FOUR PERSPECTIVES ON A CURIOUS SECONDARY TERM

FRANK THORNE

Abstract. This paper is an expanded version of the author's lecture at the Integers Conference 2011. We discuss the secondary terms in the Davenport-Heilbronn theorems on cubic fields and 3-torsion in class groups of quadratic fields. Such secondary terms had been conjectured by Datskovsky-Wright and Roberts, and proofs of these or closely related secondary terms we[['we' or 'were'? Maybe this is a stupid question and is just due to my poor English grammer]] obtained independently by Bhargava, Shankar, and Tsimerman [6], Hough [21], Zhao [40], and Taniguchi and the author [34].

In this paper we discuss the history of the problem and highlight the diverse methods used in [6, 21, 40, 34] to address it.

1. Introduction

This paper concerns the following two theorems.

Theorem 1.1. *Let $N_3^{\pm}(X)$ count the number of cubic fields K with $0 < \pm\mathrm{Disc}(K) < X$. We have*

$$N_3^{\pm}(X) = C^{\pm}\frac{1}{12\zeta(3)}X + K^{\pm}\frac{4\zeta(1/3)}{5\Gamma(2/3)^3\zeta(5/3)}X^{5/6} + o(X^{5/6}), \tag{1.1}$$

where $C^- = 3$, $C^+ = 1$, $K^- = \sqrt{3}$, and $K^+ = 1$.

Theorem 1.2. *For any quadratic field with discriminant D, let $\mathrm{Cl}_3(D)$ denote the 3-torsion subgroup of the ideal class group $\mathrm{Cl}(\mathbb{Q}(\sqrt{D}))$, we have*

$$\sum_{0<\pm D<X} \#\mathrm{Cl}_3(D) = \frac{3 + C^{\pm}}{\pi^2}X + K^{\pm}\frac{8\zeta(1/3)}{5\Gamma(2/3)^3}\prod_p\left(1 - \frac{p^{1/3}+1}{p(p+1)}\right)X^{5/6} + o(X^{5/6}), \tag{1.2}$$

and where the sum ranges over fundamental discriminants D, the product is over all primes, and the constants are as before.

The main terms are due to Davenport and Heilbronn [13], and the secondary term in Theorem 1.1 was conjectured by Roberts [29] and implicitly by Datskovsky and Wright [12].

The secondary terms were expected to be difficult to prove. However, progress has recently made in four independent works, using a variety of methods. Both results above have been proved by Bhargava, Shankar, and Tsimerman [6], using the geometry of numbers, and by Taniguchi and the present author [34], using Shintani zeta functions. Hough [21] has proved a variation of Theorem 1.2 by studying the distribution of Heegner points in the upper complex plane, and Zhao [40] has obtained a variation of Theorem 1.1 for function fields, using algebraic geometry.[1]

At the 2011 Integers Conference in Carollton, Georgia, we explained (briefly) how each of these four approaches all shed light on these secondary terms, and this paper is an expanded version of

[1]A proof of Theorem 1.2 does not appear in [6], but the authors have shown me a proof. The error terms in [21] and [40] are larger than the secondary terms at present, but both of these approaches naturally *explain* the secondary term, and both authors are currently working on refining their methods.

our lecture. We begin with some background on counting fields and on counting cubic fields in particular. The reader might skip to Section 4 for our discussion of the secondary terms.

2. Counting Fields in General

How does one count number fields? We offer only a brief discussion here; for a more detailed discussion we refer to expository accounts such as that by Cohen, Diaz y Diaz, and Olivier [10], or Chapter 6 of Bhargava's ICM proceedings article [4]. Nevertheless, this overview will allow us to draw attention to some additional interesting work on this subject.

The two most important theorems in the subject were proved by Minkowski and Hermite:

Theorem 2.1 (Hermite). *There are only finitely many number fields of bounded discriminant.*

Theorem 2.2 (Minkowski). *The discriminant of a number field K of degree n satisfies*

$$|\mathrm{Disc}(K)| \geq \left(\frac{n^n}{n!}\right)^2 \left(\frac{\pi}{4}\right)^n. \tag{2.1}$$

If we write $n = r + 2s$, where r is the number of real embeddings of K, and s is the number of pairs of complex embeddings, we may replace $(\pi/4)^n$ with $(\pi/4)^{2s}$. Moreover, by Stirling's formula $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ and therefore

$$|\mathrm{Disc}(K)| \geq \left(e^2 \cdot \frac{\pi}{4} + o(1)\right)^n = \left(5.803\cdots + o(1)\right)^n. \tag{2.2}$$

The constant $\frac{e^2\pi}{4}$ can be improved; see Odlyzko [26]. Conversely, this bound is sharp apart from the constant; Golod and Shafarevich [19] proved the existence of an *infinite class field tower* of fields K for which $|\mathrm{Disc}(K)|^{1/n}$ is constant.[2]

To prove these theorems[3], use the r embeddings $K \hookrightarrow \mathbb{R}$ and the s pairs of embeddings $K \hookrightarrow \mathbb{C}$ to embed the ring of integers of K as a lattice in n-dimensional space. As proved by Minkowski, any convex body centered around the origin, whose volume is greater than $2^n|\mathrm{Disc}(K)|$, must contain a lattice point other than zero. However, this point must have norm at least 1, and geometric considerations allow one to conclude Minkowski's lower bound.

Hermite's theorem can be proved in a similar way. We again use Minkowski's convex body theorem to find an element $\alpha \in \mathcal{O}_K$, the sizes of whose complex embeddings are all small, and which is guaranteed to generate K over $\mathbb{Q}$. The minimal polynomial of α is equal to $\prod_\sigma (X - \sigma(\alpha))$, where σ ranges over all real and complex embeddings of K, and the bounds obtained from Minkowski's theorem yield bounds on the coefficients of the minimal polynomial of α. For a fixed discriminant bound, there are only finitely many such polynomials, and thus only finitely many possibilities for K.

These classical theorems suggest that it is natural to count fields by degree. In what follows, let $N_n(X)$ be the number of fields K of degree n with $|\mathrm{Disc}(K)| < X$. (Results are also known when the sign of the discriminant, or more generally the number of real embeddings, is specified.)

We have the following results (excluding $n = 3$):

n = 1. There is only $\mathbb{Q}$.

[2]Martinet [24] gave the example of $F = \mathbb{Q}(\zeta_{11} + \zeta_{11}^{-1}, \sqrt{-46})$, which he proved to have an infinite 2-class field tower, for which $|\mathrm{Disc}(K)|^{1/n} = 92.2\cdots$ for each field K. We refer to Lemmermeyer [22] for a discussion of related issues, along with a very thorough bibliography.

[3]See [25], Ch. III.2 for complete proofs.

$n = 2$. This case is often considered trivial, but we submit that even this case is already very interesting. We have the equality of Dirichlet series

$$\sum_{[K:\mathbb{Q}]=2} |\mathrm{Disc}(K)|^{-s} = -1 + \Big(1 - 2^{-s} + 2 \cdot 4^{-s}\Big) \frac{\zeta(s)}{\zeta(2s)}, \tag{2.3}$$

and it follows that $N_2(X) \sim \frac{6}{\pi^2} + O(\sqrt{X})$.[[Should be $N_2(X) \sim \frac{6}{\pi^2}X + O(\sqrt{X})$]]

This is typically considered the end of the story, but we note several points. In the first place, note that we estimate $N_2(X)$ by counting all integers (with a 2-adic condition) and then sieve for squarefreeness. This is equivalent to counting all quadratic rings (rings which are free $\mathbb{Z}$-modules of rank 2), and sieving for maximality. We will be able to estimate $N_3(X)$ in the same way.

We also note that there are (at least) two ways to estimate $N_2(X)$. The first way is elementary; in the "key" step one observes that there are $2X/q + O(1)$ integers in $[-X, X]$ divisible by q. This is a counting problem in a lattice inside a one-dimensional vector space, and Bhargava's work hinges on higher-dimensional counting problems which are highly nontrivial.

