

# Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks

Jennifer S. Balakrishnan, Wei Ho, Nathan Kaplan, Simon Spicer,  
William Stein and James Weigandt

## ABSTRACT

Most systematic tables of data associated to ranks of elliptic curves order the curves by conductor. Recent developments, led by work of Bhargava and Shankar studying the average sizes of  $n$ -Selmer groups, have given new upper bounds on the average algebraic rank in families of elliptic curves over  $\mathbb{Q}$ , ordered by height. We describe databases of elliptic curves over  $\mathbb{Q}$ , ordered by height, in which we compute ranks and 2-Selmer group sizes, the distributions of which may also be compared to these theoretical results. A striking new phenomenon that we observe in our database is that the average rank eventually decreases as height increases.

## 1. Introduction and statement of main results

Over the past several decades, tables of elliptic curves defined over  $\mathbb{Q}$  have been very useful in number-theoretic research. A natural ordering on elliptic curves is given by their conductor. Some of the earliest tables were those in Antwerp IV [10], which include all elliptic curves of conductor at most 200. In [15], Cremona describes algorithms for listing all elliptic curves of given conductor and collects arithmetic data for these curves; these algorithms have now produced an exhaustive list of curves of conductor at most 380 000 in an ongoing project [16]. A large currently available database of elliptic curves is due to Stein and Watkins [2, 29], which includes 136 832 795 curves over  $\mathbb{Q}$  of conductor up to  $10^8$  and a table of 11 378 911 elliptic curves over  $\mathbb{Q}$  of prime conductor up to  $10^{10}$ , extending earlier tables of this type by Brumer and McGuinness [14].

It is computationally difficult to produce exhaustive lists of curves up to a given conductor. It is far easier to produce large tables by writing down elliptic curves in Weierstrass form with relatively small defining coefficients; such tables may not, however, include all curves up to a given conductor, as it occasionally happens that curves with small coefficients can have large conductor, or curves with large coefficients can have cancellation leading to a smaller than expected conductor. For example, the Stein–Watkins table contains approximately 78.5% of the elliptic curves of conductor up to 120 000 [2]. In very recent work [3], Bennett and Reznitzner pursue a different strategy for producing extensive lists of curves with good reduction outside a given prime  $p$  by using the reduction theory of binary cubic forms and solving certain Thue–Mahler equations. They find 435 893 911 isomorphism classes of curves of prime conductor up to  $10^{12}$  and explain that it is unlikely that any have been missed.

---

Received 4 February 2016.

2010 Mathematics Subject Classification 11G05 (primary), 11-04 (secondary).

Contributed to the Twelfth Algorithmic Number Theory Symposium (ANTS-XII), Kaiserslautern, Germany, 29 August–2 September 2016.

We thank the AMS Mathematics Research Communities program, where we began working on this project, for their continued financial support. The first author was partially supported by NSF grant DMS-1103831. The second author was supported by NSF grant DMS-1406066.

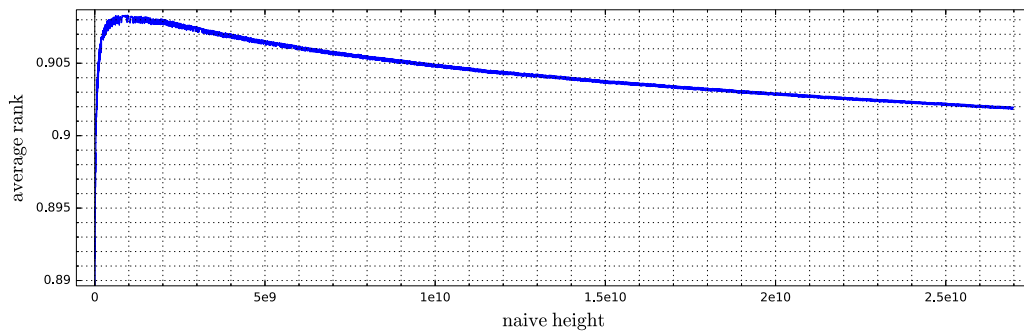


FIGURE 1. Average rank of elliptic curves up to a given height.

In this paper, we instead describe databases of curves in families ordered by *height*, a measure of the size of the coefficients of the Weierstrass equation defining the curve, since it is possible to list all curves in a specified height range. We first consider the family  $\mathcal{F}_0$  of all elliptic curves over  $\mathbb{Q}$  in short Weierstrass form

$$\mathcal{F}_0 = \{E : y^2 = x^3 + a_4x + a_6 \mid a_4, a_6 \in \mathbb{Z}, \Delta_E \neq 0\},$$

where  $\Delta_E = -16(4a_4^3 + 27a_6^2)$  denotes the discriminant of the curve  $E$ . There are two main height functions that we will consider for this family throughout this paper. The (*naive*) *height* is defined by  $H(E) := \max\{4|a_4|^3, 27a_6^2\}$ . The *uncalibrated height* is defined by  $\tilde{H}(E) := \max\{|a_4|^3, a_6^2\}$ . We believe that the naive height is the more natural of these two, but both are used in practice. Both theoretical and computational results using either height have a very similar form. Throughout this paper, we often just write ‘height’ in place of ‘naive height’.

The main result of this project is the creation of an exhaustive database of isomorphism classes of elliptic curves with naive height up to  $2.7 \cdot 10^{10}$ ; a total of 238 764 310 curves [1]. For each elliptic curve in this database, we have recorded its minimal model, torsion subgroup, conductor, Tamagawa product, rank and 2-Selmer group rank. The recorded rank for some of the elliptic curves is conditional on the generalized Riemann hypothesis (GRH), the Birch and Swinnerton-Dyer (BSD) conjecture and/or the parity conjecture (see § 2). However, the recorded rank is unconditional for at least 192 850 627 (approximately 80.77%) of the curves. Note that the largest conductor occurring in our database is 863 347 196 528.

In the databases of curves that currently exist, for example those compiled in [2, 16, 29], the average rank of elliptic curves appeared, from a distance, to be monotonically increasing as the conductor increases, although that would contradict widely believed conjectures (see § 1.1). Ours is the first database in which we can see a ‘turnaround’ point for the average rank of elliptic curves: the average rank of all curves of height up to  $X$  appears to be an increasing function of  $X$  for  $X$  up to approximately  $6 \cdot 10^8$  and then looks, from a distance, like a monotonically decreasing function (although, of course, up close it is wildly gyrating). See Figure 1 for a plot of the average rank of elliptic curves up to height  $X$ , using our database. It would be interesting to have theoretical results confirm this observed turnaround point. Of course, we cannot prove that average rank decreases monotonically (on a large scale) after this turnaround point, but this seems to be a reasonable assumption based on standard heuristics and our data (see §§ 1.1 and 1.5).

From our database, we find that the 238 764 310 curves of naive height up to  $2.7 \cdot 10^{10}$  have average rank approximately 0.901976. The proportion of these curves with each rank is

as follows.

Rank 0	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
0.32685	0.47381	0.17151	0.02615	0.00159	0.00003	0.000005

In the [Appendix](#), we include additional data about rank distribution for different height ranges. As a point of comparison, the average rank of the 126 427 408 curves with uncalibrated height up to  $10^9$  is 0.89473.

There has been much interest in finding curves of minimal conductor and given rank. Only the first few of these minimal conductors are known: 11, 37, 389, 5077, and 234 446. It is easy to extract analogous results for curves of minimal height from our database. We include a table of minimal height curves with given torsion subgroup and rank as in [Table A.6](#).

Another problem that arises when computing exhaustive tables of curves is that it is unknown how many curves they will contain if ordered by conductor or discriminant. For example, Watkins [\[31, Heuristic 4.1\]](#) suggests that the number of curves over  $\mathbb{Q}$  with conductor bounded by  $X$  is asymptotically a constant times  $X^{5/6}$ , and Brumer and McGuinness [\[14\]](#) conjecture a similar asymptotic for the number of curves over  $\mathbb{Q}$  with absolute discriminant less than  $X$ . In contrast, for curves ordered by height, the number of curves of height at most  $X$  is known to be asymptotic to a constant times  $X^{5/6}$ . For curves ordered by uncalibrated height, the constant is known to be  $4/\zeta(10)$  (see, for example, [\[20\]](#)). In our database, we observe fast convergence to this asymptotic: the number of curves of uncalibrated height at most  $10^9$  is 126 427 408, which is  $1.00049 \cdot (4/\zeta(10))(10^9)^{5/6}$ .

### 1.1. The minimalist conjecture

The data sets previously available for elliptic curves defined over  $\mathbb{Q}$  are in tension with certain widely believed conjectures (see [\[2\]](#)). We now describe some of these conjectures about the distribution of ranks.

The minimalist conjecture, inspired partly by work of Katz–Sarnak relating elliptic curves over function fields to certain random matrix statistics [\[23\]](#), and by a similar conjecture of Goldfeld for ranks of elliptic curves in families of quadratic twists [\[19\]](#), states that asymptotically half of all curves have rank zero and half have rank one, so the average rank is exactly  $1/2$ .

The sign appearing in the functional equation for the  $L$ -function associated to an elliptic curve  $E$  over  $\mathbb{Q}$  is the *parity* of  $E$ , and it is conjectured that among all curves ordered by any reasonable invariant (like conductor, discriminant or height), asymptotically half will have each parity. The parity conjecture, a consequence of the BSD conjecture, states that the sign of the functional equation is equal to the parity of the Mordell–Weil rank of  $E$ . Therefore, conjecturally, half of all curves have odd rank and thus have rank at least one, so the average rank of curves ordered by conductor is at least  $1/2$ .

