quartic forms with *integral* coefficients.

3.2. **Counting binary quartic forms using the geometry of numbers.** Methods from the geometry of numbers have been previously successful in similar counting questions, such as determining the number of equivalence classes of binary quadratic and binary cubic forms [Mer74, Sie44, Dav51b, Dav51c]. Bhargava–Shankar give an asymptotic count of the number of irreducible integral binary quartic forms, up to equivalence, of bounded height:

Theorem 3.3 ([BS10a])**.** *For $0 \le i \le 2$, let $N^{(i)}(X)$ be the number of $\mathrm{PGL}_2(\mathbb{Z})$-equivalence classes of irreducible integral binary quartic forms having $4 - 2i$ real roots and height less than X. Then*

$$N^{(0)}(X) = \frac{4}{135}\zeta(2)X^{5/6} + O(X^{3/4+\epsilon})$$
$$N^{(1)}(X) = \frac{32}{135}\zeta(2)X^{5/6} + O(X^{3/4+\epsilon})$$
and $$N^{(2)}(X) = \frac{8}{135}\zeta(2)X^{5/6} + O(X^{3/4+\epsilon}).$$

One may also impose finitely many congruence conditions on the coefficients a, b, c, d, and e of the binary quartic forms, e.g., requiring a to be 0 modulo p for a prime p. Then the number of equivalence classes of such integral binary quartic forms with height bounded by X is the total number of equivalence classes (the appropriate $N^{(i)}(X)$ from Theorem 3.3) multiplied by the p-adic density of each congruence condition imposed, with the same error term of $O(X^{3/4+\epsilon})$. This p-adic density is an easily computable fraction depending on p.

Remark 3.4. Although the exact statement of Theorem 3.3 is not strictly necessary for the proof of Theorem 3.1, the ideas and results used in the proof of Theorem 3.3 are essentially a subset of those needed for the average Selmer result.

The main idea in proving Theorem 3.3 is to reduce the question to counting lattice points in a nicely shaped domain, in which case the number of lattice points is approximately the volume of the domain. The major complication arises when the domain has cusps, which may be visualized as thin regions going off to infinity. A priori, these cuspidal regions may contain many or few integral points; see Figure 5. A clever "averaging" technique — first introduced by Bhargava in [Bha05, Bha10] for asymptotic counts of quartic and quintic rings — helps control exactly which points lie in the cusps.

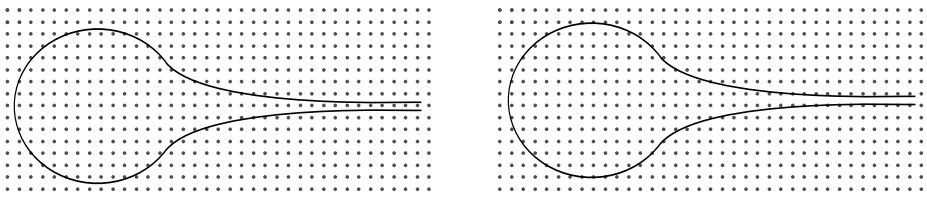


FIGURE 5. A domain with many lattice points in the cusp (left), and a domain with few lattice points in the cusp (right).

Let $V^{(i)}$ denote the subset of V corresponding to binary quartic forms with $4-2i$ roots. For the remainder of this section, we will focus on the case where the binary quartic forms have 4 real roots (that is, when $i = 0$); the other two cases are similar.

Reduction theory and fundamental domains. A **fundamental domain** or **set** for a group acting on a space is a set of elements in the space containing exactly one representative for each orbit. To count $\mathrm{PGL}_2(\mathbb{Z})$-orbits of the space $V(\mathbb{Z})$ of integral binary quartic forms, we may try to count lattice (integral) points in a fundamental domain for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $V(\mathbb{R})$. It is easier to break up this latter action into two intermediate ones, splitting the problem into two steps:

(i) find a fundamental set for the action of $\mathrm{PGL}_2(\mathbb{R}) \times \mathbb{R}^\times$ on $V(\mathbb{R})$, and
(ii) find a fundamental domain for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\mathrm{PGL}_2(\mathbb{R}) \times \mathbb{R}^\times$.