We may also estimate $N_2(X)$ using analytic number theory. In particular, Perron's formula[4] implies that

$$N_2(X) = \int_{2-i\infty}^{2+i\infty} \left[-1 + \Big(1 - 2^{-s} + 2 \cdot 4^{-s}\Big) \frac{\zeta(s)}{\zeta(2s)} \right] \frac{X^s}{s} ds. \tag{2.4}$$

The integral may be shifted to the left by analytic continuation of the zeta function, and the main term of $N_2(X)$ comes from the pole of $\zeta(s)$ at $s = 1$. We recall that the usual proof of the analytic continuation of $\zeta(s)$, using Poisson summation, reflects the fact that the integers form a (one-dimensional) lattice in a one-dimensional vector space.

These two ideas are the starting points of the geometric and analytic proofs of Theorems 1.1 and 1.2, respectively!

We note an additional point related to (2.3). Wright [38] observed that this Dirichlet series has the beautiful representation

$$\sum_{[K:\mathbb{Q}]=2} |\mathrm{Disc}(K)|^{-s} = \prod_p \Bigg(\frac{1}{2} \sum_{[K_v:\mathbb{Q}_p] \leq 2} |\mathrm{Disc}(K_v)|_p^s \Bigg), \tag{2.5}$$

and proved this in a much more general framework. (He obtained similar formulas for degree n cyclic extensions of any number field.) His results follow from considering a nontrivial 'twist' of the adelic zeta function of Tate's thesis [36]. This twist makes the affine line into a *prehomogeneous vector space*, with the action ϕ of $\mathrm{GL}(1)$ given by $\phi(t)x = t^n x$.

Essentially, a vector space is prehomogeneous if it has an action of an algebraic group G, which is transitive over $\mathbb{C}$ apart from the vanishing locus of an irreducible polynomial. This "prehomogeneous" property is essential both in Bhargava's work and in the zeta function approach pioneered by Sato and Shintani [31]; cubic, quartic, and quintic fields are parameterized by lattice points[5] up to the action of $G(\mathbb{Z})$, and these may be counted geometrically or analytically.

Wright's work on (2.5) and its generalizations mirrors his work with Datskovsky [37, 11, 12] on the Shintani zeta function associated to cubic fields, and this latter work is at the heart of the analytic approach to counting cubic fields.

[4]The following formula is valid whenever X is not an integer.

[5]Not *all* of the $G(\mathbb{Z})$-orbits correspond to fields, as we will see in the cubic case.

$n = 4$, 5. Bhargava [2, 5] proved the asymptotic formulas

$$N_4(X) \sim \frac{5}{24} \prod_p \left(1 + p^{-2} - p^{-3} - p^{-4}\right) \cdot X, \tag{2.6}$$

$$N_5(X) \sim \frac{13}{120} \prod_p \left(1 + p^{-2} - p^{-4} - p^{-5}\right) \cdot X, \tag{2.7}$$

and conjectured [3] similar formulas for $N_n(X)$. The constants have elegant interpretations (see [5, 3] as Euler products; each Euler factor represents the 'probability' that a field K of degree n has a certain localization. The beautiful proofs follow the geometric proof of the Davenport-Heilbronn theorems, which we will describe shortly.

$n > 5$. For $n > 5$ we have the conjectures of Bhargava mentioned above, as well as an upper bound $N_n(X) \ll X^{\exp(C\sqrt{\log n})}$ due to Ellenberg and Venkatesh [17]. The techniques used to obtain formulas for $N_n(X)$ for $n \leq 5$ have little prospect of generalizing to $n > 5$, as there are no prehomogeneous vector spaces parameterizing rings of rank > 5. Indeed, there is a classification theorem for prehomogeneous vector spaces due to Sato and Kimura [30], and so we check their list and verify that we are out of luck.

3. Davenport-Heilbronn, Delone-Faddeev, and the main terms

In this section we discuss the proofs of the main terms in Theorems 1.1 and 1.2. To summarize, the idea is as follows: We count cubic *fields* by counting their *maximal orders*, and so we count all cubic orders and then sieve for maximality. We count cubic orders by counting cubic *rings* and then subtracting the contribution of the reducible rings, and these correspond to binary cubic *forms*. (The space of binary cubic forms is *prehomogeneous* in the sense described previously.) Thus at last we have a geometric lattice-point counting problem, which can be solved by "brute force", so to speak. This interpretation in terms of lattice points also opens the door to the use of zeta functions.

We recommend the paper of Bhargava, Shankar, and Tsimerman [6] for the simplest self-contained account of the Davenport-Heilbronn theorems, with complete proofs.

We begin by defining cubic rings and cubic forms. A *cubic ring* is a commutative ring which is free of rank 3 as a $\mathbb{Z}$-module. The discriminant of a cubic ring is defined to be the determinant of the trace form $\langle x, y\rangle = \mathrm{Tr}(xy)$, and the discriminant of the maximal order of a cubic field is equal to the discriminant of the field.

The lattice of *integral binary cubic forms* is defined by

$$V_{\mathbb{Z}} := \{au^3 + bu^2v + cuv^2 + dv^3 \ : a, b, c, d \in \mathbb{Z}\}, \tag{3.1}$$

and the *discriminant* of such a form is given by the usual equation

$$\mathrm{Disc}(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd. \tag{3.2}$$

There is a natural action of $\mathrm{GL}_2(\mathbb{Z})$ (and also of $\mathrm{SL}_2(\mathbb{Z})$) on $V_{\mathbb{Z}}$, given by

$$(\gamma \cdot f)(u, v) = \frac{1}{\det \gamma} f((u, v) \cdot \gamma). \tag{3.3}$$

We call a cubic form f *irreducible* if $f(u, v)$ is irreducible as a polynomial over $\mathbb{Q}$, and *nondegenerate* if $\mathrm{Disc}(f) \neq 0$.

Cubic rings are related to cubic forms by the following correspondence of Delone and Faddeev, as further extended by Gan, Gross, and Savin [18]:

Theorem 3.1 ([14, 18])**.** *There is a natural, discriminant-preserving bijection between the set of* $\mathrm{GL}_2(\mathbb{Z})$*-equivalence classes of integral binary cubic forms and the set of isomorphism classes of cubic rings. Furthermore, under this correspondence, irreducible cubic forms correspond to orders in cubic fields.*

Finally, if $x \in V_{\mathbb{Z}}$ *is a cubic form corresponding to a cubic ring* R*, we have* $\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(x) \simeq \mathrm{Aut}(R)$*.*

It is therefore necessary to exclude reducible and nonmaximal rings. The nonmaximality condition requires more effort to handle, and for this Davenport and Heilbronn established the following criterion:

Proposition 3.2 ([13, 6])**.** *Under the Delone-Faddeev correspondence, a cubic ring* R *is maximal if any only if its corresponding cubic form* f *belongs to the set* $U_p \subset V_{\mathbb{Z}}$ *for all* p*, defined by the following equivalent conditions:*

- *The ring* R *is not contained in any other cubic ring with index divisible by* p*.*
- *The cubic form* f *is not a multiple of* p*, and there is no* $\mathrm{GL}_2(\mathbb{Z})$*-transformation of* $f(u,v) = au^3 + bu^2v + cuv^2 + dv^3$ *such that* a *is a multiple of* p^2 *and* b *is a multiple of* p*.*

In particular, the condition U_p only depends on the coordinates of f modulo p^2.

Therefore one can count cubic fields as follows. One obtains an asymptotic formula for the number of cubic rings of bounded discriminant by counting lattice points in fundamental domains for the action of $\mathrm{GL}_2(\mathbb{Z})$, bounded by the constraint $|\mathrm{Disc}(x)| < X$. The fundamental domains may be chosen so that almost all reducible rings correspond to forms with $a = 0$, and so these may be excluded from the count.

One then multiplies this asymptotic by the product of all the local densities of the sets U_p. This gives a heuristic argument for the main term in (1.1), and one incorporates a sieve to obtain a proof.