The minimalist conjecture also follows from this conjectural equidistribution of parity and the idea that elliptic curves generally have as small a rank as allowed by parity (see also [\[2\]](#) for an extended discussion). A consequence of the minimalist conjecture is that zero per cent of curves (asymptotically) should have rank at least two. We discuss heuristics for higher rank curves in [§ 3.1.1](#).

### 1.2. Selmer groups

Recent breakthroughs involving orbit parametrizations of genus one curves and the geometry of numbers have led to new unconditional bounds on average ranks of elliptic curves ordered by naive height. These rank bounds are consequences of results on Selmer groups of elliptic curves.

For each integer  $n \geq 2$ , the  $n$ -Selmer group  $S_n(E)$  of an elliptic curve  $E$  over  $\mathbb{Q}$  fits into an exact sequence

$$0 \rightarrow E(\mathbb{Q})/nE(\mathbb{Q}) \rightarrow S_n(E) \rightarrow \text{III}(E)[n] \rightarrow 0, \quad (1.1)$$

where  $\text{III}(E)[n]$  denotes the  $n$ -torsion subgroup of the Tate–Shafarevich group  $\text{III}(E)$  of  $E$  over  $\mathbb{Q}$ . If  $p$  is a prime, then  $S_p(E)$  is an elementary abelian  $p$ -group, whose dimension as an  $\mathbb{F}_p$ -vector space is called the  $p$ -Selmer rank of  $E$ .

**THEOREM 1.1** (Bhargava and Shankar [6–9]). *When all elliptic curves  $E/\mathbb{Q}$  are ordered by naive height, for  $n \leq 5$ , the average size of  $S_n(E)$  is  $\sigma(n)$ , the sum of the divisors of  $n$ .*

In § 3.2, we discuss the 2-Selmer group sizes for elliptic curves in our database, ordered by height. We believe that this is the first large-scale database of Selmer group information to compare with these theoretical results. The average size of  $S_2(E)$  for all curves of height at most  $2.7 \cdot 10^{10}$  is 2.6656 and seems to be increasing towards the theoretical asymptotic average of 3.

A consequence of the Selmer group result for  $n = 5$  in Theorem 1.1 is an upper bound on the average Mordell–Weil rank.

**COROLLARY 1.2** (Bhargava and Shankar [9]). *When all elliptic curves over  $\mathbb{Q}$  are ordered by height, their average rank is at most 0.885.*

Our data, especially the samples at larger height (see § 1.5), suggest that, in fact, the average rank of elliptic curves is well below 0.885. Extending Theorem 1.1 to all  $n$  would imply the minimalist conjecture, and such a generalization is supported by heuristics such as [5, Conjecture 1.3], obtained by modeling the exact sequence (1.1) of  $\mathbb{Z}_p$ -modules via random maximal isotropic spaces.

### 1.3. Other families

Bhargava and Ho have adapted these Selmer group arguments to apply to families of curves with marked points. We define the family of elliptic curves

$$\mathcal{F}_1 := \{E : y^2 + a_3y = x^3 + a_2x^2 + a_4x \mid a_2, a_3, a_4 \in \mathbb{Z}, \Delta_E \neq 0\},$$

where  $\Delta_E$  denotes the discriminant of the elliptic curve  $E$ . For this family, there is a natural height function  $H_1(E) := \max\{a_2^6, a_3^4, |a_4|^3\}$  for  $E \in \mathcal{F}_1$ . In [4], they show that if elliptic curves in  $\mathcal{F}_1$  are ordered by height  $H_1$ , then the average size of the 2-Selmer groups is bounded above by 6, the average size of the 3-Selmer groups is 12, and the average rank is bounded by  $13/6$ .

We have created a database of all isomorphism classes of elliptic curves in  $\mathcal{F}_1$  with height  $H_1 \leq 10^8$  and computed the same invariants, such as rank and 2-Selmer rank. Note that only 693 601 (approximately 19.3%) of these curves are in the main database. This database is discussed in more detail in § 3.5.

In [4], several other families of elliptic curves with marked points are also studied with similar results on the average sizes of Selmer groups and bounds on average ranks. It would be interesting to create databases for these families to compare with the theoretical results.

### 1.4. Other properties

Our database also includes several other invariants of elliptic curves for which we can give results similar to those above, for example the number of curves with a given torsion subgroup and rank. Table A.5 gives the number of elliptic curves  $E$  of naive height at most  $2.7 \cdot 10^{10}$  having certain interesting properties, as well as the proportion of different ranks. We also

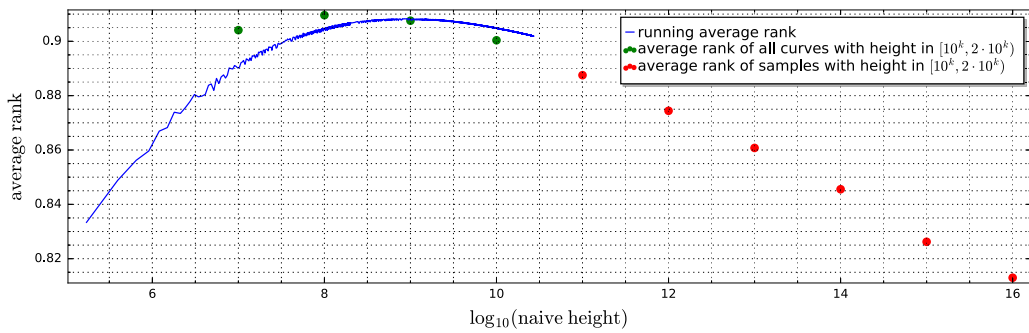


FIGURE 2. Average rank of elliptic curves ( $\log_{10}$  scale).

may compute averages related to each of these properties: for example, the average 2-rank of  $\text{III}(E)$  for the curves in the main database is 0.23912 and the average rank of all curves having complex multiplication in this database is 0.89848.

We note that 19.99% of the curves in our database have positive discriminant and that the rank distributions for curves of positive and negative discriminant appear to have different behavior: the average rank of curves in our database with  $\Delta_E < 0$  is 0.88694 and the average rank of curves with  $\Delta_E > 0$  is 0.961245. We discuss some of these issues further in §3.1.

### 1.5. Samples

We have also created small databases of random samples of elliptic curves at larger heights. In particular, for each  $k \in [11, 16]$ , we chose 100 000 curves from a uniform distribution of all curves in the height range  $[10^k, 2 \cdot 10^k)$  and computed the same invariants. These are not exhaustive datasets, but still provide some evidence for the behavior of various quantities as height increases. For example, we see the average rank decreases rapidly: see Table A.3 and Figure 2, where the red points denote the average ranks for the samples, the green points denote the average rank for all curves in the height range  $[10^k, 2 \cdot 10^k)$  and the blue curve represents the running average rank of all curves up to a given height.

We also record the distribution of the orders of the 2-Selmer groups in Table A.4; note the rapid convergence of the average 2-Selmer size to the theoretical average of 3. See Figures 4, 6, and 7 for the proportion of rank two curves, the average 2-Selmer size, and the average 2-rank of  $\text{III}[2]$ , respectively. Each figure includes the relevant values for these samples, denoted by red points; the green dots denote the corresponding values for all curves in the height range  $[10^k, 2 \cdot 10^k)$  for  $k = 7, 8, 9, 10$  and the blue curve represents the running average or proportion.

## 2. Computing ranks of elliptic curves

In this section, we describe our methods for populating our databases and computing information about Mordell–Weil groups and Selmer groups for the elliptic curves in the databases. There are two challenges involved when computing the rank: one must exhibit explicit rational points on  $E$  while simultaneously showing that the rank of  $E(\mathbb{Q})$ , denoted by  $\text{rk } E(\mathbb{Q})$ , is no more than the rank of the subgroup generated by the known points.

There are two fundamentally different ways of obtaining upper bounds on the rank. The first relies on computing the  $n$ -Selmer group  $S_n(E)$  for various integers  $n$ . This method gives the correct answer whenever it terminates, but its termination is conditional on the conjecture that  $\text{III}(E)$  is finite. The second relies on computing upper bounds on the order of vanishing of the  $L$ -function attached to  $E$  and is conditional on the Birch and Swinnerton-Dyer conjecture.

### 2.1. Populating databases

It is a straightforward task to write down all pairs of integers  $(a_4, a_6)$  such that the naive height  $\max\{4|a_4|^3, 27a_6^2\}$  of the corresponding curve  $y^2 = x^3 + a_4x + a_6$  is in any chosen range. For each such pair, we check that the discriminant of the corresponding curve does not vanish, that is, that the curve is non-singular. A curve of this form is isomorphic to one of smaller height if and only if there is a prime  $p$  such that  $p^4 \mid a_4$  and  $p^6 \mid a_6$ , and it is straightforward to remove non-minimal duplicates. This process gives an exhaustive list of all isomorphism classes of curves in the desired height range. For ease of computation and data analysis, we store the main database in approximately 30 shards, most corresponding to a height range of size  $10^9$ .

We use a similar process to create an exhaustive list of isomorphism classes of curves in  $\mathcal{F}_1$  of height at most  $10^8$  using the modified height function  $H_1$  for that family.