For (i), such a fundamental set is easy to explicitly construct. One checks that a binary quartic form with all real roots and invariants I and J defines a unique $\mathrm{PGL}_2(\mathbb{R})$-orbit in $V(\mathbb{R})$ with those invariants. Thus, a fundamental set L consists of real binary quartic forms whose invariants range over all I and J, up to scaling (because of the action of $\mathbb{R}^\times$). Note that for any $h \in \mathrm{PGL}_2(\mathbb{R}) \times \mathbb{R}^\times$, the set hL is also a fundamental set. It is crucial that we may choose L such that hL is always a compact set.

Example 3.5. In fact, we may choose representatives for a fundamental set with height 1. One fundamental set for $V^{(0)}(\mathbb{R})$ is

$$L = \left\{ f_t(x_1, x_2) = x_1^3 x_2 - \frac{1}{3} x_1 x_2^3 - \frac{t}{27} x_2^4 : -2 \le t \le 2 \right\},$$

where $I(f_t) = 1$, $J(f_t) = t$, and discriminant $\Delta(f_t) = 4 - t^2 > 0$.

For (ii), there is a standard decomposition, due to Gauss, of a fundamental domain $\mathcal{F}$ for $\mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})\times\mathbb{R}^\times$. This description also gives explicit coordinates for $\mathcal{F}$.

Combining (i) and (ii) shows that the set $\mathcal{F}hL$, for any $h\in\mathrm{PGL}_2(\mathbb{R})\times\mathbb{R}^\times$, contains a representative from each $\mathrm{PGL}_2(\mathbb{Z})$-orbit of $V^{(0)}(\mathrm{R})$. In fact, when viewed as a multiset, $\mathcal{F}hL$ overcounts each orbit — by the size of the stabilizer in $\mathrm{PGL}_2(\mathbb{R})$ of the binary quartic divided by the size of its stabilizer in $\mathrm{PGL}_2(\mathbb{Z})$. For binary quartics in $V^{(0)}(\mathbb{R})$, this quotient is $4/1=4$ almost always (in a sense that may be made precise), so it suffices to assume that each orbit is counted four times. In other words, the set $\mathcal{F}hL$ is (almost) a union of four fundamental domains for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $V^{(0)}(\mathbb{R})$.

We are now interested in counting the number of integer points in $\mathcal{F}hL$ of bounded height (and dividing by 4).

Averaging and volumes. As alluded to earlier, the number of lattice points in a domain like $\mathcal{F}hL$ is essentially the volume of the domain, by ideas of Minkowski and refinements by Davenport [Dav51a, Dav64], but one needs to control the points in the cusp of this domain.

In order to thicken the cusp for better control, we take not just a single domain $\mathcal{F}hL$, but a small ball's worth of such domains by letting the element h vary in a compact set. To obtain the final answer, the number of lattice points in the union of these domains $\mathcal{F}hL$, counted with multiplicity, must be divided by the volume of this compact set.

This larger domain with a thicker cusp may be split into two parts, the main body and the cusp; a clever choice of where to exactly separate the two will give the desired estimates. In particular, let the cusp be the part of the fundamental domain containing the binary quartic forms $f(x_1,x_2)$ from (9) for which the absolute value of the coefficient a of x_1^4 is strictly less than 1. Then any integral binary quartic form in the cusp has a equal to 0 and hence is reducible!

The volume of the main body then approximates the number of lattice points in it, and one may show that it contains a negligible number of reducible binary quartic forms.

Remark 3.6. Theorem 3.3 only concerns irreducible binary quartic forms, but when we return in §3.3 to counting binary quartic forms corresponding to 2-Selmer elements, we will include the reducible binary quartic forms found in the cusp.