The main term of Theorem 1.2 may be proved similarly. By class field theory, there is a bijection between subgroups of $\mathrm{Cl}_3(\mathbb{Q}(\sqrt{D}))$ of index 3, and cubic fields of discriminant D, which are precisely those which are not totally ramified at any prime. This ramification condition may also be detected by reducing cubic forms modulo p^2, and thus may be incorporated into Proposition 3.2. The remainder of the proof is the same.

3.1. **The work of Belabas, Bhargava, and Pomerance.** In [1], Belabas, Bhargava, and Pomerance (BBP) introduced improvements to Davenport and Heilbronn's method, and obtained an error term of $O(X^{7/8+\epsilon})$ in (1.1) (and also in (1.2)). They begin by observing that

$$N_3^{\pm}(X) = \sum_{q \geq 1} \mu(q) N^{\pm}(q, X), \tag{3.4}$$

where $N^{\pm}(q, X)$ counts the number of cubic orders of discriminant $0 < \pm D < X$ which are nonmaximal at every prime dividing q. [[Here and in the following q is squarefree.]] This simple observation has proved to be quite useful! For large q, BBP prove that $N^{\pm}(q, X) \ll X 3^{\omega(q)}/q^2$ using reasonably elementary methods. Therefore, one has

$$N_3^{\pm}(X) = \sum_{q \leq Q} \mu(q) N^{\pm}(q, X) + O(X/Q^{1-\epsilon}). \tag{3.5}$$

For small q, BBP estimate $N^{\pm}(q, X)$ with explicit error terms using geometric methods. These error terms are good enough to allow them to take the sum in (3.4) up to $(X \log X)^{1/8}$, which yields a final error term of $O(X^{7/8+\epsilon})$.

In addition, their methods extend to counting *quartic* fields, where they obtain a main term of C_4X with error $\ll X^{23/24+\epsilon}$.

4. The "Four Approaches"

This, then, brings us to Theorems 1.1 and 1.2. Theorem 1.1 was conjectured in print by Roberts [29], and implicitly in the foundational work of Datskovsky and Wright [37, 11, 12]. Roberts gave substantial and convincing numerical evidence for his conjecture, as well as the heuristic argument described in Section 5.

In this paper we will discuss four very different approaches, all independent, to these secondary terms.

Shintani zeta functions (Taniguchi-T. [34, 35]).Taniguchi and the author further developed the theory of Shintani zeta functions, and gave proofs of Theorems 1.1 and 1.2, with error terms of $O(X^{7/9+\epsilon})$ and $O(X^{18/23+\epsilon})$, respectively. These proofs closely follow the heuristic arguments of Roberts and Datskovsky-Wright.

The Shintani zeta function approach has proved quite flexible, and in [34] we also prove a variety of generalizations of the main theorems. Perhaps the most interesting of these is a generalization to counting cubic field discriminants (or 3-torsion elements of class groups) in arithmetic progressions. For example, when counting cubic fields [[Add 'with discriminants'? I am not sure whether we need add this!]] $\equiv a \pmod 7$, the secondary term in Theorem 1.1 is different for every residue class $a \pmod 7$.

A refined geometric approach (Bhargava, Shankar, and Tsimerman [6]). Bhargava, Shankar, and Tsimerman gave the first proof of Theorem 1.1, with an error term of $O(X^{13/16+\epsilon})$. Their proof follows the lines of Davenport and Heilbronn's original work, counting lattice points via geometric arguments, and introduces several substantial simplifications and improvements. Indeed, these improvements are already evident in their proofs of the classical Davenport-Heilbronn theorems.

Their proof of the secondary term is based on a "slicing" argument which we describe briefly in Section 6. They also introduce a useful correspondence between cubic rings. This correspondence has also proved useful in the context of Shintani zeta functions, and a combined approach ([7], in progress) has yielded an error term of $O(X^{2/3+\epsilon})$.

Equidistribution of Heegner points (Hough [21]). Hough proves a statement closely related to Theorem 1.2. His scope is limited to counting 3-torsion elements in imaginary quadratic fields $\mathbb{Q}(\sqrt{D})$ with $D \equiv 2 \pmod 4$; in this context, he "proves" a version of Theorem 1.2, but with an error term larger than $X^{5/6}$. In followup work (in preparation), he proves the secondary term for a *smoothed* of this sum.

Hough's main results concern the equidistribution of the Heegner points associated to the 3-part of class groups of imaginary quadratic fields; his proof of the Davenport-Heilbronn theorem appears as a bonus, and without reference to binary cubic forms. His techniques also apply to the k-parts of these class groups, for odd $k \geq 5$, where he obtains secondary terms (of order $X^{1/2+1/k}$). His error terms are currently larger than both the secondary and the main terms for $k \geq 5$, but his work sheds light on a notoriously difficult problem.

Trigonal curves and the Maroni invariant (Zhao [40]). Zhao enumerates cubic extensions of the rational function field $\mathbb{F}_q(t)$ using algebraic geometry. Such extensions are in bijection with [[add 'isomorphic classes of ']] smooth trigonal curves, that is, smooth 3-fold covers of $\mathbb{P}^1$, and these may be counted by embedding them in [[Hirzebruch]] surfaces F_k. He obtains a secondary term as a consequence of a bound for the integer k, called the Maroni invariant; for now his error terms are larger than this secondary term.

In all of these approaches, the secondary term is easier to see than it is to prove. In particular, each of these approaches yield the secondary term in a natural way, but it is not *a priori* evident that the error terms can be made smaller than $X^{5/6}$!

In what follows, we will describe, insofar as we can in a couple of pages, why each of these approaches naturally yields a secondary term. For the Shintani zeta function approach, we will say only a very little bit about our treatment of the error term, and for the other three approaches we will say nothing about the error. Indeed, we will use the notation $O(\cdots)$ whenever the details are best left to the respective papers.

5. The Shintani Zeta Function Approach

Taniguchi and the author [34] proved Theorems 1.1 and 1.2 using the analytic theory of Shintani zeta functions. The *Shintani zeta functions* associated to the space of binary cubic forms are defined by the Dirichlet series

$$\xi^{\pm}(s) := \sum_{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash V_{\mathbb{Z}}} \frac{1}{|\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(x)|} |\mathrm{Disc}(x)|^{-s} = \sum_R \frac{1}{|\mathrm{Aut}(R)|} |\mathrm{Disc}(R)|^{-s}. \tag{5.1}$$

In the former sum, $V_{\mathbb{Z}}$ is the lattice defined in (3.1), the sum is over elements of positive or negative discriminant respectively, and $\mathrm{Stab}(x)$ is the stabilizer of x in $\mathrm{GL}_2(\mathbb{Z})$. The latter sum is over isomorphism classes of cubic rings. The equality of these two sums follows from the Delone-Faddeev correspondence (Theorem 3.1).

This is an example of a zeta function associated to a prehomogeneous vector space; in this case the word 'prehomogeneous' reflects the fact that $\mathrm{GL}_2(\mathbb{R})$ acts transitively on the positive- and negative-discriminant loci $V_{\mathbb{R}}^{\pm}$. Sato and Shintani [31] developed a general theory of zeta functions associated to prehomogeneous vector spaces, and for the space of binary cubic forms Shintani proved [32] that these zeta functions enjoy analytic continuation and an explicit functional equation (see [32] or [34] for details). These zeta functions have poles at $s = 1$, and, anomalously, at $s = 5/6$.

Shintani's work opens the door to the study of cubic fields using analytic number theory. In particular, by Perron's formula and standard techniques we have

$$\sum_{\substack{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash V_{\mathbb{Z}} \\ \pm \mathrm{Disc}(x) < X}} \frac{1}{|\mathrm{Stab}(x)|} = \int_{2-i\infty}^{2+i\infty} \xi^{\pm}(s) \frac{X^s}{s} ds = \mathrm{Res}_{s=1} \xi^{\pm}(s) X + \frac{6}{5} \mathrm{Res}_{s=5/6} \xi^{\pm}(s) X^{5/6} + O(X^{3/5+\epsilon}). \tag{5.2}$$

The left side is not the counting function of cubic fields, but for the first time *we see the $X^{5/6}$ secondary term.* This, therefore, gives an explanation for the secondary terms in Theorems 1.1 and 1.2: because the Shintani zeta functions have secondary poles.[6]

To count cubic fields, the most important step is sieving for maximality. This is possible by work of Datskovsky and Wright [37, 11, 12], who gave Shintani's work an adelic formulation, along the lines of Tate's thesis [36]. This allowed them to incorporate a variety of conditions into the definition of the Shintani zeta functions, including the Davenport-Heilbronn maximality conditions (given in Proposition 3.2).