To create each of the larger height samples of 100 000 curves, for each integer  $k \in [11, 16]$ , we repeatedly sample integers  $a_4$  and  $a_6$  uniformly from the appropriate ranges such that  $4|a_4|^3 < 2 \cdot 10^k$  and  $27a_6^2 < 2 \cdot 10^k$ . If the curve  $y^2 = x^3 + a_4x + a_6$  is non-singular, minimal, and has naive height at least  $10^k$ , then it is entered in the database.

### 2.2. General procedure for computing rank

The goal of this section is to explain how we compute the Mordell–Weil rank for each curve in our databases. We assume several conjectures during these computations: Birch and Swinnerton-Dyer (BSD), generalized Riemann hypothesis (GRH), and parity, although not all of them are needed for every curve. Recall that the parity conjecture, which follows from BSD, states that for an elliptic curve  $E$  over  $\mathbb{Q}$ , the root number is  $(-1)^{\text{rk } E(\mathbb{Q})}$ .

For each curve, we first compute standard arithmetical data, such as the conductor, root number, Tamagawa product and torsion subgroup. To obtain rank, the first major step is to run Cremona’s `mwrnk` program with default parameters, which searches for points of low height and runs a 2-descent. For each curve, `mwrnk` yields the 2-Selmer rank and upper and lower bounds for the Mordell–Weil rank.

If the bounds agree, we of course may determine the rank immediately, and if the difference between the upper and lower bound obtained from `mwrnk` is 1, then the root number combined with the parity conjecture gives the value of the rank. However, for many curves, the interval between the `mwrnk` lower and upper bounds contains at least two integers of the ‘correct’ parity, for example curves with even parity, lower bound 0, and upper bound 2.

In these cases, we attempt to improve the upper bound by applying the analytic technique described in § 2.3. The upper bounds coming from this method are conditional on GRH. In Corollary 2.2, for any positive real parameter  $\Delta$ , we obtain an expression in terms of  $\Delta$  that is an upper bound for the analytic rank of  $E$  and which converges to the analytic rank from above. Assuming BSD, the analytic rank is equal to  $\text{rk } E(\mathbb{Q})$ , so applying this bound with large enough  $\Delta$  converges to the correct value of the rank. Unfortunately, this method becomes computationally infeasible for large values of  $\Delta$ . We compute this upper bound with successively larger values of  $\Delta$ , usually between 1 and 3, by the Sage function `analytic_rank_upper_bound`, and stop the process and conclude that we have determined  $\text{rk } E(\mathbb{Q})$  whenever the upper bound is within 1 of the `mwrnk` lower bound. For a small number of curves in our larger height samples, we used this method with values of  $\Delta$  up to 3.9, which took several days for each curve at the highest values of  $\Delta$ .

This process allows us to conclude  $\text{rk } E(\mathbb{Q})$  in the vast majority of cases. For the remaining curves, we use methods in Magma to conclude the correct rank by either finding additional rational points to improve the lower bound or by computing the Cassels–Tate pairing between Selmer group elements to improve the upper bound. These techniques are described in § 2.4.

2.3. Analytic upper bounds

The analytic rank of an elliptic curve may be bounded from above by a certain explicit formula-derived sum over the non-trivial zeros of  $L_E(s)$ , at the expense of having to assume GRH. We reproduce [11, Lemma 2.1], which is a version of the explicit formula for elliptic curve  $L$ -functions that is akin to the Weil formulation of the Riemann-von Mangoldt explicit formula for  $\zeta(s)$ ; a proof may be found in [22, Theorem 5.12].

LEMMA 2.1. Assume GRH. Let  $E$  be an elliptic curve over  $\mathbb{Q}$ , and let

$$b_n(E) = \begin{cases} -(p^e + 1 - \#\tilde{E}(\mathbb{F}_{p^e})) \cdot \log(p), & n = p^e \text{ a prime power,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

where  $\#\tilde{E}(\mathbb{F}_{p^e})$  is the number of points on the (possibly singular) curve over the finite field of  $p^e$  elements obtained by reducing  $E$  modulo  $p$ . Let  $\gamma$  range over the imaginary parts of non-trivial zeros of  $E$  and let  $c_n = c_n(E) = b_n(E)/n$ . Suppose that  $f(z)$  is an entire function such that:

- there exists a  $\delta > 0$  such that  $f(x + iy) = O(x^{-(1+\delta)})$  for  $|y| < 1 + \epsilon$  for some  $\epsilon > 0$ ; and
- the Fourier transform of  $f$ , given by  $\hat{f}(y) = \int_{-\infty}^{\infty} e^{-ixy} f(x) dx$ , exists and is such that  $\sum_{n=1}^{\infty} c_n \cdot \hat{f}(\log n)$  converges absolutely.

Then

$$\sum_{\gamma} f(\gamma) = \frac{1}{\pi} \left[ \log \left( \frac{\sqrt{N_E}}{2\pi} \right) \hat{f}(0) + \Re \int_{-\infty}^{\infty} F(1 + it) f(t) dt + \frac{1}{2} \sum_{n=1}^{\infty} c_n (\hat{f}(\log n) + \hat{f}(-\log n)) \right],$$

where  $F(z)$  denotes the digamma function, the logarithmic derivative of  $\Gamma(z)$ .

We may use the above to provide computationally effective upper bounds on the analytic rank of an elliptic curve by choosing an appropriate test function  $f$  whose Fourier transform has compact support. The method appears to have first been formulated by Mestre in [24], and used by Brumer in [13] to prove that, conditional on GRH, the average rank of elliptic curves is at most 2.3. The method was further refined to produce an upper bound of 2 by Heath-Brown in [21] and then 25/14 by Young in [32].

Specifically, we use the parametrized Fejér kernel as used by Mestre, Brumer and Heath-Brown in the publications above and by Bober in [11]: that is,

$$f_{\Delta}(x) = \text{sinc}^2(\Delta x) = \left( \frac{\sin(\Delta \pi x)}{\Delta \pi x} \right)^2, \tag{2.2}$$

where  $\Delta > 0$  is the tightness parameter. Its Fourier transform is the triangular function

$$\hat{f}_{\Delta}(y) = \begin{cases} \frac{1}{\Delta} \left( 1 - \frac{|y|}{2\pi\Delta} \right), & |y| \leq 2\pi\Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Moreover, the integral  $\Re \int_{-\infty}^{\infty} F(1 + it) f_{\Delta}(t) dt$  can be computed explicitly in terms of known constants and special functions:

$$\Re \int_{-\infty}^{\infty} F(1 + it) \cdot f_{\Delta}(t) dt = -\frac{\eta}{\pi\Delta} + \frac{1}{2\pi^2\Delta^2} \left( \frac{\pi^2}{6} - \text{Li}_2(e^{-2\pi\Delta}) \right). \tag{2.4}$$

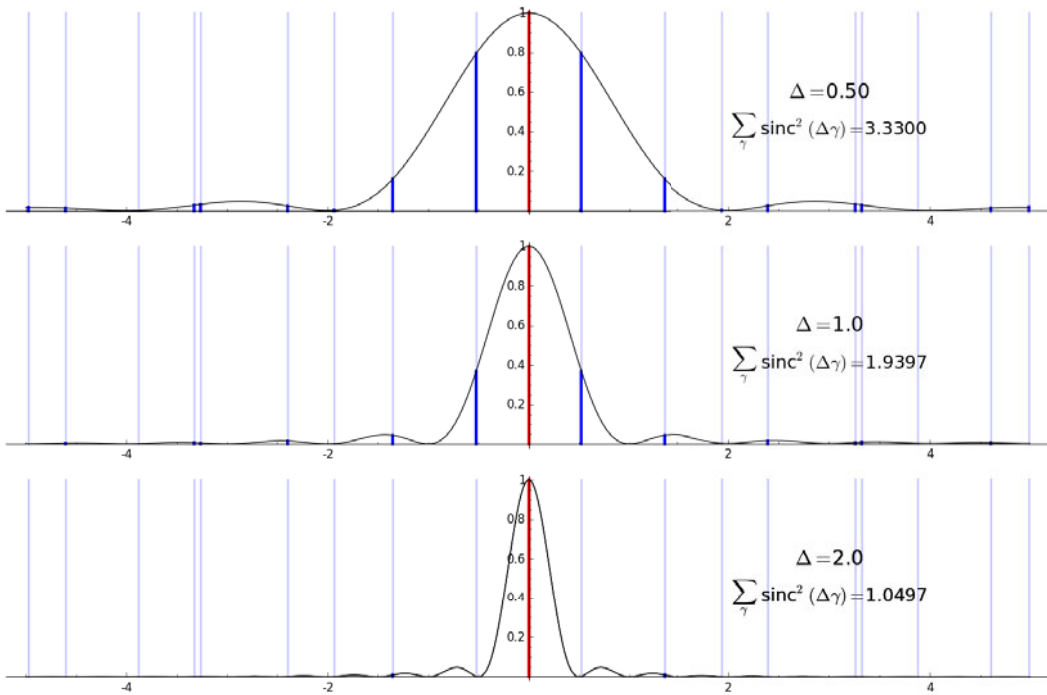


FIGURE 3. A graphic representation of the  $\text{sinc}^2$  sum for the elliptic curve  $E : y^2 = x^3 - 18x + 51$ , a rank one curve with conductor  $N_E = 750384$ , for three increasing values of the parameter  $\Delta$ . Vertical lines have been plotted at  $x = \gamma$  whenever  $L_E(1 + i\gamma) = 0$  (red for the single central zero and blue for non-central zeros); the height of the darkened portion of each line is given by the black curve  $\text{sinc}^2(\Delta x)$ . Summing up the lengths of the dark vertical lines thus gives the value of the  $\text{sinc}^2$  sum. As  $\Delta$  increases, the contribution from the blue lines, corresponding to non-central zeros, goes to zero, while the contribution from the central zero in red remains at 1. Thus the sum must approach 1 as  $\Delta$  increases.

