Higher composition laws and applications

Manjul Bhargava*

Abstract. In 1801 Gauss laid down a remarkable law of composition on integral binary quadratic forms. This discovery, known as *Gauss composition*, not only had a profound influence on elementary number theory but also laid the foundations for ideal theory and modern algebraic number theory. Even today, Gauss composition remains one of the best ways of understanding ideal class groups of quadratic fields.

The question arises as to whether there might exist similar laws of composition on other spaces of forms that could shed light on the structure of other algebraic number rings and fields. In this article we present several such higher analogues of Gauss composition, and we describe how each of these composition laws can be interpreted in terms of ideal classes in appropriate rings of algebraic integers. We also discuss several applications of these composition laws, including the resolution of a critical case of the Cohen–Lenstra–Martinet heuristics, and a solution of the long-standing problem of counting the number of quartic and quintic fields of bounded discriminant. In addition, we describe the mysterious relationship between these various composition laws and the exceptional Lie groups. Finally, we discuss prospects for future work and conclude with several open questions.

1. Introduction

Gauss published his seminal treatise *Disquisitiones Arithmeticae* in 1801. One of the primary subjects of this work was the (integral) *binary quadratic form*, i.e., any expression $f(x, y) = ax^2 + bxy + cy^2$ where $a, b, c \in \mathbb{Z}$.[1] The group $\mathrm{SL}_2(\mathbb{Z})$ acts naturally on the space of binary quadratic forms by linear substitution of variable: if $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, then one defines

$$(\gamma \cdot f)(x, y) = f((x, y)\gamma).$$

Gauss studied this action of $\mathrm{SL}_2(\mathbb{Z})$ on binary quadratic forms f in terms of the *discriminant* $\mathrm{Disc}(f) = b^2 - 4ac$, as it is easily seen that this discriminant remains

---

[1]Gauss actually considered only the forms where b is even; however, from the modern point of view it is more natural to assume a, b, c are arbitrary integers.

invariant under the action of $SL_2(\mathbb{Z})$. In fact, one can show that *any* polynomial $P(a, b, c)$ invariant under the action of $SL_2(\mathbb{Z})$ on the space of binary quadratic forms $ax^2 + bxy + cy^2$ must be a polynomial in the discriminant $b^2 - 4ac$ (see e.g. [28]).

It follows that the binary quadratic forms of any fixed discriminant D also naturally break up into orbits under the action of $SL_2(\mathbb{Z})$. We say a quadratic form $ax^2 + bxy + cy^2$ is *primitive* if a, b, c are relatively prime. Then $SL_2(\mathbb{Z})$ evidently preserves primitivity, so that the primitive forms of a given discriminant also break up into $SL_2(\mathbb{Z})$-orbits. Gauss's remarkable discovery regarding these primitive $SL_2(\mathbb{Z})$-orbits was the following:

Theorem 1.1 (Gauss). *Let $D \equiv 0$ or 1 modulo 4. Then the set of $SL_2(\mathbb{Z})$-orbits of primitive binary quadratic forms having discriminant D naturally possesses the structure of a finite abelian group.*

What is particularly remarkable about this theorem is that Gauss proved this result before the notion of group formally existed! Theorem 1.1 is quite a deep fact, and has a number of beautiful interpretations. Classically, the theorem generalizes the identity of Brahmagupta [12]:

$$(x_1^2 + Dy_1^2)(x_2^2 + Dy_2^2) = x_3^2 + Dy_3^2,$$

where $x_3 = x_1x_2 + Dy_1y_2$ and $y_3 = x_1y_2 - y_1x_2$. Gauss's theorem describes all identities of the form

$$(a_1x_1^2 + b_1x_1y_1 + c_1y_1^2)(a_2x_2^2 + b_2x_2y_2 + c_2y_2^2) = (a_3x_3^2 + b_3x_3y_3 + c_3y_3^2) \quad (1)$$

where x_3 and y_3 are bilinear functions of (x_1, y_1) and (x_2, y_2) with integer coefficients. Because of the bilinearity condition on (x_3, y_3), the existence of an identity of the type (1) depends only on the $SL_2(\mathbb{Z})$-equivalence classes of the three forms. Remarkably, the ensemble of all such identities turns the set of $SL_2(\mathbb{Z})$-equivalence classes of primitive quadratic forms of discriminant D into a group for any eligible value of D. This is precisely the group described by Gauss in Theorem 1.1, showing in particular that the theorem is compatible with the multiplicative structure of the values taken by the forms.