A bogus proof of Roberts' conjecture is as follows. For a set of primes $\mathcal{P}$, define the *$\mathcal{P}$-maximal Shintani zeta function* by the Dirichlet series (5.1), with the added condition that the only cubic

[6] Of course, this begs the question of why the Shintani zeta functions have poles at $s = 5/6$. This can of course be explained by the various calculations in Shintani's work, but for now a satisfying highbrow argument is unknown to the author.

forms counted are those in the set U_p for each $p \in \mathcal{P}$. Therefore, letting $\mathcal{P}$ be the set of primes $< X$, it follows that
(5.3)

$$\sum_{\pm\mathrm{Disc}(x)<R}{}' \frac{1}{|\mathrm{Aut}(R)|} = \int_{2-i\infty}^{2+i\infty} \xi_{\mathcal{P}}^{\pm}(s)\frac{X^s}{s}ds = \mathrm{Res}_{s=1}\xi_{\mathcal{P}}^{\pm}(s)X + \frac{6}{5}\mathrm{Res}_{s=5/6}\xi_{\mathcal{P}}^{\pm}(s)X^{5/6} + O(X^{3/5+\epsilon}),$$

where the sum is over all *maximal* cubic rings[7], or equivalently cubic fields, with $\pm\mathrm{Disc}(R) < X$. Unfortunately the error term depends on the particular zeta function $\xi_{\mathcal{P}}$, and therefore the implied constant depends on X. Indeed, when carried out rigorously this argument yields an error term exponential in X.

However, this flawed argument has proved very useful. Datskovsky and Wright [12] used a variant of this argument to give an analytic proof of the Davenport-Heilbronn theorem. The key step[8] is to use (5.3) but with $\mathcal{P}$ equal to the set of all primes $< Y$, where Y tends to infinity very slowly with X. Given their adelic framework, their work yielded an asymptotic formula for the number of cubic extensions of any global field (of characteristic not 2 or 3).

Moreover, Roberts [29] computed the limit of the secondary terms in (5.3) and found an excellent match with numerical data. This led him to conjecture that (5.3) (equivalently, (1.1)) was correct with some undetermined error term smaller than $X^{5/6}$.

He closed his paper with the following (paraphrased very slightly):

> The pessimistic discussion in [12] suggests to us that the way may be difficult. However, one ingredient of a proof might be the functional equation of $\xi_{\mathcal{P}}^{\pm}(s)$ with respect to $s \to 1-s$, studied in [37, 11]. Another ingredient might be [[31], Thm. 3], which concerns growth of arithmetic functions whose associated Dirichlet series satisfy such a functional equation.

This is actually a good summary of our proofs of Theorems 1.1 and 1.2! To get the ball rolling, we needed to incorporate the sieve used by Belabas, Bhargava, and Pomerance in Section 3.1. In place of the $\mathcal{P}$-maximal zeta function we introduced the *q-nonmaximal zeta function*, which counts only cubic rings *not* maximal at each prime dividing q. [[Here again q is square free]] This allowed us to count maximal cubic rings, and the irreducible such rings correspond to cubic fields.

Remark. We count 3-torsion elements of class groups by a variation of this argument: Such elements correspond to cubic fields not totally ramified at any prime, and so we expand the definition of the q-nonmaximal zeta function to count cubic rings either nonmaximal or totally ramified at every prime dividing q.

In what follows we will oversimplify somewhat and describe what is "morally" true. The equation (5.3) holds for each q, and the q-dependence of the error term is given by the exponential sum

$$\sum_{x \in V_{\mathbb{Z}/q^2\mathbb{Z}}} |\widehat{\Phi}_q(x)| := \sum_{x \in V_{\mathbb{Z}/q^2\mathbb{Z}}} \left| \frac{1}{q^8} \sum_{y \in V_{\mathbb{Z}/q^2\mathbb{Z}}} \Phi_q(y)\exp(2\pi i[x,y]/q^2) \right|, \quad (5.4)$$

[7]Reducible maximal cubic rings correspond to fields of degree ≤ 2, which can be separately counted and therefore subtracted with good error terms.

[8]We have adapted their argument somewhat to our point of view. Datskovsky and Wright also need, as did Davenport and Heilbronn, a bound for the number of cubic rings nonmaximal at any prime $> Y$, and they need a separate estimate for the number of quadratic fields K with $|\mathrm{Disc}(K)| < X$.

where $\Phi_q(x)$ is the characteristic function of the Davenport-Heilbronn q-nonmaximality condition, and

$$[x, y] = x_4 y_1 - \frac{1}{3} x_3 y_2 + \frac{1}{3} x_2 y_3 - x_1 y_4. \tag{5.5}$$

Each inner sum appears naturally when we shift the contour in (5.3) to the left of $\Re(s) = 0$ and apply the functional equation. Trivially, the sum in (5.4) is bounded by q^8; this is enough to allow us to take $Q = X^{1/25}$ in (3.5) and obtain an error term of $X^{24/25+\epsilon}$ in the Davenport-Heilbronn theorems.

This, then, was the starting point of [34], and we incorporated several improvements to lower the error terms in Theorems 1.1 and 1.2 to $O(X^{7/9+\epsilon})$ and $O(X^{18/23+\epsilon})$, respectively. One of these improvements is an analysis of the sum in (5.4), carried out in [35]. For each x we obtained exact formulas for the inner sum in (5.4), proving that the outer sum is $\ll q^{1+\epsilon}$.

6. A Refined Geometric Approach

We begin by describing Bhargava, Shankar, and Tsimerman's proof [6] of the main term in Theorem 1.1. Later, we will describe how the secondary term appears in their work, but only in the context of irreducible cubic forms (corresponding to cubic orders). Finally, we describe a correspondence for cubic forms which they use to improve their error terms.

The argument in [6] largely follows Davenport and Heilbronn's original work. Fix a sign, and let $V_{\text{irr}}^{\pm}$ denote the irreducible points $x \in V_{\mathbb{Z}}$ with $\pm\text{Disc}(x) > 0$. Write $n_+ = 6$, $n_- = 2$ for the order of the stabilizers of the action of $\text{GL}_2(\mathbb{R})$ on $V_{\mathbb{R}}^{\pm}$. The equality $n_+ = 3n_-$ reflects the fact that cubic fields of negative discriminant are three times as common as those of positive discriminant.

Write $\mathcal{F}$ for a certain fundamental domain for $\text{GL}_2(\mathbb{Z})\backslash\text{GL}_2(\mathbb{R})$ in $\text{GL}_2(\mathbb{R})$ (see (12) of [6]), originally constructed by Gauss. For any $z \in V^{\pm}$, consider the multiset $\mathcal{F}v$, where the multiplicity of an element $x \in V^{\pm}$ is equal to the number of $g \in \mathcal{F}$ for which $gv = x$. It is readily checked that each element $x \in G_{\mathbb{Z}}\backslash V_{\mathbb{Z}}^{\pm}$ is represented in this multiset $n_i/m(x)$ times, where $m(x)$ denotes the size of the stabilizer of x in $\text{GL}_2(\mathbb{Z})$. We conclude that

$$N(V^{\pm}; X) := \sum_{\substack{0 < \pm\text{Disc}(x) < X \\ x \ irred.}} \frac{1}{|\text{Stab}(x)|} = \frac{1}{n_i} \{x \in \mathcal{F}v \cap V_{\text{irr}}^{\pm} \ : \ |\text{Disc}(x)| < X\}. \tag{6.1}$$

The irreducible[[should be 'reducible']] points can be handled without too much difficulty: The number of [[reducible]] integral binary cubic forms $au^3 + bu^2v + cuv^2 + dv^3$ in the multiset $\mathcal{R}_X(v) := \{w \in \mathcal{F}v \ : \ |\text{Disc}(w)| < X\}$, satisfying $a \neq 0$, is $\ll X^{3/4+\epsilon}$. Conversely, all cubic forms with $a = 0$ are reducible. This condition $a \neq 0$ meshes well with the geometric arguments that follow.