Here  $\eta \approx 0.57722$  is the Euler–Mascheroni constant and  $\text{Li}_2(x) = \sum_{n \geq 1} (x^n/n^2)$  is the logarithmic integral function. Combining (2.2)–(2.4) with Lemma 2.1, we obtain the following corollary.

**COROLLARY 2.2.** *Assume GRH. Let  $\gamma$  range over the imaginary parts of the non-trivial zeros of  $L_E(s)$  and let  $\Delta > 0$ . Then*

$$\sum_{\gamma} \text{sinc}^2(\Delta\gamma) = \frac{1}{\Delta\pi} \left[ \left( -\eta + \log\left(\frac{\sqrt{N_E}}{2\pi}\right) \right) + \frac{1}{2\pi\Delta} \left( \frac{\pi^2}{6} - \text{Li}_2(e^{-2\pi\Delta}) \right) + \sum_{n < e^{2\pi\Delta}} c_n \cdot \left( 1 - \frac{\log n}{2\pi\Delta} \right) \right]$$

and since  $\text{sinc}^2(0) = 1$  and  $\text{sinc}^2(x) \rightarrow 0$  as  $x \rightarrow \infty$ , the sum converges to the analytic rank of  $E$  from above as  $\Delta \rightarrow \infty$ . See Figure 3.

What is notable about the above formula is that evaluation of the right-hand side is a finite computation and only requires knowledge of the conductor of the elliptic curve and a finite number of  $a_p$  values. See [28, p. 67] for a more detailed derivation of this method. We may, therefore (assuming BSD), obtain efficient upper bounds on the rank of an elliptic curve by choosing an appropriate value of  $\Delta$  and performing this finite calculation.



### 2.4. Magma techniques

For a small number of curves in the main database and in the sample databases, we also use additional methods in Magma [12]. While we assume GRH to speed up the construction of 2-coverings and 4-coverings of these curves, the rank bounds coming from these methods are unconditional, since we computed  $S_2(E)$  using `mwrnk` and do not need to use a full list of 2-coverings or 4-coverings.

The Magma procedure begins by carrying out a 2-descent to compute the specific 2-coverings of  $E$  corresponding to elements of  $S_2(E)$  and then searches for rational points of small height on the 2-coverings.

If the rank is not determined at this stage, then we compute the Cassels–Tate pairing on elements of  $S_2(E)/E(\mathbb{Q})[2]$ . Recall that the Cassels–Tate pairing  $\Gamma$  is an alternating bilinear pairing on  $\text{III}(E)$  taking values in  $\mathbb{Q}/\mathbb{Z}$ ; if  $\text{III}(E)$  is finite, then it is non-degenerate. When restricted to  $\text{III}(E)[2]$ , this gives a non-degenerate alternating bilinear pairing on  $\text{III}(E)[2]/2\text{III}(E)[4]$  or, equivalently, on  $S_2(E)/\text{im}(S_4(E))$ , which takes values in  $\mathbb{Z}/2\mathbb{Z} = \{0, 1\}$ . In particular, if  $C$  and  $D$  are 2-coverings of  $E$  with  $\Gamma(C, D) = 1$ , then  $C$  and  $D$  correspond to elements of order two in  $\text{III}(E)$ , which gives an improved upper bound on rank.

If necessary, we next use a 4-descent to find explicit 4-coverings of  $E$  corresponding to some elements of  $S_4(E)$  and then search for rational points of small height on these 4-coverings, refining lower bounds on the rank. For almost all curves, these methods, combined with the parity conjecture, are enough to determine the rank.

For seven curves in our sample database at height  $10^{16}$ , all these techniques, including computing analytic upper bounds with very large values of  $\Delta$  and extensive point searches, do not determine rank: for each of these curves  $E$ , the 2-rank of  $S_2(E)$  is 2 and  $E(\mathbb{Q})$  has trivial torsion. We use Magma to compute the value  $\Gamma(C, D) \in \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$  of the Cassels–Tate pairing for a 4-covering  $C \in S_4(E)$  and a 2-covering  $D \in S_2(E)$ . We find an explicit pair  $(C, D)$  with  $\Gamma(C, D) = 1$ , which implies that  $S_2(E) \cong \text{III}(E)[2] \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$  and thus  $\text{rk } E(\mathbb{Q}) = 0$ . Two such curves for which this method is needed are  $y^2 = x^3 + 169304x + 25788938$  and  $y^2 = x^3 + 77108x - 22146514$ .

## 3. Data analysis

### 3.1. Distribution of rank

As mentioned in § 1, we see from our main database that the average rank of elliptic curves of height up to  $X$  increases with small  $X$  and then decreases (on a large scale); see Figure 1 for a plot of the average rank of all elliptic curves of height at most  $2.7 \cdot 10^{10}$  and Tables A.1 and A.2 for a more detailed distribution. Moreover, in our larger height samples, the average rank decreases as height increases, with the average of the height  $10^{16}$  sample approximately 0.813 (see Table A.3 and Figure 2 for details). For comparison, we note that the average rank of all curves of conductor up to 360 000 is 0.72759 (using data from [16]) and the average rank of all curves of conductor at most  $10^8$  in the Stein–Watkins database is 0.865 [2].

3.1.1. *Higher rank curves.* There are many proposed heuristics for predicting the asymptotics of rank two curves. For example, when curves are ordered by conductor, Watkins [31] predicts, based on ideas from random matrix theory, that the number of isomorphism classes of elliptic curves with conductor at most  $X$  and rank two is  $O(X^{19/24}(\ln X)^{3/8})$ . For curves ordered by height, the recent heuristics of Park, Poonen, Voight and Wood [25] predict that, for  $2 \leq r \leq 20$ , the number of curves with height up to  $X$  and rank at least  $r$  is asymptotic to  $X^{(21-r)/24+o(1)}$  (but are not fine enough to predict the logarithmic term).

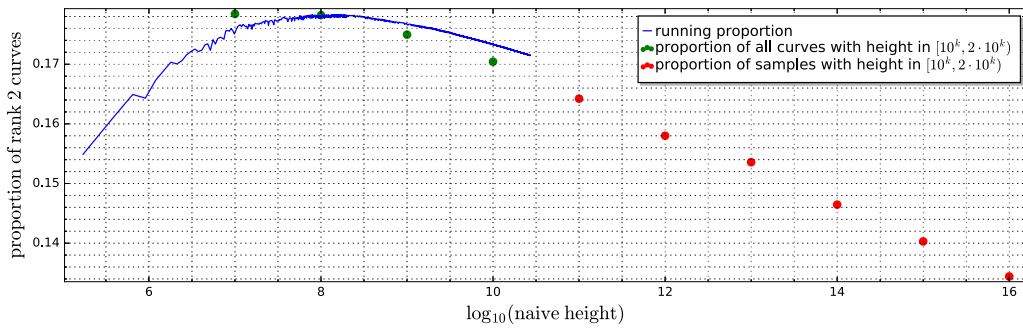


FIGURE 4. Proportion of rank two curves, including samples ( $\log_{10}$  scale).

In our main database, there are 40 949 307 rank two curves, approximately 17.15% of all of the curves (and 34.3% of the even rank curves). The number of rank two curves in the main database is approximately  $0.0686 \cdot (2.7 \cdot 10^{10})^{19/24} (\ln(2.7 \cdot 10^{10}))^{3/8}$  and we note that the constant seems to slightly increase as height increases. However, the proportion of curves of height up to  $X$  having rank two decreases as  $X$  increases (for  $X$  larger than approximately  $10^8$ ): see Figure 4 and Table A.3.

3.1.2. *Positive versus negative discriminant.* It is believed that, asymptotically, the distribution of ranks of elliptic curves  $E$  with height at most  $X$  and  $\Delta_E > 0$  should be the same as the distribution of ranks of elliptic curves  $E$  with height at most  $X$  and  $\Delta_E < 0$ . However, for small values of  $X$ , these distributions initially appear different. Brumer and McGuinness [14] note that, in their database of 310 716 curves of prime conductor up to  $10^8$ , the average rank of those with  $\Delta_E > 0$  is 1.04, while the average rank for those with  $\Delta_E < 0$  is 0.94. In [2], the authors point out that the relationship between the sign of the discriminant and the average rank is a little subtle, computing far enough to find a crossing point in the graphs of average rank of curves of conductor at most  $10^8$  in the Stein–Watkins database with given sign of  $\Delta_E$ .

By the form of our height function, a curve  $E$  has  $\Delta_E > 0$  if and only if  $4|a_4|^3 > 27a_6^2$  and  $a_4 < 0$ . This accounts for exactly half of curves for which  $4|a_4|^3 > 27a_6^2$ , but less than half (in fact, 19.99%) of all curves in our main database. Among all curves of naive height at most  $2.7 \cdot 10^{10}$ , we see that the average rank of those with  $\Delta_E > 0$  is 0.961245 while the average rank of those with  $\Delta_E < 0$  is 0.88694. In fact, rank is weakly correlated with the sign of the discriminant, with a correlation coefficient  $r = 0.03856$ ; while this correlation value is small, it is still significant given the large size of the database. Note also that the fact that the sizes of these two sets of curves are not close to being equal does not explain this discrepancy in rank. It would be interesting to have a theoretical explanation for these numerical observations. Figure 5 plots the average rank for all curves with each sign of  $\Delta_E$  and height less than a given value.