In modern language, the group described in Theorem 1.1 is simply the narrow class group of the unique quadratic ring $S(D)$ of discriminant D (see Section 2.1). This connection with ideal class groups was in fact one of the original motivations for Dedekind to introduce "ideal numbers", or what are now called ideals. Thus Theorem 1.1 really lies at the foundations of modern algebraic number theory. Moreover, Gauss composition still remains one of the best methods for understanding narrow class groups of quadratic fields, and it is certainly still the best way of computing with them.

Of course, Gauss's composition law is related in this way only to field extensions of $\mathbb{Q}$ of degree two, and it would be desirable to have similar ways to understand cubic, quartic, and higher degree fields. The question thus arises: do there exist analogous

composition laws on other spaces of forms, which could be used to shed light on the structure of higher degree fields?

2. The parametrization of algebraic structures

2.1. Gauss composition and rings of rank 2. An alternate way of viewing Gauss composition is as a parametrization result. To describe this, we need some simple definitions. First, define a *ring of rank n* to be any commutative ring with identity whose underlying additive group is isomorphic to $\mathbb{Z}^n$. For example, an order in a number field of degree n is a ring of rank n. Rings of rank 2, 3, 4, 5, and 6 are called *quadratic, cubic, quartic, quintic*, and *sextic* rings respectively. In general, a ring $\mathcal{R}$ of rank n is said to be an *order* in a $\mathbb{Q}$-algebra K if $\mathcal{R} \otimes \mathbb{Q} = K$.

Given a ring $\mathcal{R}$ of rank n, there are two simple functions $\mathcal{R} \to \mathbb{Z}$ called the *trace* and the *norm*, denoted by Tr and N respectively. Given $\alpha \in \mathcal{R}$, we define $\mathrm{Tr}(\alpha)$ (resp. $\mathrm{N}(\alpha)$) as the trace (resp. determinant) of the linear map $\mathcal{R} \xrightarrow{\times \alpha} \mathcal{R}$ given by multiplication by α. The function $x, y \mapsto \mathrm{Tr}(xy)$ defines an inner product on $\mathcal{R}$. If $\langle \alpha_0, \ldots, \alpha_{n-1} \rangle$ is a $\mathbb{Z}$-basis of $\mathcal{R}$, then the *discriminant* $\mathrm{Disc}(\mathcal{R})$ is defined to be the determinant $\mathrm{Det}(\mathrm{Tr}(\alpha_i \alpha_j)_{0 \leq i, j \leq n-1})$. In basis-free terms, the discriminant of $\mathcal{R}$ is the co-volume of the lattice $\mathcal{R}$ with respect to this inner product, and forms the most important invariant of a ring of rank n. It turns out that the discriminant is always an integer congruent to 0 or 1 (mod 4).

It is easy to describe what all quadratic rings are in terms of the discriminant. Namely, for every integer D congruent to 0 or 1 modulo 4, there is a unique quadratic ring $S(D)$ having discriminant D (up to isomorphism), given by

$$S(D) = \begin{cases} \mathbb{Z}[x]/(x^2) & \text{if } D = 0, \\ \mathbb{Z} \cdot (1, 1) + \sqrt{D} \cdot (\mathbb{Z} \oplus \mathbb{Z}) & \text{if } D \geq 1 \text{ is a square,} \\ \mathbb{Z}[(D + \sqrt{D})/2] & \text{otherwise.} \end{cases} \qquad (2)$$

Therefore, if we denote by $\mathbb{D}$ the set of elements of $\mathbb{Z}$ that are congruent to 0 or 1 (mod 4), we may say that isomorphism classes of quadratic rings are parametrized by $\mathbb{D}$. The case $D = 0$ is called the *degenerate* case.