The final step is to compute the volume of this main body, which may be done explicitly. A critical lemma in this computation involves changing from the standard Euclidean measure on the space $V(\mathbb{R})$ to the product of the Haar measure on the group $\mathrm{PGL}_2(\mathbb{R})$ and the measures given by the invariants I and J. This Jacobian computation mirrors the intuitive idea that $V(\mathbb{R})$ is roughly a product of $\mathrm{PGL}_2(\mathbb{R})$ and the quotient $\mathrm{PGL}_2(\mathbb{R})\backslash V(\mathbb{R})$.

3.3. **Sieves and uniformity estimates.** For Theorems 2.2 and 3.1, the relevant count is for *rational* equivalence classes of binary quartic forms corresponding to *locally soluble* genus one curves. Thus, we need to add several steps to the ideas from §3.2:

(a) As mentioned in Remark 3.6, the reducible binary quartic forms in the cusp must be incorporated.

(b) Find an integral representative for each $\mathrm{PGL}_2(\mathbb{Q}) \times \mathbb{Q}^\times$-orbit of $V(\mathbb{Q})^{\mathrm{ls}}$ with integral[8] invariants (and determine exactly how many each rational orbit contains).
(c) Impose the necessary local conditions — via sieve methods — to restrict to the space $V(\mathbb{Q})^{\mathrm{ls}}$ of locally soluble binary quartic forms.

Part (a) is important but straightforward. As mentioned earlier, the cusp region contains binary quartic forms that have a linear factor; these exactly correspond to the identity elements in the 2-Selmer groups! As these and other reducible forms do not appear often in the main body, the main body counts only irreducible binary quartic forms, corresponding to non-identity elements of the Selmer groups.

The first part of (b) is a standard fact in this case [BSD63, CFS10]; it is essentially a local computation. That is, given a rational binary quartic form f in $V(\mathbb{Q})^{\mathrm{ls}}$ with integral invariants, then for all primes p, there exists an element $g_p \in \mathrm{PGL}_2(\mathbb{Q}_p)$ such that the binary quartic form $g_p \cdot f$ has coefficients in $\mathbb{Z}_p$. Then we may use the idea of weak approximation to "glue" together all of these g_p's into an element $g \in \mathrm{PGL}_2(\mathbb{Q})$, as PGL_2 has class number one; the binary quartic form $g \cdot f$ then has integral coefficients.

In general, however, the orbit of such an $f \in V(\mathbb{Q})^{\mathrm{ls}}$ may contain many $\mathrm{PGL}_2(\mathbb{Z})$-orbits, so we need to weight each integral orbit by $1/n$, where n is the number of integral orbits for that rational orbit. This weighting may in fact be incorporated into the sieve for part (c). Again using the fact that the group PGL_2 has class number one, this last step is a local computation; the global weight is a product of local weights, which are related to the size of the stabilizers of the binary quartic forms in $\mathrm{PGL}_2(\mathbb{Q}_p)$.

This "geometric sieve," originating in work of Ekedahl [Eke91] and extended by Poonen [Poo03, Poo04] and Bhargava [Bha11], is the final step. Imposing *finitely* many congruence conditions on the binary quartic forms translates into multiplying the original count by the local densities for each condition, as mentioned after Theorem 3.3. However, we now need to impose a condition for every prime p, so to obtain an actual limit (as opposed to only a limsup), a certain *uniformity estimate* is needed.[9] In particular, one shows that the binary quartic forms that are "bad" at a prime p are rare as p approaches infinity, so they may be safely ignored.

In the end, the product of all these local factors simplifies[10] to be an invariant of the group PGL_2, called the **Tamagawa number** $\tau(\mathrm{PGL}_2)$. In other words, the limit as $X \to \infty$ of the weighted number of irreducible integral binary quartic forms in $V(\mathbb{Z})^{\mathrm{ls}}$ with height $< X$, divided by the number of elliptic curves of height $< X$, is $\tau(\mathrm{PGL}_2) = 2$. For the average for a family $\mathscr{F}$ of elliptic curves defined by finitely many congruence conditions, the local factors would affect the numerator and denominator equally, so the (limit of the) quotient would not change.