### 3.2. Selmer groups and Tate–Shafarevich groups

Theorem 1.1 says that the average size of  $S_2(E)$  among all elliptic curves converges to 3. Since the rank of  $S_2(E)$  gives an upper bound on  $\text{rk } E(\mathbb{Q})$  and the average rank of small height is larger than the conjectured asymptotic value of  $1/2$ , it may seem reasonable to guess that the average size of  $S_2(E)$  exceeds the theoretical average. However, the average size of  $S_2(E)$  for all curves of height up to  $X$  where  $X \leq 2.7 \cdot 10^{10}$  appears to be increasing towards the predicted value of 3. In our samples, the average size of  $S_2(E)$  increases in each larger height sample, with the average size in the  $10^{16}$  sample already up to 2.90311 (see Table A.4 and Figure 6).

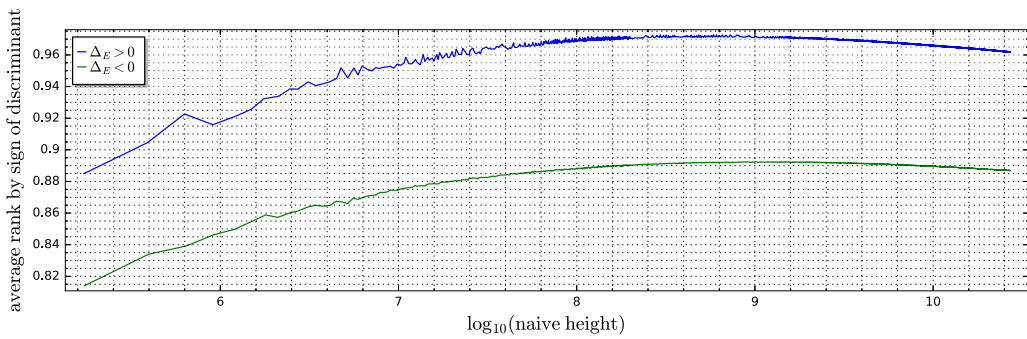


FIGURE 5. Average rank of curves with positive discriminant versus negative discriminant ( $\log_{10}$  scale).

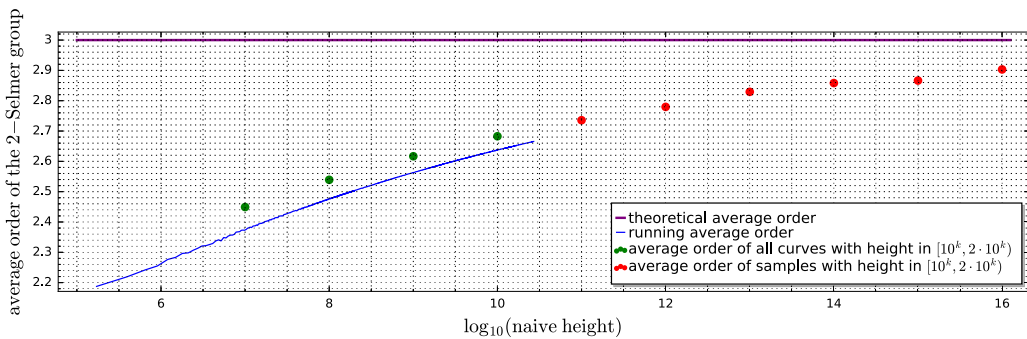


FIGURE 6. Average order of the 2-Selmer groups, including samples ( $\log_{10}$  scale).

Conjecture 1.1 in [26] predicts that, as we vary over all elliptic curves  $E$  over  $\mathbb{Q}$  ordered by height,

$$\text{Prob}(\dim_{\mathbb{F}_p} S_p(E) = d) = \left( \prod_{j \geq 0} (1 + p^{-j})^{-1} \right) \left( \prod_{j=1}^d \frac{p}{p^j - 1} \right), \tag{3.1}$$

which is compatible with the more general conjectures of [5]. For  $p = 2$ , equation (3.1) predicts that the proportion of curves with  $\dim_{\mathbb{F}_2} S_2(E) = 0, 1$  and  $2$  should be approximately  $0.2097, 0.4194$ , and  $0.2796$ , respectively. We see that the proportion of curves in our main database with  $\dim_{\mathbb{F}_2} S_2(E) = 0, 1$  and  $2$  are approximately  $0.2381, 0.4449$  and  $0.2578$ , respectively, which are quite close to the predicted values.

Since we record  $\text{rk } E(\mathbb{Q})$ , the size of  $S_2(E)$  and the torsion subgroup  $E(\mathbb{Q})_{\text{tors}}$  for each elliptic curve  $E$  in our database, we also easily deduce the size of the 2-torsion part  $\text{III}(E)[2]$  of the Tate–Shafarevich group of  $E$ . Delaunay gives a conjecture for the asymptotic distribution of  $\text{rk}_{p^j} \text{III}(E)$  in analogy with the Cohen–Lenstra heuristics for class groups of number fields [17]. More precisely, as we vary over all curves  $E$  over  $\mathbb{Q}$  up to isomorphism of rank  $r$  ordered by conductor, he predicts that

$$\text{Prob}(\dim_{\mathbb{F}_p} \text{III}(E)[p] = 2n) = p^{-n(2r+2n-1)} \frac{\prod_{i=n+1}^{\infty} (1 - p^{-(2r+2i-1)})}{\prod_{i=1}^n (1 - p^{-2i})}.$$

See [26, Conjecture 5.1] for this version of the statement, where it is noted that it is reasonable to expect the same result to hold for curves ordered by height, or [18, Conjecture 4] for a slightly different phrasing. For example, these heuristics predict that for rank zero curves, the proportion with 2-rank of  $\text{III}(E)$  equal to  $0, 2$  or  $4$  is equal to  $0.4194, 0.5592$  or  $0.0213$ ,

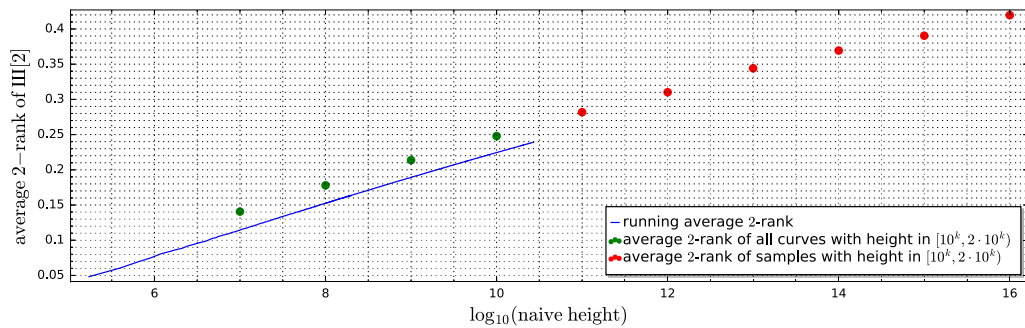


FIGURE 7. Average 2-rank of  $\text{III}[2]$ , including samples ( $\log_{10}$  scale).

respectively, and for rank one curves, the proportion with 2-rank 0 or 2 is 0.8388 or 0.1598, respectively. The moments of the conjectured distribution of  $|\text{III}(E)(p^j)|$  are then computed by Delaunay and Jouhet as [18, Conjecture 3]: the expected value of  $|\text{III}(E)[2]|$  for curves of rank  $r$  is  $1 + 2^{-(2r-1)}$ .

In our main database, the proportions of curves with rank  $r = 0$  and  $\dim_{\mathbb{F}_2} \text{III}(E)[2] = 2n$  for  $n = 0, 1$  and  $2$  are 0.7294, 0.2695 and 0.0011, respectively, and with  $r = 1$  and  $n = 0, 1$  are 0.9393 and 0.0607, respectively. The average size of  $|\text{III}(E)[2]|$  for these rank zero curves is 1.825 and for rank one is 1.182. We see that the rank zero distribution of  $\text{III}(E)[2]$  is not particularly close to the theoretical predictions, but that the data fit more closely for curves of rank one. In each case, the average size of  $\text{III}(E)[2]$  is significantly smaller than expected, which helps to explain why the average size of  $S_2(E)$  appears to approach the asymptotic value of 3 from below, even though the average rank seems to approach the asymptotic value of  $1/2$  from above.

See Figure 7 for a plot of the average 2-rank of  $\text{III}[2]$  up to a given height and Figure A.1 for a plot of the average size of  $\text{III}[2]$  up to a given height for rank zero and rank one curves separately.

### 3.3. Other invariants: root number and torsion subgroups

One may ask whether the convergence of other arithmetic invariants appears to be faster than the convergence of average rank or average size of the 2-Selmer group. For example, it is natural to conjecture that unless there is a good reason to believe that they must be biased, the root numbers of elliptic curves in families should be equidistributed. As noted earlier, in our main database the proportion of curves with root number 1 is 0.49995, already quite close to the conjectured value of  $1/2$ . Figure A.2 shows how the average root number appears to quickly converge to the expected theoretical value of 0.