Gauss composition concerns the parametrization of narrow (or oriented) ideal classes in oriented quadratic rings. An *oriented* quadratic ring is a quadratic ring in which one of the two choices for a square root $\sqrt{D}$ of D has been distinguished, where D denotes the discriminant of the ring.[2] An *oriented* ideal of a nondegenerate quadratic ring S is a pair (I, ε), where I is any ideal of S having rank 2 over $\mathbb{Z}$ and $\varepsilon = \pm 1$ gives the *orientation* of I. Multiplication of oriented ideals is defined

---

[2]The advantage of this point of view is that any two oriented quadratic rings of the same discriminant are then canonically isomorphic; to construct this isomorphism, one simply sends the distinguished $\sqrt{D}$ in one ring to that in the other. Note that a choice of $\sqrt{D}$ amounts to a choice of generator of $\wedge^2 S$, namely $1 \wedge \left(\frac{D+\sqrt{D}}{2} \right)$ — hence the name *oriented* quadratic ring.

componentwise. Similarly, for an element $\kappa \in K = S \otimes \mathbb{Q}$, the product $\kappa \cdot (I, \varepsilon)$ is defined to be the ideal $(\kappa I, \text{sgn}(N(\kappa))\varepsilon)$. Two oriented ideals (I_1, ε_1) and (I_2, ε_2) are said to be in the same *class* if $(I_1, \varepsilon_1) = \kappa \cdot (I_2, \varepsilon_2)$ for some invertible $\kappa \in K$. In practice, we will denote an oriented ideal (I, ε) simply by I, with the orientation $\varepsilon = \varepsilon(I)$ on I being understood.

In this language, Gauss composition states:

Theorem 2.1. *There is a canonical bijection between the set of* $\text{SL}_2(\mathbb{Z})$*-equivalence classes of nondegenerate binary quadratic forms and the set of isomorphism classes of pairs* (S, I)*, where S is a nondegenerate oriented quadratic ring and I is an oriented ideal class of S.*

The map from oriented ideal classes to binary quadratic forms is easily described. Given an oriented ideal $I \subset S$, let $\langle \alpha_1, \alpha_2 \rangle$ be a correctly oriented basis of I, i.e., a basis such that the determinant of the change-of-basis matrix from $\langle 1, \sqrt{D} \rangle$ to $\langle \alpha_1, \alpha_2 \rangle$ has the same sign as $\varepsilon(I)$;[3] this determinant is called the *norm* of I and is denoted $N(I)$. To the oriented ideal I, one then associates the binary quadratic form

$$Q(x, y) = \frac{N(\alpha_1 x + \alpha_2 y)}{N(I)}. \tag{3}$$

One readily verifies that $Q(x, y)$ is an integral binary quadratic form and that it is well-defined up to the action of $\text{SL}_2(\mathbb{Z})$. What is remarkable about Theorem 2.1 is not just that every oriented ideal class of a quadratic ring yields an integral binary quadratic form, but that every integral binary quadratic form arises in this way! Another remarkable aspect of the correspondence (3) of Theorem 2.1 is that it is *discriminant-preserving*: under the bijection, the discriminant of a binary quadratic form is equal to the discriminant of the corresponding quadratic ring. That is, oriented ideal classes in $S(D)$ correspond to $\text{SL}_2(\mathbb{Z})$-equivalence classes of binary quadratic forms having discriminant D.

An oriented ideal I of the oriented quadratic ring $S(D)$ is said to be *invertible* if there exists a (fractional) oriented ideal I' such that the product II' is $(S(D), +1)$. It is known that the set of invertible oriented ideals modulo multiplication by scalars forms a finite abelian group $\text{Cl}^+(S(D))$, called the *oriented* (or *narrow*) *class group*.[4] One checks that invertible oriented ideals correspond to primitive forms via (3). Gauss's group structure on classes of primitive forms of discriminant D arises from the fact that the invertible oriented ideal classes of a quadratic ring $S(D)$ form a group under multiplication.

In the statement of the theorem, we have used the word "nondegenerate" to mean "nonzero discriminant". Theorem 2.1 could also be extended to zero discriminant, although this would require a rather more involved notion of "oriented ideal class", so in what follows we always restrict ourselves to the nondegenerate case.

---

[3]Evidently, for any basis $\langle \alpha_1, \alpha_2 \rangle$ of I, either $\langle \alpha_1, \alpha_2 \rangle$ or $\langle \alpha_2, \alpha_1 \rangle$ will be correctly oriented.

[4]The usual *class group* is a quotient of the oriented class group, and may be obtained by "forgetting" all orientations.