In addition, the weight $\frac{1}{|\text{Stab}(x)|}$ occurring in (6.1) can be neglected, as there are $\ll X^{1/2}$ points x with $|\text{Disc}(x)| < X$ with nontrivial stabilizer.

The formula in (6.1) does not depend on v, and one innovation of [6] is to average over many v. In particular, define $B := \{w = (a, b, c, d) \in V : \ 3a^2 + b^2 + c^2 + 3d^2 \leq 10, \ |\text{Disc}(w)| \geq 1\}$, and we have

$$N(V^{\pm}; X) = \frac{\int_{v \in B \cap V^{\pm}} \frac{1}{n_i} \{x \in \mathcal{F}v \cap V_{\text{irr}}^{\pm} \ : \ |\text{Disc}(x)| < X\} |\text{Disc}(v)|^{-1} dv}{\int_{v \in B \cap V^{\pm}} |\text{Disc}(v)|^{-1} dv}. \tag{6.2}$$

The authors describe this step as "thickening the cusp". As $|\mathrm{Disc}(v)|^{-1}[[|\mathrm{Disc}(v)|^{-1}dv]]$ is $\mathrm{GL}_2(\mathbb{R})$-invariant, we may rewrite this as

$$N(V^{\pm};X) = M^{\pm}\int_{g\in\mathcal{F}} \#\{x \in V_{\mathrm{irr}}^{\pm}\cap gB : |\mathrm{Disc}(x)| < X\}dg, \tag{6.3}$$

for a certain constant $M^{\pm}$ depending on B. This is rewritten as

$$N(V^{\pm},X) = M^{\pm}\int_{g\in N'(a)A'\Lambda} \#\{x \in V_{\mathrm{irr}}^{\pm}\cap B(n,t,\lambda,X)\}t^{-2}dnd^{\times}td^{\times}\lambda \tag{6.4}$$

using a standard decomposition of $\mathcal{F}$ ((12) of [6]), where $B(n,t,\lambda,X)$ is the region $\{x \in gB : |\mathrm{Disc}(x)| < X\}$. A proposition of Davenport establishes that in this situation, the count of lattice points above is well approximated by the volume of $B(n,t,\lambda,X)$, provided that t is not too large. Conversely, when t is large all the lattice points are in the cusp $a = 0$, and hence reducible and not counted. Thus, the above is reduced to

$$N(V^{\pm},X) = M^{\pm}\int_{\substack{g\in N'(a)A'\Lambda\\ t<C^{1/3}\lambda^{1/3}}} \mathrm{Vol}(B(n,t,\lambda,X))t^{-2}dnd^{\times}td^{\times}\lambda + O(\cdots). \tag{6.5}$$

The condition on t is now removed, subject to an error term, and this integral is equal to

$$N(V^{\pm},X) = \frac{1}{2\pi M^{\pm}}\int_{v\in B\cap V^{\pm}} \mathrm{Vol}(\mathcal{R}_X(v))|\mathrm{Disc}(v)|^{-1}dv + O(\cdots). \tag{6.6}$$

This volume does not in fact depend on v, so that we have (excluding error terms)

$$N(V^{\pm},X) = \frac{1}{n_i}\mathrm{Vol}(\mathcal{R}_X(v)) = \frac{\pi^2}{12n_i}X. \tag{6.7}$$

6.1. **Origin of the secondary term.** We now explain how Bhargava, Shankar, and Tsimerman refine these calculations to obtain a secondary term. Our brief explanation will necessarily be somewhat vague, which should encourage the reader to read [6]. We focus on their count of irreducible binary cubic *forms*; as in [1, 34], they also incorporate a sieve to count cubic fields.

They begin by incorporating several tweaks to (6.4). For technical reasons, they count discriminants in dyadic intervals $[X/2,X]$. In addition, they introduce a smooth function $\Psi_0(t)$ on $\mathbb{R}_{\geq 0}$, such that $\Psi_0(t) = 0$ for $t \leq 2$ and $\Psi_0(t) = 1$ for $t \geq 3$. They thus rewrite (6.4) (restricted to a dyadic interval $[X/2,X]$) as

$$N(V^{\pm},[X/2,X]) = \frac{1}{M^{\pm}}\int_{g\in N'(a)A'\Lambda}\left(\Psi_0\left(\frac{t\kappa}{\lambda^{1/3}}\right) + \Psi\left(\frac{t\kappa}{\lambda^{1/3}}\right)\right)\times \#\{x \in V_{\mathrm{irr}}^{\pm}\cap B(n,t,\lambda,[X/2,X])t^{-2}dnd^{\times}td^{\times}\lambda, \tag{6.8}$$

where $\Psi := 1 - \Psi_0$, and κ is a parameter to be chosen later (to minimize error terms). The decomposition $\Psi + \Psi_0 = 1$ splits this integral into two, and we restrict our attention to the Ψ_0 part, which yields the secondary term.

They now "slice" the count of binary cubic forms by the first coordinate. In particular, for $a \in \mathbb{Z}$, let $B_a(n,t,\lambda,[X/2,X])$ denote the set of binary cubic forms in $B_a(n,t,\lambda,[X/2,X])$ whose u^3 coefficient is equal to a. Then, we have

$$\#\big\{x \in V_{\mathrm{irr}}^{\pm}\cap B(n,t,\lambda,[X/2,X])\big\} = \sum_{\substack{a\in\mathbb{Z}\\ a\neq 0}}\#\big\{x \in V_{\mathrm{irr}}^{\pm}\cap B_a(n,t,\lambda,[X/2,X])\big\}. \tag{6.9}$$

Skipping ahead in their work, the Ψ_0-contribution to (6.8) is equal to

$$\frac{2}{3M^{\pm}}\sum_{a=1}^{\infty}\int_{\lambda=(\sqrt{3/2})^3/C}^{X^{1/4}}\int_{u>0}\Phi_0\Big(\frac{u^{1/3}\kappa}{a^{1/3}}\Big)\frac{\lambda^{10/3}u^{1/3}}{a^{1/3}}\mathrm{Vol}\big(B_u([X/(2\lambda^4),X/\lambda^4])\big)d^{\times}ud^{\times}\lambda, \tag{6.10}$$

where $B_a([Y/2,Y])$ denotes the set of cubic forms in B with first coordinate a and discriminant in $[Y/2,Y]$. By Mellin inversion, we have

$$\sum_{a=1}^{\infty}a^{-1/3}\Phi_0\Big(\frac{u^{1/3}\kappa}{a^{1/3}}\Big)=\zeta(1/3)+3\widetilde{\Phi}_0(-2)(\kappa^3u)^{2/3}+O(\cdots), \tag{6.11}$$

and it is this (negative) $\zeta(1/3)$ which contributes the secondary term. In contrast, the second term of (6.11) (which "looks larger", especially in light of the eventual choice $\kappa=X^{1/12}$) contributes a term of order X. This second term is incorporated into (6.10) and then combined with the Ψ term of (6.8) to obtain the main term in (6.7). In contrast, when the $\zeta(1/3)$ term is plugged into (6.10), the resulting integral can be evaluated without undue difficulty, and it has order $X^{5/6}$.

Remark. As Shankar has explained to the author, their original proof was more complicated but more geometrically intuitive. In particular, the above argument relies on the analytic continuation of the zeta function to show that the secondary term is negative, but this can also be "seen" by studying the geometry of the cusp.