Another example comes from studying the torsion subgroups of elliptic curves. We know that, as  $X \rightarrow \infty$ , the average size of the torsion subgroup of an elliptic curve of height up to  $X$  approaches 1 (see Figure A.3 where this convergence appears to be quite fast). We recall a more precise theorem.

**THEOREM 3.1** (Harron and Snowden [20]). *Consider all elliptic curves in  $\mathcal{F}_0$  ordered by uncalibrated height. Let  $N_G(X)$  be equal to the number of isomorphism classes of elliptic curves of height up to  $X$  with torsion subgroup  $G$ . Then*

$$N_{\text{trivial}}(X) \sim \frac{4}{\zeta(10)} X^{5/6}, \quad N_{\mathbb{Z}/2\mathbb{Z}}(X) \sim c_2 X^{1/2} \quad \text{and} \quad N_{\mathbb{Z}/3\mathbb{Z}}(X) \sim c_3 X^{1/3},$$

where  $c_2 \approx 3.1969$  and  $c_3 \approx 1.5221$ .

We emphasize that this result uses the uncalibrated height function; our database includes all 12 6427 408 elliptic curves of uncalibrated height at most  $10^9$ . Table A.7 includes the number and average rank of these elliptic curves with each possible torsion subgroup.

The number of curves with trivial torsion and uncalibrated height at most  $10^9$  is approximately  $0.9995 \cdot (4/\zeta(10))(10^9)^{5/6}$ . Similarly, the numbers with uncalibrated height at most  $10^9$  and  $E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z}$  and with  $E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/3\mathbb{Z}$  are approximately  $1.2125 \cdot 3.1969(10^9)^{1/2}$  and  $0.4993 \cdot 1.5221(10^9)^{1/3}$ , respectively. The number of curves with trivial torsion and bounded uncalibrated height appears to converge very quickly to the theoretical value, with the convergence being slower for torsion subgroups that occur less frequently in  $\mathcal{F}_0$ .

We note that if we restrict to curves with a given torsion subgroup, the analogue of the minimalist conjecture is expected to hold, implying that half of all curves have rank zero and half have rank one. We find that the average rank of all curves of naive height at most  $2.7 \cdot 10^{10}$  and  $E(\mathbb{Q})_{\text{tors}} = \mathbb{Z}/2\mathbb{Z}$  is 0.79895 and the average rank of those with  $E(\mathbb{Q})_{\text{tors}} = \mathbb{Z}/3\mathbb{Z}$  is 0.60882. Figures A.4 and A.5 show plots of average rank of curves with naive height up to  $X$  and torsion subgroups  $\mathbb{Z}/2\mathbb{Z}$  and  $\mathbb{Z}/3\mathbb{Z}$ , respectively.

### 3.4. Elliptic curves with complex multiplication

In [2], the authors also consider average rank statistics for elliptic curves with complex multiplication (CM), which have endomorphism ring (over  $\mathbb{C}$ ) strictly larger than  $\mathbb{Z}$ . In the Stein–Watkins database of curves with conductor at most  $10^8$ , the proportion of the set of 135 226 curves with CM that have rank zero is 0.411, significantly larger than the overall proportion 0.336 of curves of rank zero. The average rank of the CM curves in that database is 0.687.

In our main database of curves with naive height at most  $2.7 \cdot 10^{10}$ , there are 65 732 curves with CM, with only 32.819% of them having rank zero. In fact, the rank distribution for the CM curves looks approximately like that of the entire main database, as expected, and the average rank of these CM curves is 0.89848. Figure A.6 gives a plot of average rank up to a given height for these CM curves.

### 3.5. Family of elliptic curves with one marked point

For the family  $\mathcal{F}_1$  of elliptic curves with a marked point, the rank and 2-Selmer rank distribution for the 3594 891 isomorphism classes of elliptic curves in  $\mathcal{F}_1$  with height at most  $10^8$  are as follows.

Rank	No. of curves	% of curves	2-Selmer rank	No. of curves
0	15 783	0.44	0	364
1	1239 600	34.48	1	1145 633
2	1724 209	47.96	2	1727 290
3	564 784	15.71	3	657 323
4	49 642	1.38	4	63 235
5	872	0.024	5	1 045
6	1	0.000028	6	1

The average rank of these curves is 1.83185 and the average size of the 2-Selmer group is 4.31296. Note that the 15 783 rank 0 curves here all have non-trivial torsion, although the marked point is asymptotically a non-torsion point 100% of the time by Hilbert irreducibility.

An analogue of the minimalist conjecture predicts that the average rank among all curves in  $\mathcal{F}_1$  converges to  $3/2$ . Just as in the family  $\mathcal{F}_0$  in our main database, we see that the average rank of curves with ‘small’ height is larger than the expected asymptotic value. Notably, despite

having many fewer curves in this database than in our main database, the distribution here is closer to the asymptotic expectation, for example there are 15.71% rank three curves here, compared with 17.15% rank two curves in the main database.

Appendix. *Additional tables and plots*

TABLE A.1. Rank distribution for isomorphism classes of elliptic curves of naive height  $\leq X$ .

$X$	Rank 0	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
$10^8$	722 275	1 073 502	400 769	51 258	1 551	7	0
$10^9$	4 930 963	7 268 430	2 706 491	384 928	16 975	137	0
$10^{10}$	33 944 219	49 473 528	18 099 044	2 727 260	153 537	2119	1
$2 \cdot 10^{10}$	60 667 897	88 095 239	31 992 709	4 871 438	289 954	4654	5
$2.7 \cdot 10^{10}$	78 039 852	113 128 980	40 949 307	6 259 159	380 519	6481	12

TABLE A.2. Average rank of isomorphism classes of elliptic curves of naive height  $\leq X$ .

$X$	Average rank of elliptic curves of naive height $\leq X$
$10^8$	0.904724540
$10^9$	0.908338779
$10^{10}$	0.904965606
$2 \cdot 10^{10}$	0.902949521
$2.7 \cdot 10^{10}$	0.901975777

TABLE A.3. Rank distribution in samples of 100 000 elliptic curves of height between  $10^k$  and  $2 \cdot 10^k$ .

$k$	Rank 0	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Average rank
11	33 318	47 547	16 422	2495	213	5	0.88753
12	34 018	47 470	15 801	2483	219	9	0.87442
13	34 481	47 665	15 357	2298	192	7	0.86076
14	35 000	47 991	14 647	2180	178	4	0.84557
15	35 941	47 856	14 029	1994	174	6	0.82622
16	36 407	48 105	13 442	1885	155	6	0.81294

TABLE A.4. 2-Selmer ranks in samples of 100 000 elliptic curves of height between  $10^k$  and  $2 \cdot 10^k$ .

$k$	2-rank of 2-Selmer group						Average size of 2-Selmer group
	0	1	2	3	4	5	
11	23 058	44 020	26 363	6015	532	12	2.73566
12	22 829	43 541	26 608	6392	605	25	2.77959
13	22 231	43 257	27 069	6692	729	22	2.82925
14	21 973	43 177	27 073	6968	777	32	2.85819
15	22 162	42 750	27 193	7077	786	32	2.86650
16	21 613	42 631	27 553	7329	836	38	2.90311

TABLE A.5. The number of elliptic curves  $E$  in the main database with various properties and the proportion of curves with each rank out of those with the specified property.

Property	No. of curves (% of database)	Rank 0	Rank 1	Rank 2	Rank $\geq 3$
$E(\mathbb{Q})_{\text{tors}}$ trivial	238 528 817 (99.901%)	0.327	0.474	0.172	0.028
$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z}$	233 153 (0.098%)	0.359	0.492	0.141	0.008
$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/3\mathbb{Z}$	1020	0.463	0.466	0.072	0
$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/4\mathbb{Z}$	257	0.521	0.463	0.016	0
$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/6\mathbb{Z}$	23	0.870	0.130	0	0
$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	1035	0.453	0.496	0.051	0
$\Delta_E > 0$	47 738 800 (19.994%)	0.305	0.466	0.192	0.036
$\Delta_E < 0$	191 025 510 (80.006%)	0.332	0.476	0.166	0.026
$\text{rk}_2(\text{III}(E)[2]) = 0$	210 301 413 (88.079%)	0.271	0.505	0.192	0.032
$\text{rk}_2(\text{III}(E)[2]) = 2$	28 374 370 (11.884%)	0.741	0.242	0.017	0.00024
$\text{rk}_2(\text{III}(E)[2]) = 4$	88 527 (0.037%)	0.978	0.022	0	0
$E$ has CM	65 732	0.328	0.474	0.170	0.028
$E$ has conductor $\leq 10^8$	4 908 673	0.305	0.474	0.193	0.027

TABLE A.6. Elliptic curves  $E$  of the form  $y^2 = x^3 + a_4x + a_6$  with minimal naive height for the specified rank and torsion subgroup.