Finally, adding in the cusp contribution (for the identity elements in the 2-Selmer groups) implies the average size of the 2-Selmer group is

$$2 + 1 = 3.$$

[8]For simplicity, we are ignoring some factors of 2 and 3 throughout the discussion of this part.

[9]In the original paper [BS10a], obtaining this uniformity estimate is the most difficult and technical part, but the refined geometric sieve in [Bha11] significantly simplifies the computation needed here.

[10]See also [Poo12] for an explanation of this fact by computing an adelic volume instead.

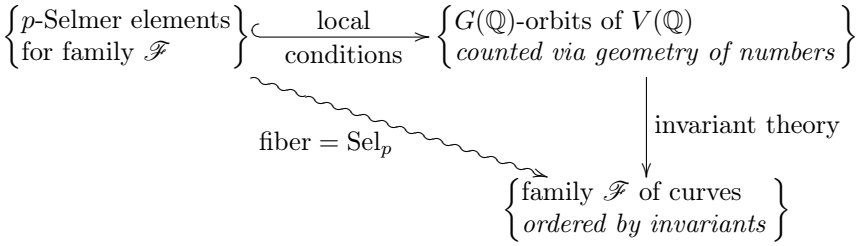
4. Generalizations and corollaries

We now outline generalizations of Theorem 2.2 and the methods discussed in §3 to other p-Selmer groups, other families of elliptic curves, and even families of higher genus curves. In §4.2, we also explain some corollaries for densities of low rank elliptic curves.

The strategy for proving Theorem 2.2 presented in §3 relies heavily on a description of 2-Selmer elements as equivalence classes of binary quartic forms with certain local properties. The geometry-of-numbers techniques apply to the situation after reducing the question to counting lattice points in a fundamental domain for the action of a group on a vector space.

Generalizing these methods thus depends on relating elements of Selmer groups to the orbits of a vector space V under the action of a group G; these orbits may then be counted as before. We modify diagram (8) to reflect the more general goal:

(12)



The family $\mathscr{F}$ for Theorem 2.2 consists of elliptic curves in short Weierstrass form. More generally, one may choose $\mathscr{F}$ to be other families of elliptic curves or even higher genus curves, whose Jacobians have analogously defined p-Selmer groups.

Finding appropriate groups G and vector spaces V related to the Selmer elements is still a relatively ad hoc process. For elliptic curves, we generally use the geometric description of elements of p-Selmer groups — as locally soluble torsors with degree p line bundles — to find such G and V.

Remark 4.1. The method summarized in diagram (12) was also previously used by Davenport–Heilbronn [DH69] and Bhargava [Bha05] to prove two of the only known cases of the Cohen-Lenstra-Martinet heuristics on distributions of ideal class groups of number fields. In those cases, the family $\mathscr{F}$ is replaced by a family of number fields (quadratic fields and cubic fields, respectively, ordered by discriminant), and the p-torsion of the ideal class group is the analogue of the p-Selmer group.

4.1. **Other Selmer groups for elliptic curves.** We survey recent results on average sizes of Selmer groups for elliptic curves; the methods behind these theorems all arise from the ideas highlighted in diagram (12).

In [BS10b], Bhargava–Shankar extend their methods from [BS10a] to 3-Selmer groups, by using the classical description of 3-Selmer elements as locally soluble curves cut out by ternary cubic forms, up to equivalence.

Theorem 4.2 (Bhargava–Shankar 2010)**.** *The average size of the* 3*-Selmer group for elliptic curves over* $\mathbb{Q}$*, ordered by height, is* 4*.*

The average for 3-Selmer groups gives an improved upper bound of 7/6 for the limsup of the average rank of elliptic curves.

In fact, Bhargava–Shankar have work in progress, using similar methods, to show that the average size of the 4- and 5-Selmer groups for elliptic curves, ordered by height, is 7 and 6, respectively. With some additional work, they are able to use these averages to show that the limsup of the average rank is in fact bounded above by 0.89.