6.2. **A correspondence for cubic forms.** To count cubic fields, Bhargava, Shankar, and Tsimerman need to sieve for maximality, and so they must introduce the Davenport-Heilbronn conditions given in Proposition 3.2. These could be treated in a naive manner, but the authors improve their error terms by introducing a useful correspondence for cubic forms.[9]

We introduce some notation. Let $N^{\pm}(V_{\mathbb{Z}};X)$ count all cubic forms v with $0<\pm\mathrm{Disc}(v)<X$. For a prime p, let $N^{\pm}(U_p^c;X)$ count those cubic forms which are nonmaximal at p, i.e., which do not lie in the set U_p defined in Proposition 3.2. (In referring to cubic forms as "nonmaximal" we are appealing to the Delone-Faddeev correspondence.) Finally, for any $\alpha\in\mathbb{P}^1(\mathbb{Z}/p\mathbb{Z})$, let $N^{\pm}(V_{p,\alpha};X)$ count those cubic forms v such that the reduction of v modulo p has α as a root.

With this notation, we have the following correspondence:

Proposition 6.1. [6] *We have the identity*

$$N^{\pm}(U_p^c;X)=\sum_{\alpha\in\mathbb{P}^1(\mathbb{F}_p)}N^{\pm}(V_{p,\alpha};X/p^2)-\sum_{\alpha\in\mathbb{P}^1(\mathbb{F}_p)}N^{\pm}(V_{p,\alpha};X/p^4)+N^{\pm}(V_{\mathbb{Z}};X/p^4). \tag{6.12}$$

This identity generalizes in a straightforward way from prime p to squarefree q, where we count cubic forms not in U_p for any $p|q$. The identity amounts to a combinatorial argument, enumerating ways in which a nonmaximal cubic ring can be replaced with a larger ring.

This identity is a crucial step in [6], and it can also be translated into an identity for the q-nonmaximal Shintani zeta function! In work in progress by Bhargava, Taniguchi, and the author [7] we are developing this combined approach. We have proved an error term of $O(X^{2/3+\epsilon})$ in Theorem 1.1 as well as a secondary term for relative cubic extensions of quadratic base fields, and we are currently working on further extensions and generalizations.

[9] As Shankar explained to me, in the case of cubic fields it is this correspondence which allows them to cross the $X^{5/6}$ barrier.

7. Equidistribution of Heegner Points

Remarkably, Hough [21] demonstrated that the Delone-Faddeev correspondence is not the only path to the Davenport-Heilbronn theorem. Hough studied the distribution of Heegner points associated to 3-torsion elements of ideal class groups of quadratic fields, and he obtained a result related to Theorem 1.2 as a consequence.

For simplicity, Hough worked only with imaginary discriminants $D \equiv 2 \pmod 4$; note that any such discriminant is not fundamental,[[Why this is not fundamental? Use the defination in Bhargava, Belbas..., $4D$ is a discriminant, so it is fundamental. Did I say something stupid?]] and the discriminant of the field $\mathbb{Q}(\sqrt{D})$ is $4D$.[10] He proves that

$$\sum_{\substack{0<D<X \\ D\equiv 2 \pmod 4}} \#\mathrm{Cl}_3(-D) = \frac{4}{\pi^2}X + O(X^{19/20+\epsilon}), \tag{7.1}$$

and for "good"[11] test functions ϕ, he proves that

$$\sum_{\substack{0<D \\ D\equiv 2 \pmod 4}} \#\mathrm{Cl}_3(-D)\phi(D/X) = \frac{4}{\pi^2}\widehat{\phi}(1)X + O(X^{7/8+\epsilon}). \tag{7.2}$$

The correct secondary term of order $X^{5/6}$ appears in both formulas, in spite of the larger error terms, and in in followup work (in progress) he obtains an error less than $X^{5/6}$ in (7.2), thereby obtaining another proof of a secondary term.

Hough's work has very different prospects for generalization from the approaches described previously. In particular, his approach can be used to study k-torsion in class groups for odd $k > 3$, and he makes the following conjecture:

Conjecture 1 (Hough [21])**.** *For good test functions ϕ and odd $k \geq 3$, we have*

$$\sum_{\substack{0<D \\ D\equiv 2 \pmod 4}} \#\mathrm{Cl}_k(-D)\phi(D/X) = \frac{4}{\pi^2}\widehat{\phi}(1)X + C_{1,k}\widehat{\phi}\left(\frac{1}{2}+\frac{1}{k}\right)X^{1/2+1/k} + o(X^{1/2+1/k}), \tag{7.3}$$

where

$$C_{1,k} := \frac{1}{6k}\frac{\zeta(1-\frac{2}{k})}{\zeta(2)}\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}-\frac{1}{k})}{\Gamma(1-\frac{1}{k})}\left(1-2^{1/k}+2^{1-1/k}\right)\times \prod_{p>2}\left(1+\frac{1}{p+1}\left(p^{-1/k}-p^{-1+2/k}-p^{-1+1/k}-p^{-1}\right)\right). \tag{7.4}$$

Indeed he "proves" his conjecture, naturally obtaining the two main terms above but (for now) obtaining error terms larger than X. Note that obtaining even the main term for the single case $k = 5$ would be a major achievement. Indeed, the average size of $\mathrm{Cl}_p(-D)$ was conjectured by Cohen and Lenstra [8] to be 2 for each prime $p \geq 3$, and Cohen and Lenstra further conjectured that the p-part of $\mathrm{Cl}_p(-D)$ is isomorphic to a fixed p-group H with probability proportional to

[10] It seems that the restrictions that $D \equiv 2 \pmod 4$ and that D be negative could both be lifted with some effort. The restriction on the sign is the more serious of the two, as elements of class groups of $\mathbb{Q}(\sqrt{D})$ for positive D correspond to closed geodesics rather than Heegner points. However, Duke's result holds for either sign, so in [21] Hough expresses optimism that the positive discriminant case of Theorem 1.2 could be handled as well.

[11] Compact support and infinitely differentiable.

$\frac{1}{|\mathrm{Aut}(H)|}$. No case of these conjectures is currently known for any $p \geq 5$, and essentially nothing is known for $p = 3$ beyond Theorem 1.2.

7.1. **Heegner points and equidistribution.** Hough's main result concerns the equidistribution of Heegner points associated to the 3-part of the class group. We recall the relevant background. Suppose that $\mathfrak{a}$ is an ideal of an imaginary quadratic field $\mathbb{Q}(\sqrt{-D})$. Then writing $\mathfrak{a} = [x_1, x_2]$ as a $\mathbb{Z}$-module, where reordering if necessary we ask that $x_1/x_2 \in \mathbb{H} := \{z \in \mathbb{C} \ : \ \Im(z) > 0\}$, we define x_1/x_2 to be the *Heegner point* $z_{\mathfrak{a}}$ *associated to the ideal* $\mathfrak{a}$. A change of basis for $\mathfrak{a}$ corresponds to a linear fractional transformation $z_{\mathfrak{a}} \to \frac{az_{\mathfrak{a}}+b}{cz_{\mathfrak{a}}+d}$ with $ad - bc = 1$, and therefore the Heegner point $z_{\mathfrak{a}}$ is well-defined in the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. One checks that $z_{\mathfrak{a}} = z_{\alpha\mathfrak{a}}$ for any $\alpha \in \mathbb{Q}(\sqrt{-D})$, and hence $z_{\mathfrak{a}}$ depends only on the class of $\mathfrak{a}$ in $\mathrm{Cl}(\sqrt{-D})$. One similarly checks that $z_{\mathfrak{a}} \neq z_{\mathfrak{b}}$ if $\mathfrak{a}$ and $\mathfrak{b}$ are inequivalent in $\mathrm{Cl}(\sqrt{-D})$.