$E(\mathbb{Q})_{\text{tors}}$	Rank	$a_4$	$a_6$	$H(E)$	Conductor( $E$ )
Trivial	0	-1	-1	27	368
Trivial	1	-1	1	27	92
		1	-1	27	248
		1	1	27	496
Trivial	2	-4	1	256	916
Trivial	3	-13	4	8 788	66 848
Trivial	4	-19	151	615 627	4 705 528
Trivial	5	-217	1585	67 830 075	107 827 292
Trivial	6	-1126	6796	5 710 513 504	35 708 014 976
$\mathbb{Z}/2\mathbb{Z}$	0	0	1	4	64
$\mathbb{Z}/2\mathbb{Z}$	1	-2	0	32	256
$\mathbb{Z}/2\mathbb{Z}$	2	7	8	1 728	4 960
$\mathbb{Z}/2\mathbb{Z}$	3	-82	0	2 205 472	430 336
$\mathbb{Z}/2\mathbb{Z}$	4	1030	6396	4 370 908 000	76 983 424
$\mathbb{Z}/3\mathbb{Z}$	0	0	4	432	108
$\mathbb{Z}/3\mathbb{Z}$	1	0	9	2 187	972
$\mathbb{Z}/3\mathbb{Z}$	2	0	225	1 366 875	24 300
$\mathbb{Z}/4\mathbb{Z}$	0	-2	1	32	40
$\mathbb{Z}/4\mathbb{Z}$	1	-2	21	11 907	760
$\mathbb{Z}/4\mathbb{Z}$	2	-191	-510	27 871 484	7 832
$\mathbb{Z}/5\mathbb{Z}$	0	-43	8208	1 819 024 128	11

Table A.6 (Continued).

$E(\mathbb{Q})_{\text{tors}}$	Rank	$a_4$	$a_6$	$H(E)$	Conductor( $E$ )
$\mathbb{Z}/6\mathbb{Z}$	0	0	1	27	36
$\mathbb{Z}/6\mathbb{Z}$	1	-348	2497	168 576 768	1 260
$\mathbb{Z}/7\mathbb{Z}$	0	-43	166	744 012	26
$\mathbb{Z}/9\mathbb{Z}$	0	-219	1654	73 864 332	54
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	0	-1	0	4	32
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	1	-21	-20	37 044	288
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	2	-73	72	1 556 068	19 040
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	0	-351	1890	172 974 204	24

TABLE A.7. The number of elliptic curves  $E$  in  $\mathcal{F}_0$  with uncalibrated height at most  $10^9$  and specified torsion subgroup.

$E(\mathbb{Q})_{\text{tors}}$	Number of curves	Average rank of these curves
Trivial	126 303 317	0.894838
$\mathbb{Z}/2\mathbb{Z}$	122 574	0.7832
$\mathbb{Z}/3\mathbb{Z}$	760	0.59079
$\mathbb{Z}/4\mathbb{Z}$	188	0.48936
$\mathbb{Z}/5\mathbb{Z}$	1	0
$\mathbb{Z}/6\mathbb{Z}$	16	0.125
$\mathbb{Z}/7\mathbb{Z}$	1	0
$\mathbb{Z}/8\mathbb{Z}$	0	
$\mathbb{Z}/9\mathbb{Z}$	1	0
$\mathbb{Z}/10\mathbb{Z}$	0	
$\mathbb{Z}/12\mathbb{Z}$	0	
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	549	0.56466
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$	1	0
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$	0	
$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$	0	

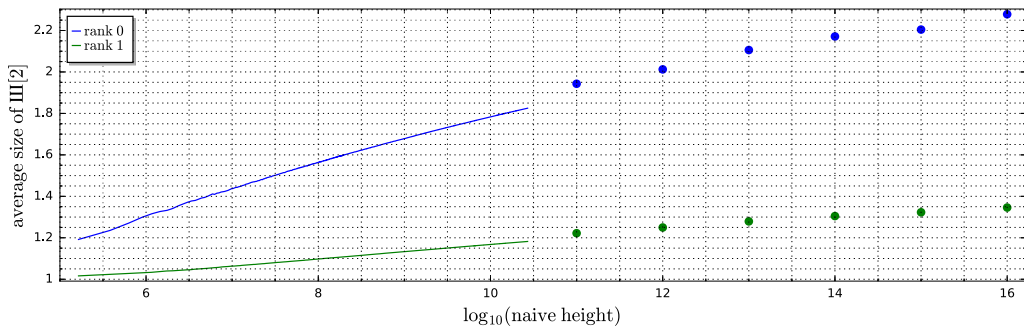


FIGURE A.1. Average size of  $\text{III}[2]$  for curves of rank zero and rank one, including samples ( $\log_{10}$  scale).



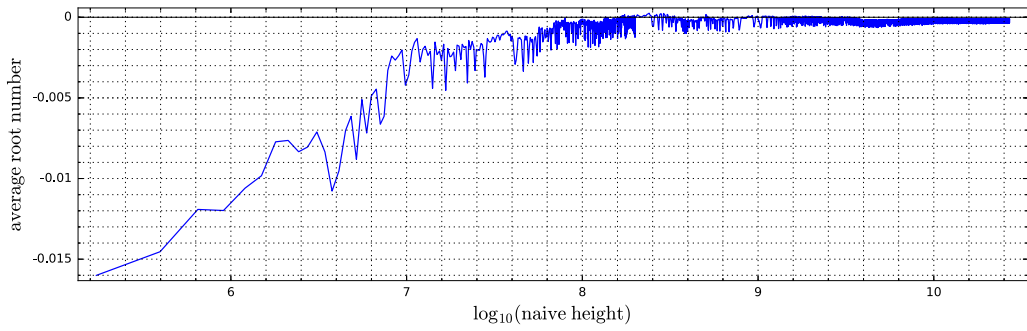


FIGURE A.2. Average root number ( $\log_{10}$  scale).

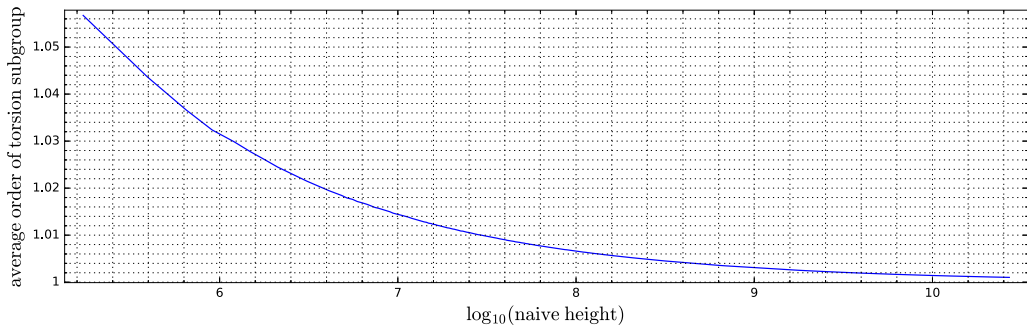


FIGURE A.3. Average order of the torsion subgroup ( $\log_{10}$  scale).

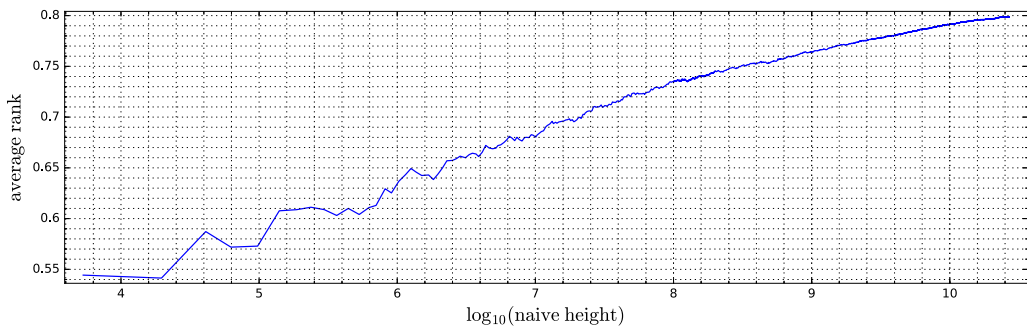


FIGURE A.4. Average rank of curves with torsion subgroup  $\mathbb{Z}/2\mathbb{Z}$  ( $\log_{10}$  scale).

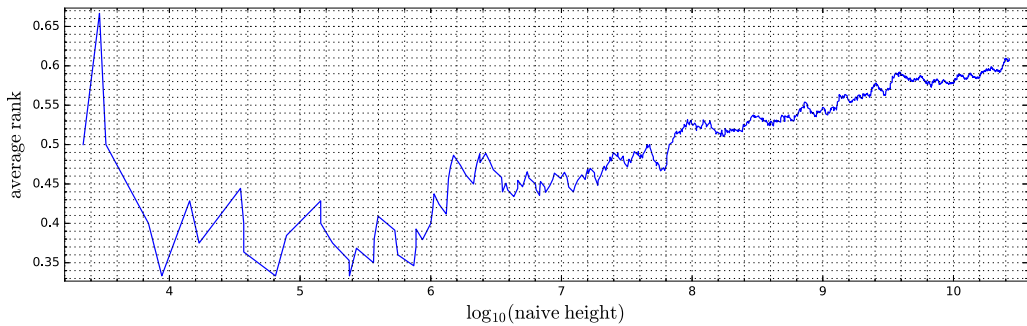
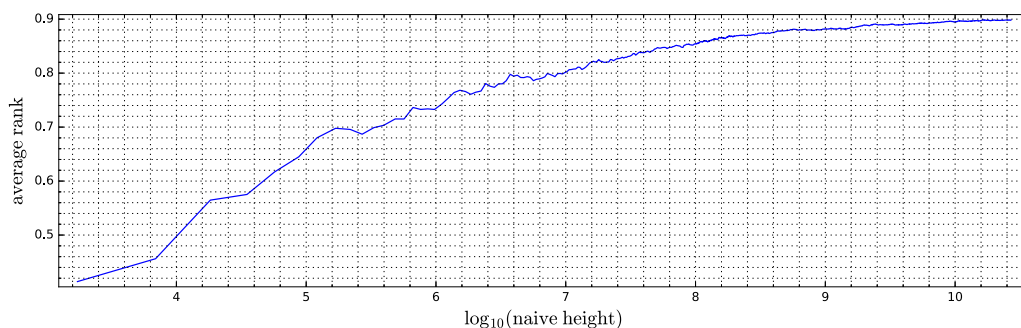
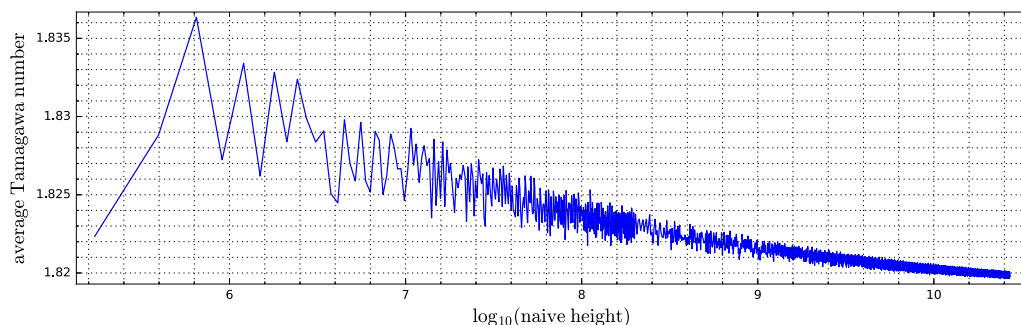