In joint work with Bhargava [BH12b], we find the average sizes of 2- and/or 3-Selmer groups for various families of elliptic curves, such as the family

$$\mathscr{F}_1 := \left\{y^2 + a_3y = x^3 + a_2x^2 + a_4x : a_2, a_3, a_4 \in \mathbb{Z}, \Delta \neq 0\right\} \tag{13}$$

of elliptic curves with one marked point, ordered by analogous notions of height. These averages rely on explicit descriptions of Selmer elements for these families as orbits of certain representations [BH12a]. Upper bounds on average ranks of elliptic curves in these families are also obtained in the same way.

For all the families considered in [BH12b], we find that the marked points on the elliptic curves essentially act independently. For example, for the family $\mathscr{F}_1$, independence would imply that the single marked point should increase the p-rank of the p-Selmer group by 1, and indeed, the 2- and 3-Selmer groups have average sizes $3 \cdot 2 = 6$ and $4 \cdot 3 = 12$, respectively.

4.2. **Lots of rank 0 and rank 1 curves.** Using the average size of 3-Selmer groups, one may deduce the existence of many elliptic curves with rank 0 and as a result, many for which the BSD conjecture is true!

Dokchitser–Dokchitser [DD10] prove the **p-parity conjecture** over $\mathbb{Q}$, which states that the root number of an elliptic curve over $\mathbb{Q}$ is determined by the parity of its p-Selmer rank. By using congruence conditions to construct a positive-density family of elliptic curves with equidistributed root number, Bhargava–Shankar combine the p-parity conjecture with Theorem 4.2 to show that a positive density[11] of all elliptic curves, when ordered by height, have rank exactly 0.

In addition, applying Skinner–Urban's results [SU06, SU10] on the main conjecture of Iwasawa theory for GL_2 shows that a positive proportion of elliptic curves have *analytic* rank exactly 0. Since the BSD conjecture is known for curves of analytic rank 0 by Kolyvagin's work [Kol88], Bhargava–Shankar conclude that a positive proportion of elliptic curves over $\mathbb{Q}$ satisfy the BSD conjecture.

Moreover, with the assumption that $Ш(E)$ for any elliptic curve E is finite (or the weaker assumption that the 3-torsion subgroup $Ш(E)[3]$ is always a square), they find a positive density of elliptic curves with rank 1.

Similar results hold for the family $\mathscr{F}_1$ from (13): a positive density of elliptic curves in $\mathscr{F}_1$ have rank 1, and conditional on the finiteness of Ш, a positive density have rank 2.

4.3. **Higher genus curves.** As mentioned in §1.1, while curves of genus at least 2 have finitely many rational points, determining the number of rational points is still quite difficult. Given an ordering of curves in a particular family, one may ask similar questions as for elliptic curves, e.g., for any finite number N, what is the density of curves with N points? What is the average number of rational points, if finite?

[11] All of these statements on having a positive density or proportion of curves with a given property are, more precisely, about the lower density of such curves being positive.

The techniques discussed in §3, surprisingly, give some results towards these questions, at least for hyperelliptic curves with rational Weierstrass points. Bhargava–Gross [BG12a] first find a description of the 2-Selmer elements here using rational orbits of a certain representation of odd orthogonal groups (see also work of Thorne [Tho12] for more such parametrizations for higher genus curves using Lie theory). They then compute the average size of the 2-Selmer group:

Theorem 4.3 (Bhargava–Gross 2012)**.** *Fix $g \geq 1$. Then the average size of the 2-Selmer group for Jacobians of genus g hyperelliptic curves over $\mathbb{Q}$ with a rational Weierstrass point, ordered by height, is 3.*

Not only does this give an upper bound of 3/2 for the limsup of the average Mordell-Weil rank of the Jacobians of such curves, but it also may be used — along with the method of Chabauty and Coleman [Cha41, Col85] —- to show that there are many curves with very few points. Poonen and Stoll [PS] have recently improved upon the results from [BG12a] of this type:

Corollary 4.4 (Poonen–Stoll 2012)**.** *Fix $g \geq 3$. Then a positive proportion of genus g hyperelliptic curves over $\mathbb{Q}$ with a rational Weierstrass point have no other rational points, and a majority of such curves have at most 7 rational points.*

In fact, Poonen–Stoll show that as the genus g tends to infinity, the lower density of these curves for which the given Weierstrass point is the only rational point tends to 1.