Therefore, there are $h(-D)$ Heegner points associated to $\mathbb{Q}(\sqrt{-D})$ in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, and in [15] Duke proves that they become equidistributed with respect to the hyperbolic measure $\frac{dx\,dy}{y^2}$ as $D \to \infty$. Hough proves the same for the Heegner points associated to the 3-part of the ideal class group:

Theorem 7.1 (Hough [21])**.** *Let B be a bounded Borel measurable subset of $\mathbb{H}$ having boundary of measure zero. Then as $D \to \infty$,*

$$\sum_{\substack{0>-D>-X \\ d\equiv 2 \pmod 4 \\ \text{squarefree}}} \#\big\{\mathfrak{a} \text{ primitive, nonprincipal} : [\mathfrak{a}] \in Cl_3(\mathbb{Q}(\sqrt{-D})), z_{\mathfrak{a}} \in B\big\} \sim \frac{6X}{\pi^3}\mathrm{vol}(B). \tag{7.5}$$

Here we take $B \in \mathbb{H}$ rather than $B \in \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, so that each Heegner point is counted once for each representative of its $\mathrm{SL}_2(\mathbb{Z})$ orbit in B.

Straightaway one can see that this is an equidistribution statement. But it also implies the Davenport-Heilbronn theorem! Taking B to be the standard (or any other) fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$, one counts each Heegner point (and thus each nontrivial element of $\mathrm{Cl}_3(\mathbb{Q}(\sqrt{-D}))$) exactly once. The Davenport-Heilbronn theorem follows from the classical computation $\mathrm{vol}(B) = \frac{3}{\pi}$.

We will follow Hough's approach sufficiently far as to see the secondary term. His work begins with the following parameterization of ideals (and therefore Heegner points), essentially due to Soundararajan [33].

Proposition 7.2. [21, 33] *Let $D \equiv 2 \pmod 4$ be squarefree and let $k \geq 3$ be odd. The set*

$$\{(l, m, n, t) \in (\mathbb{Z}^+)^4 \ : \ lm^k = l^2n^2 + t^2D, \ l|D, \ (m, ntD) = 1\} \tag{7.6}$$

is in bijection with primitive ideal pairs $(\mathfrak{a}, \overline{\mathfrak{a}})$ with $\mathfrak{a} \neq 1$ and $\mathfrak{a}^k$ principal in $\mathbb{Q}(\sqrt{-D})$. Explicitly, the ideal $\mathfrak{a}$ is given as a $\mathbb{Z}$-module by

$$\mathfrak{a} = [lm, lnt^{-1} + \sqrt{-D}]$$

where $\mathbb{N}\mathfrak{a} = lm$ and t^{-1} is the inverse of t modulo m.

The proof is relatively straightforward. For example, if $\mathfrak{a}^k$ is principal and coprime to D, then $\mathfrak{a}^k = (n + t\sqrt{-D})$ and taking norms yields $(\mathbb{N}\mathfrak{a})^k = n^2 + t^2D$, a solution to (7.6). Ideals $\mathfrak{a}$ not coprime to D yield solutions to (7.6) with $l > 1$.

As one of the two main steps in his proof, Hough now applies this parameterization to count Heegner points in the region

$$R_Y := \Big\{z \in \mathbb{H} \ : -\frac{1}{2} < \Im(z) < \frac{1}{2}, \ \ \Re(z) > \frac{1}{Y}\Big\}. \tag{7.7}$$

[[Should it be $\left\{z \in \mathbb{H} \ : -\frac{1}{2} < \Re(z) < \frac{1}{2}, \ \ \Im(z) > \frac{1}{Y}\right\}$?]]
(This establishes the vertical distribution of Heegner points, and he separately establishes that they are equidistributed horizontally.) He proves that the number of Heegner points in R_Y is equal to

$$\frac{6}{\pi^3} YX + C_{5/6} X^{5/6} + O(\cdots), \tag{7.8}$$

for a fairly complicated error term which depends on both X and Y. The secondary term appears here for the following reason: By the parameterization above, we may write an ideal $\mathfrak{a}$ as $[lm, lnt^{-1} + \sqrt{-D}]$. We have $lm^3 = l^2n^2 + t^2 D$, so that $lm \geq D^{1/3}$, implying that the Heegner point $z_{\mathfrak{a}}$ has imaginary part $\frac{\sqrt{D}}{lm} \leq D^{1/6} \leq X^{1/6}$. In other words, the Heegner points equidistribute as $X \to \infty$, but for fixed X they do not appear high in the cusp.

The hyperbolic volume of the subset of R_Y with $\Im(z) > X^{1/6}$ is equal to

$$\int_{x=-1/2}^{1/2} \int_{y=O(D^{1/6})}^{\infty} \frac{dx\ dy}{y^2} = X^{-1/6}. \tag{7.9}$$

Accounting (much more carefully than done here!) for the fact that this subset contains no Heegner points yields the secondary negative term in (7.8), and an analogous argument explains the second term in (7.3).

8. Hirzebruch Surfaces and the Maroni Invariant

Zhao's work [40] is still in preparation, so we offer only a very brief overview here. Using algebraic geometry, Zhao estimates the number of cubic extensions of the *rational function field* $\mathbb{F}_q(t)$.

Fix a finite field $\mathbb{F}_q$ of characteristic not equal to 2 or 3, and let $S(2N)$ be the number of isomorphism classes of cubic extensions $K/\mathbb{F}_q(t)$ of degree 3, such that $|\mathrm{Disc}(F)| = q^{2N}$. In [12], Datskovsky and Wright proved that

$$S(2N) = 2\frac{\mathrm{Res}_{s=1}\zeta_{\mathbb{F}_q(t)}(1)}{\zeta_{\mathbb{F}_q(t)}(3)} q^{2N} + o(2N), \tag{8.1}$$

$$[[S(2N) = 2\frac{\mathrm{Res}_{s=1}\zeta_{\mathbb{F}_q(t)}(s)}{\zeta_{\mathbb{F}_q(t)}(3)} q^{2N} + o(q^{2N})]]$$

where $\zeta_{\mathbb{F}_q(t)}(s)$ is the Dedekind zeta function of $\mathbb{F}_q(t)$, or equivalently the zeta function of the algebraic curve $\mathbb{P}^1_{\mathbb{F}_q}$.

Datskovsky and Wright's proof uses the adelic Shintani zeta function; and in [40] Zhao obtains another proof of (8.1) using algebraic geometry. A *trigonal curve* is a 3-fold cover of $\mathbb{P}^1$, and there is a bijection between [[isomorphic classes of]] smooth [[trigonal]] curves and cubic rational function fields. Already this perspective is enlightening; for example, the Riemann-Hurwitz formula implies that the discriminant has even degree, explaining why (8.1) is an equation for $S(2N)$ and not $S(N)$.

The problem, then, is to count smooth trigonal curves. These may be counted by embedding the curves into *Hirzebruch surfaces* F_k; for each trigonal curve C there is a unique nonnegative integer k such that C embeds into [[$F_k \equiv \mathbb{P}_{\mathbb{P}^1}(\mathcal{O} \oplus \mathcal{O}(-k))$. This integer k is called the Maroni Invariant of the trigonal curve.]]

Zhao now counts curves on each surface F_k and sieves for smoothness, using [[a seive method]] related to those of Poonen [28]. [[However, the main techinque is different and new.]] The main term in (8.1) comes from estimating [['summing' is better]] the results for all [[possible]] integers k. However, the Maroni invariant of a [[irreducible trigonal]] curve C [[with fixed brach degree i.e.

fixed absolute discriminant]] can be given a sharp upper bound in terms of the discriminant [[q^{2N}]], and limiting our sum to the F_k for

$$[[0 \leq k \leq \frac{N}{3}]]$$

naturally introduces a secondary term of order $q^{5N/3}$ into (8.1)!

Zhao's method, like all other methods described in this paper, comes with error terms; at present he has obtained an error term of $O(q^{7N/4})$ in (8.1). He has some optimism that this error term can be reduced below $q^{5N/3}$, thereby proving the secondary term.

9. Conclusion

Cubic fields have seen a great deal of recent study recently; we particularly recommend the interesting papers of Cohen and Morra [9] and Martin and Pollack [23] for recent studies of cubic fields from different perspectives.

Many interesting open questions remain in the subject. To name one, what is the "correct" error term in the Davenport-Heilbronn theorems? Even if we cannot prove it, it would still be interesting to determine the expected order of magnitude. To get the best error terms it seems that we should count fields K of degree *at most* 3, each weighted by $\frac{1}{|\mathrm{Aut}(K)|}$. The data in [29] suggests that the true error may be smaller than $X^{1/2}$, and a comparison with the divisor problem suggests that the error might be on the order of $X^{3/8}$. Nevertheless, the comparison with the divisor problem is not exact, and the numerical data is inconclusive.