FIGURE A.5. Average rank of curves with torsion subgroup  $\mathbb{Z}/3\mathbb{Z}$  ( $\log_{10}$  scale).

FIGURE A.6. Average rank of CM curves ( $\log_{10}$  scale).FIGURE A.7. Average Tamagawa number ( $\log_{10}$  scale).

*Acknowledgements.* We used the open-source software SageMath [30] and SageMathCloud [27] extensively throughout this project and we thank Harald Schilly for always keeping a close eye on our SMC projects. We are very grateful to the University of Michigan Advanced Research Computing Technology Services and the University of Oxford for providing further computing resources. We also thank Steve Donnelly for discussions about computing rank in Magma [12]. Finally, we thank the anonymous referees who provided many helpful suggestions for improving the paper.

#### References

1. J. S. BALAKRISHNAN, W. HO, N. KAPLAN, S. SPICER, W. STEIN and J. WEIGANDT, <http://wstein.org/papers/2016-height/>.
2. B. BEKTEMIROV, B. MAZUR, W. STEIN and M. WATKINS, ‘Average ranks of elliptic curves: tension between data and conjecture’, *Bull. Amer. Math. Soc. (N.S.)* 44 (2007) no. 2, 233–254.
3. M. BENNETT and A. RECHNITZER, ‘Computing elliptic curves over  $\mathbb{Q}$ : bad reduction at one prime’, Preprint.
4. M. BHARGAVA and W. HO, ‘On the average sizes of Selmer groups in families of elliptic curves’, Preprint.
5. M. BHARGAVA, D. M. KANE, H. W. LENSTRA JR., B. POONEN and E. RAINS, ‘Modeling the distribution of ranks, Selmer groups, and Shafarevich-Tate groups of elliptic curves’, *Camb. J. Math.* 3 (2015) no. 3, 275–321.
6. M. BHARGAVA and A. SHANKAR, ‘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.
7. M. BHARGAVA and A. SHANKAR, ‘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.
8. M. BHARGAVA and A. SHANKAR, ‘The average number of elements in the 4-Selmer groups of elliptic curves is 7’, Preprint, 2013, [arXiv:1312.7333](https://arxiv.org/abs/1312.7333).
9. M. BHARGAVA and A. SHANKAR, ‘The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1’, Preprint, 2013, [arXiv:1312.7859](https://arxiv.org/abs/1312.7859).

10. B. J. BIRCH and W. KUYK (eds), *Modular functions of one variable, Vol. IV*, Lecture Notes in Mathematics 476 (Springer, Berlin–New York, 1975).
11. J. W. BOBER, ‘Conditionally bounding analytic ranks of elliptic curves’, *ANTS X—Proceedings of the Tenth Algorithmic Number Theory Symposium*, Open Book Series 1 (Mathematical Sciences Publishers, Berkeley, CA, 2013) 135–144.
12. W. BOSMA, J. CANNON and C. PLAYOUST, ‘The Magma algebra system. I. The user language’, *J. Symbolic Comput.* 24 (1997) no. 3–4, 235–265, *Computational algebra and number theory* (London, 1993).
13. A. BRUMER, ‘The average rank of elliptic curves. I’, *Invent. Math.* 109 (1992) no. 3, 445–472.
14. A. BRUMER and O. MCGUINNESS, ‘The behavior of the Mordell-Weil group of elliptic curves’, *Bull. Amer. Math. Soc. (N.S.)* 23 (1990) no. 2, 375–382.
15. J. CREMONA, *Algorithms for modular elliptic curves*, 2nd edn (Cambridge University Press, Cambridge, 1997).
16. J. CREMONA, ecdata: 2016-02-07, <http://dx.doi.org/10.5281/zenodo.45657>, February 2016.
17. C. DELAUNAY, ‘Heuristics on Tate-Shafarevitch groups of elliptic curves defined over  $\mathbb{Q}$ ’, *Experiment. Math.* 10 (2001) no. 2, 191–196.
18. C. DELAUNAY and F. JOUHET, ‘ $p^\ell$ -torsion points in finite abelian groups and combinatorial identities’, *Adv. Math.* 258 (2014) 13–45.
19. D. GOLDFELD, ‘Conjectures on elliptic curves over quadratic fields’, *Number theory*, Carbondale, Proceedings of Southern Illinois Conference, Southern Illinois University, Carbondale, Ill, 1979, Lecture Notes in Mathematics 751 (Springer, Berlin, 1979) 108–118.
20. R. HARRON and A. SNOWDEN, ‘Counting elliptic curves with prescribed torsion’, *J. reine angew. Math.*, to appear, doi:[10.1515/crelle-2014-0107](https://doi.org/10.1515/crelle-2014-0107).
21. D. R. HEATH-BROWN, ‘The average analytic rank of elliptic curves’, *Duke Math. J.* 122 (2004) no. 3, 591–623, 04.
22. H. IWANIEC and E. KOWALSKI, *Analytic number theory*, American Mathematical Society Colloquium Publications 53 (American Mathematical Society, Providence, RI, 2004).
23. N. M. KATZ and P. SARNAK, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications 45 (American Mathematical Society, Providence, RI, 1999).
24. J.-F. MESTRE, ‘Formules explicites et minorations de conducteurs de variétés algébriques’, *Compositio Math.* 58 (1986) no. 2, 209–232.
25. J. PARK, B. POONEN, J. VOIGHT and M. WOOD, ‘Heuristics for boundedness of ranks of elliptic curves over  $\mathbb{Q}$ ’, Preprint, 2016, [arXiv:1602.01431](https://arxiv.org/abs/1602.01431).
26. B. POONEN and E. RAINS, ‘Random maximal isotropic subspaces and Selmer groups’, *J. Amer. Math. Soc.* 25 (2012) no. 1, 245–269.
27. SageMath, Inc. *SageMathCloud*, 2016. <https://cloud.sagemath.com>.
28. S. V. SPICER, The zeros of elliptic curve  $L$ -functions: analytic algorithms with explicit time complexity, PhD Thesis, University of Washington, 2015.
29. W. A. STEIN and M. WATKINS, ‘A database of elliptic curves—first report’, *Algorithmic number theory* (Sydney, 2002), Lecture Notes in Computer Science 2369 (Springer, Berlin, 2002) 267–275.
30. The SageMath Developers. Sage Mathematics Software (Version 6.9), 2015. <http://www.sagemath.org>.
31. M. WATKINS, ‘Some heuristics about elliptic curves’, *Experiment. Math.* 17 (2008) no. 1, 105–125.
32. M. YOUNG, ‘Low-lying zeros of families of elliptic curves’, *J. Amer. Math. Soc.* 19 (2006) no. 1, 205–250.

Jennifer S. Balakrishnan  
 Mathematical Institute  
 University of Oxford  
 Woodstock Road  
 Oxford OX2 6GG  
 United Kingdom

[balakrishnan@maths.ox.ac.uk](mailto:balakrishnan@maths.ox.ac.uk)

Nathan Kaplan  
 Department of Mathematics  
 University of California, Irvine  
 Irvine, CA 92697  
 USA

[nckaplan@math.uci.edu](mailto:nckaplan@math.uci.edu)

Wei Ho  
 Department of Mathematics  
 University of Michigan  
 Ann Arbor, MI 48109  
 USA

[weiho@umich.edu](mailto:weiho@umich.edu)

Simon Spicer  
 Facebook Inc.  
 1 Hacker Way  
 Menlo Park, CA 94025  
 USA

[mlungu@fb.com](mailto:mlungu@fb.com)

*William Stein*  
*Department of Mathematics*  
*University of Washington*  
*Seattle, WA 98195-4350*  
*USA*  
[wstein@uw.edu](mailto:wstein@uw.edu)

*James Weigandt*  
*Institute for Computational and*  
*Experimental Research in Mathematics*  
*Brown University*  
*Providence, RI 02912*  
*USA*  
[james\\_weigandt@brown.edu](mailto:james_weigandt@brown.edu)