Finally, Bhargava and Gross [BG12b] have very recently shown, using methods analogous to [BG12a], that most hyperelliptic curves over $\mathbb{Q}$ have zero rational points!

$$* \sim * \sim *$$

Acknowledgments. We thank Manjul Bhargava, Bhargav Bhatt, Arul Shankar, and Brian Street for reading and commenting on previous drafts. We thank Simon Spicer for his help with collecting data for ranks and Selmer groups, and we thank Nick Katz for suggesting relevant references. All computations were done with SAGE [S$^+$12].

References

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* **Q***: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

[BG12a] Manjul Bhargava and Benedict H. Gross, *The average size of the 2-Selmer group of jacobians of hyperelliptic curves having a rational Weierstrass point*, 2012, `http://arxiv.org/abs/1208.1007`.

[BG12b] ———, *Most hyperelliptic curves over $\mathbb{Q}$ have no rational points*, in preparation, 2012.

[BH12a] Manjul Bhargava and Wei Ho, *Coregular spaces and genus one curves*, preprint, 2012.

[BH12b] ———, *On the average sizes of Selmer groups in families of elliptic curves*, preprint, 2012.

[Bha05] Manjul Bhargava, *The density of discriminants of quartic rings and fields*, Ann. of Math. (2) **162** (2005), no. 2, 1031–1063.

[Bha10] ———, *The density of discriminants of quintic rings and fields*, Ann. of Math. (2) **172** (2010), no. 3, 1559–1591.

[Bha11] ———, *The geometric squarefree sieve and unramified nonabelian extensions of quadratic fields*, preprint, 2011.

[BKL+] Manjul Bhargava, Daniel Kane, Hendrik Lenstra, Bjorn Poonen, and Eric Rains, *Modeling the distribution of Selmer groups, Shafarevich–Tate groups, and ranks of elliptic curves*, in preparation. Extended abstract available in pp. 45–48 of `http://www.mfo.de/document/1232/OWR_2012_38.pdf`.

[BM90] Armand Brumer and Oisín McGuinness, *The behavior of the Mordell-Weil group of elliptic curves*, Bull. Amer. Math. Soc. (N.S.) **23** (1990), no. 2, 375–382.

[BMSW07] Baur Bektemirov, Barry Mazur, William Stein, and Mark Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 2, 233–254.

[Bom90] Enrico Bombieri, *The Mordell conjecture revisited*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **17** (1990), no. 4, 615–640.

[Bru92] Armand Brumer, *The average rank of elliptic curves. I*, Invent. Math. **109** (1992), no. 3, 445–472.

[BS10a] Manjul Bhargava and Arul Shankar, *Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves*, 2010, `http://arxiv.org/abs/1006.1002`.

[BS10b] ———, *Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0*, 2010, `http://arxiv.org/abs/1007.0052`.

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

[BSD65] ———, *Notes on elliptic curves. II*, J. Reine Angew. Math. **218** (1965), 79–108.

[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.

[Cha41] Claude Chabauty, *Sur les points rationnels des courbes algébriques de genre supérieur à l'unité*, C. R. Acad. Sci. Paris **212** (1941), 882–885.

[CKRS02] J. B. Conrey, J. P. Keating, M. O. Rubinstein, and N. C. Snaith, *On the frequency of vanishing of quadratic twists of modular L-functions*, Number theory for the millennium, I (Urbana, IL, 2000), A K Peters, Natick, MA, 2002, pp. 301–315.

[CL84] H. Cohen and H. W. Lenstra, Jr., *Heuristics on class groups of number fields*, Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983), Lecture Notes in Math., vol. 1068, Springer, Berlin, 1984, pp. 33–62.