Perhaps the most compelling open question concerns quartic and quintic fields. Asmyptotic formulas were proved by Bhargava [2, 5]; should these formulas have secondary terms? It is generally believed that they should; however, so far we lack even a good conjecture. The zeta function approach seems likely to work, and Sato and Shintani [31] and Yukie [39] have made some progress in this direction; the geometric approach seems likely to work as well. However, for now the difficulties with either approach appear rather severe.

Another open question concerns the *multiplicity* of cubic fields. It is believed that there should be $\ll n^{\epsilon}$ cubic fields of discriminant $\pm n$, but the best bound known is $O(n^{1/3+\epsilon})$, due to Ellenberg and Venkatesh [17]. Nontrivial bounds were also obtained by Helfgott and Venkatesh [20] and Pierce [27], using a variety of methods. Any of the methods described in the present paper could potentially yield improvements, but none of them have succeeded to date. (The present author has attempted improvements by means of Shintani zeta functions, which have led to a number of instructve and interesting failures.)

Finally, this paper begs the question of whether unexpected connections might be found between the four perspectives presented here, and some promising preliminary results have been obtained in this direction. We anticipate that this and further related questions will be addressed in the near future.

10. Acknowledgments

I would like to thank the many people who have shared their insights on these secondary terms, especially Manjul Bhargava, Jordan Ellenberg, Bob Hough, Arul Shankar, Kannan Soundararajan, Takashi Taniguchi, Melanie Matchett Wood, and Yongqiang Zhao. Most of all I would like to thank Akshay Venkatesh who originally brought this topic to my attention.

In addition, I would like to thank the organizers of Integers 2011 for hosting an outstanding conference.

References

[1] K. Belabas, M. Bhargava, and C. Pomerance, *Error estimates in the Davenport-Heilbronn theorems*, Duke Math. J. **153** (2010), no. 1, 173–210.

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

[3] M. Bhargava, *Mass formulae for extensions of local fields, and conjectures on the density of number field discriminants*, Int. Math. Res. Not. (2007), no. 17, Art. ID rnm052, 20 pp.

[4] M. Bhargava, *Higher composition laws and applications*, Proceedings of the International Congress of Mathematicians, Vol. II, 271??94, Eur. Math. Soc., Zürich, 2006.

[5] M. Bhargava, *The density of discriminants of quintic rings and fields*, Ann. Math. (2) **172** (2010), no. 3, 1559-1591.

[6] M. Bhargava, A. Shankar, and J. Tsimerman, *On the Davenport-Heilbronn theorem and second order terms*, preprint; available at `http://arxiv.org/abs/1005.0672`.

[7] M. Bhargava, T. Taniguchi, and F. Thorne, work in preparation.

[8] H. Cohen and H. Lenstra, *Heuristics on class groups of number fields*, Number theory, Noordwijkerhout 1983, 33–62, Lecture Notes in Math., 1068, Springer, Berlin, 1984.

[9] H. Cohen and A. Morra, *Counting cubic extensions with given quadratic resolvent*, J. Algebra **325** (2011), 461–478.

[10] H. Cohen, F. Diaz y Diaz, and M. Olivier, *Counting discriminants of number fields*, J. Théor. Nombres Bordeaux **18** (2006), no. 3, 573–593.

[11] B. Datskovsky and D. Wright, *The adelic zeta function associated to the space of binary cubic forms. II. Local theory*, J. Reine Angew. Math. **367** (1986), 27–75.

[12] B. Datskovsky and D. Wright, *Density of discriminants of cubic extensions*, J. Reine Angew. Math. **386** (1988), 116–138.

[13] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic fields. II*, Proc. Roy. Soc. London Ser. A **322** (1971), no. 1551, 405–420.

[14] B. N. Delone and D. K. Faddeev, *The theory of irrationalities of the third degree (English translation*, AMS, Providence, 1964.

[15] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math., **92** (1):73–90, 1988.

[16] J. Ellenberg and A. Venkatesh, *The number of extensions of a number field with fixed degree and bounded discriminant*, Ann. of Math. **163** (2006), no. 2, 723–741.

[17] J. Ellenberg and A. Venkatesh, *Reflection principles and bounds for class group torsion*, Int. Math. Res. Not. no. 1 (2007), Art. ID rnm002.

[18] W. T. Gan, B. Gross, and G. Savin, *Fourier coefficients of modular forms on G_2*, Duke Math. J. **115** (2002), 105–169.

[19] E. Golod and I. Shafarevich, *On the class field tower* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **28** (1964), 261–272.

[20] H. Helfgott and A. Venkatesh, *Integral points on elliptic curves and 3-torsion in class groups*, J. Amer. Math. Soc. **19** (2006), no. 3, 527–550.

[21] B. Hough, *Equidistribution of Heegner points associated to the 3-part of the class group*, preprint; available at `http://arxiv.org/abs/1005.1458`.

[22] F. Lemmermeyer, *Class field towers*, `http://www.rzuser.uni-heidelberg.de/~hb3/publ/pcft.pdf`.

[23] G. Martin and P. Pollack, *The average least character nonresidue and further variations on a theme of Erdős*, preprint; available at `http://arxiv.org/abs/1112.1175`.

[24] J. Martinet, *Petits discriminants*, Ann. Inst. Fourier **29** (1979), 159–170.

[25] J. Neukirch, *Algebraic number theory*, Springer-Verlag, Berlin, 1999.

[26] A. Odlyzko, *Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: a survey of recent results.*, Sém. Théor. Nombres Bordeaux (2) **2** (1990), no. 1, 119–141.

[27] L. Pierce, *The 3-part of class numbers of quadratic fields*, J. London Math. Soc. (2) **71** (2005), no. 3, 579–598.

[28] B. Poonen, *Bertini theorems over finite fields*, Ann. Math. (2) **160** (2004), no. 3, 1099??127.

[29] D. Roberts, *Density of cubic field discriminants*, Math. Comp. **70** (2001), no. 236, 1699–1705.

[30] M. Sato and T. Kimura, *A classification of irreducible prehomogeneous vector spaces and their relative invariants.* Nagoya Math. J. **65** (1977), 1–155.

[31] M. Sato and T. Shintani, *On zeta functions associated with prehomogeneous vector spaces*, Ann. of Math. (2) **100** (1974), 131–170.

[32] T. Shintani, *On Dirichlet series whose coefficients are class numbers of integral binary cubic forms*, J. Math. Soc. Japan **24** (1972), 132–188.

[33] K. Soundararajan, *Divisibility of class numbers of imaginary quadratic fields*, J. London Math. Soc. (2) **61** (2000), 681–690.

[34] T. Taniguchi and F. Thorne, *Secondary terms in counting functions for cubic fields*, submitted; available at `http://arxiv.org/abs/1102.2914`.

[35] T. Taniguchi and F. Thorne, *Orbital L-functions for the space of binary cubic forms*, submiited; available at `http://arxiv.org/abs/1112.5030`.

[36] J. Tate, *Fourier analysis in number fields and Hecke's zeta-functions*, thesis (Princeton, 1950); reprinted in J. W.S. Cassels and A. Fröhlich, eds., *Algebraic number theory,* Academic Press, London, 1986.

[37] D. Wright, *The adelic zeta function associated to the space of binary cubic forms. I. Global theory*, Math. Ann. **270** (1985), no. 4, 503–534.

[38] D. Wright, *Twists of the Iwasawa-Tate zeta function*, Math. Z. **200** (1989), 209–231.

[39] A. Yukie, *Shintani zeta functions*, London Mathematical Society Lecture Note Series, 183, Cambridge University Press, Cambridge, 1993.

[40] Y. Zhao, doctoral thesis, University of Wisconsin, in preparation.

Department of Mathematics, University of South Carolina, 1523 Greene Street, Columbia, SC 29208

E-mail address: `thorne@math.sc.edu`