[CM87] H. Cohen and J. Martinet, *Class groups of number fields: numerical heuristics*, Math. Comp. **48** (1987), no. 177, 123–137.

[Col85] Robert F. Coleman, *Effective Chabauty*, Duke Math. J. **52** (1985), no. 3, 765–770.

[Cre06] John Cremona, *The elliptic curve database for conductors to 130000*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 4076, Springer, Berlin, 2006, pp. 11–29.

[Cre12] John Cremona, ***mwrank*** *program*, 2012, `http://homepages.warwick.ac.uk/~masgaj/mwrank/`.

[Dav51a] H. Davenport, *On a principle of Lipschitz*, J. London Math. Soc. **26** (1951), 179–183.

[Dav51b] ———, *On the class-number of binary cubic forms. I*, J. London Math. Soc. **26** (1951), 183–192.

[Dav51c] ———, *On the class-number of binary cubic forms. II*, J. London Math. Soc. **26** (1951), 192–198.

[Dav64] ———, *Corrigendum: "On a principle of Lipschitz"*, J. London Math. Soc. **39** (1964), 580.

[DD10] Tim Dokchitser and Vladimir Dokchitser, *On the Birch-Swinnerton-Dyer quotients modulo squares*, Ann. of Math. (2) **172** (2010), no. 1, 567–596.

[Del01] Christophe Delaunay, *Heuristics on Tate-Shafarevitch groups of elliptic curves defined over* $\mathbb{Q}$, Experiment. Math. **10** (2001), no. 2, 191–196.

[Del07] ———, *Heuristics on class groups and on Tate-Shafarevich groups: the magic of the Cohen-Lenstra heuristics*, Ranks of elliptic curves and random matrix theory, London Math. Soc. Lecture Note Ser., vol. 341, Cambridge Univ. Press, Cambridge, 2007, pp. 323–340.

[DH69] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic fields*, Bull. London Math. Soc. **1** (1969), 345–348.

[dJ02] A. J. de Jong, *Counting elliptic surfaces over finite fields*, Mosc. Math. J. **2** (2002), no. 2, 281–311, Dedicated to Yuri I. Manin on the occasion of his 65th birthday.

[Eke91] Torsten Ekedahl, *An infinite version of the Chinese remainder theorem*, Comment. Math. Univ. St. Paul. **40** (1991), no. 1, 53–59.

[Fal83] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366.

[Fal91] Gerd Faltings, *Diophantine approximation on abelian varieties*, Ann. of Math. (2) **133** (1991), no. 3, 549–576.

[Gol79] Dorian Goldfeld, *Conjectures on elliptic curves over quadratic fields*, Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979), Lecture Notes in Math., vol. 751, Springer, Berlin, 1979, pp. 108–118.

[HB93] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem*, Invent. Math. **111** (1993), no. 1, 171–195.

[HB94] ———, *The size of Selmer groups for the congruent number problem. II*, Invent. Math. **118** (1994), no. 2, 331–370, With an appendix by P. Monsky.

[HB04] ———, *The average analytic rank of elliptic curves*, Duke Math. J. **122** (2004), no. 3, 591–623.

[Kan12] Daniel M. Kane, *On the ranks of the 2-Selmer groups of twists of a given elliptic curve*, 2012, `http://arxiv.org/abs/1009.1365`.

[KMR11] Zev Klagsbrun, Barry Mazur, and Karl Rubin, *Selmer ranks of quadratic twists of elliptic curves*, 2011, `http://arxiv.org/abs/1111.2321`.

[Kol88] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $Ш(E, \mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671.

[KS99] Nicholas M. Katz and Peter Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, vol. 45, American Mathematical Society, Providence, RI, 1999.

[KS00] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(1/2+it)$*, Comm. Math. Phys. **214** (2000), no. 1, 57–89.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186.

[Mer74] F. Mertens, *Ueber einige asymptotische Gesetze der Zahlentheorie*, J. Reine Angew. Math. **77** (1874), 289–338.

[Mor22] Louis J. Mordell, *On the rational solutions of the indeterminate equation of the third and fourth degrees*, Proc. Cambridge Philos. Soc. **21** (1922), 179–192.

[MR10] Barry Mazur and Karl Rubin, *Ranks of twists of elliptic curves and Hilbert's tenth problem*, Invent. Math. **181** (2010), no. 3, 541–575.

[Poo03] Bjorn Poonen, *Squarefree values of multivariable polynomials*, Duke Math. J. **118** (2003), no. 2, 353–373.

[Poo04] ———, *Bertini theorems over finite fields*, Ann. of Math. (2) **160** (2004), no. 3, 1099–1127.

[Poo12] ———, *Average rank of elliptic curves*, Séminaire Bourbaki, 2011-2012, 64ème année no. 1049.

[PR12] Bjorn Poonen and Eric Rains, *Random maximal isotropic subspaces and Selmer groups*, J. Amer. Math. Soc. **25** (2012), no. 1, 245–269.

[PS] Bjorn Poonen and Michael Stoll, *Chabauty's method proves that most odd degree hyperelliptic curves have only one rational point*, in preparation.

[S$^+$12] W. A. Stein et al., *Sage Mathematics Software (Version 5.3)*, The Sage Development Team, 2012, `http://www.sagemath.org`.

[SD08] Peter Swinnerton-Dyer, *The effect of twisting on the 2-Selmer group*, Math. Proc. Cambridge Philos. Soc. **145** (2008), no. 3, 513–526.

[Sie44] Carl Ludwig Siegel, *The average measure of quadratic forms with given determinant and signature*, Ann. of Math. (2) **45** (1944), 667–685.

[Sie66] ———, *Über einige Anwendungen diophantischer Approximationen (1929)*, Gesammelte Abhandlungen. Bände I, II, III, Springer-Verlag, 1966, pp. 209–266.

[Sil07] A. Silverberg, *The distribution of ranks in families of quadratic twists of elliptic curves*, Ranks of elliptic curves and random matrix theory, London Math. Soc. Lecture Note Ser., vol. 341, Cambridge Univ. Press, Cambridge, 2007, pp. 171–176.

[SU06] Christopher Skinner and Eric Urban, *Vanishing of L-functions and ranks of Selmer groups*, International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 473–500.

[SU10] Christopher Skinner and Eric Urban, *The Iwasawa main conjectures for GL(2)*, preprint. Available at `http://www.math.columbia.edu/~urban/eurp/MC.pdf`, 2010.

[SW02] William A. Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, pp. 267–275.

[Tho12] Jack Thorne, *The arithmetic of simple singularities*, Ph.D. thesis, Harvard University, 2012.

[TW95] Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[Voj91] Paul Vojta, *Siegel's theorem in the compact case*, Ann. of Math. (2) **133** (1991), no. 3, 509–548.

[Wat07] Mark Watkins, *Rank distribution in a family of cubic twists*, Ranks of elliptic curves and random matrix theory, London Math. Soc. Lecture Note Ser., vol. 341, Cambridge Univ. Press, Cambridge, 2007, pp. 237–246.

[Wat08] ———, *Some heuristics about elliptic curves*, Experiment. Math. **17** (2008), no. 1, 105–125.

[Wil95] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[You06] Matthew P. Young, *Low-lying zeros of families of elliptic curves*, J. Amer. Math. Soc. **19** (2006), no. 1, 205–250.

[Yu05] Gang Yu, *Average size of 2-Selmer groups of elliptic curves. II*, Acta Arith. **117** (2005), no. 1, 1–33.

[Yu06] ———, *Average size of 2-Selmer groups of elliptic curves. I*, Trans. Amer. Math. Soc. **358** (2006), no. 4, 1563–1584 (electronic).

[ZK87] D. Zagier and G. Kramarz, *Numerical investigations related to the L-series of certain elliptic curves*, J. Indian Math. Soc. (N.S.) **52** (1987), 51–69 (1988).

Department of Mathematics, Columbia University, New York, NY 10027
E-mail address: `who@math.columbia.edu`