2.2. Parametrization and rings of rank n. In terms of Theorem 2.1, it becomes easier to see what we might mean by "generalizations" of Gauss composition. Namely, we seek an algebraic group G and a representation V, defined over $\mathbb{Z}$, such that the set $G(\mathbb{Z}) \setminus V(\mathbb{Z})$ of integral orbits are in canonical bijection with interesting algebraic objects – such as rings of rank n, modules over these rings, and maps among them. In Gauss's case, the group G is SL_2 and V is the space of binary quadratic forms, and we have seen that the integral orbits parametrize oriented ideal classes (or oriented rank 1 modules) in quadratic rings. In general, we have the following question:

Question 2.2. For what pairs (G, V) does $G(\mathbb{Z}) \setminus V(\mathbb{Z})$ parametrize rings, modules, maps, etc.?

If other such pairs (G, V) do in fact exist, where do we go about looking for them? One thing to notice about the action of $\mathrm{GL}_2(\mathbb{C})$ on the vector space of binary quadratic forms over $\mathbb{C}$ is that there is essentially one (Zariski open) orbit – i.e., any binary quadratic form of nonzero discriminant can be taken to any other such form via an element of $\mathrm{GL}_2(\mathbb{C})$. It is also possible to see this from the point of view of Gauss composition: by "base change" the proof of Theorem 2.1 shows that nondegenerate orbits over $\mathbb{C}$ must be in one-to-one correspondence with quadratic rings over $\mathbb{C}$ – which must take the form $\mathbb{C} \oplus \mathbb{C}$ – and ideal classes over such rings – which also must take the form $\mathbb{C} \oplus \mathbb{C}$ (up to isomorphism). So over $\mathbb{C}$, there is essentially just one object of the form (S, I), namely $S = I = \mathbb{C} \oplus \mathbb{C}$.

By the same argument, if we are to get a parametrization result of a simple form like Gauss composition, where objects being parametrized are rings of rank n, ideal classes, etc. (so that there is only one such nondegenerate object over $\mathbb{C}$), then the pair (G, V) must also have the property that there is just one open orbit over $\mathbb{C}$. Such representations having just one open orbit over $\mathbb{C}$ have come up for numerous authors in various contexts, and they are known as "prehomogeneous vector spaces".

Definition 2.3. A *prehomogeneous vector space* is a pair (G, V) where G is an algebraic group and V is a rational vector space representation of G such that the action of $G(\mathbb{C})$ on $V(\mathbb{C})$ has just one Zariski open orbit.

In a monumental work, Sato and Kimura [33] gave a classification of all "reduced, irreducible" prehomogeneous vector spaces. Namely, they showed that there are essentially 36 of them! A few of these 36 are in fact infinite families. In another beautiful work, Wright and Yukie [41] studied these spaces over fields and found that the K-orbits for a field K frequently correspond to field extensions of K; for example, the nondegenerate $\mathbb{Q}$-orbits on the space of binary quadratic forms are naturally in bijection with quadratic extensions of $\mathbb{Q}$. So that gives us some hope, and obtaining the answer to Question 2.2 thus translates into the following goal:

Goal 2.4. Understand $G(\mathbb{Z}) \setminus V(\mathbb{Z})$ for prehomogeneous vector spaces (G, V).

Of course, some of these spaces are quite large – thirty or more dimensions – so to just go in and analyze the integer orbits is somewhat daunting. Even Gauss's space,

which is only three-dimensional, is (as we have seen!) far from trivial. Gauss's own treatment of Gauss composition took numerous pages to describe.

   To make further progress, we wish to have a different – and perhaps also simpler – perspective on Gauss composition that might lend itself more naturally to generalization to other spaces. Following [4], we give such a perspective in terms of $2 \times 2 \times 2$ cubes of integers. As we will see, the space of $2 \times 2 \times 2$ cubes not only gives an elementary description of Gauss composition, but also leads to composition laws and analogues of Theorem 2.1 for numerous other prehomogeneous vector spaces.

3. The story of the cube

Suppose we put integers on the corners of a cube:

$$
\begin{array}{c}
\text{(4)}
\end{array}
$$

Notice that any such cube A of integers may be sliced into two 2×2 matrices, and in essentially three different ways, corresponding to three different planes of symmetry of a cube. More precisely, the integer cube A given by (4) can be sliced into the following pairs of 2×2 matrices:

$$M_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad N_1 = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

$$M_2 = \begin{bmatrix} a & c \\ e & g \end{bmatrix}, \quad N_2 = \begin{bmatrix} b & d \\ f & h \end{bmatrix} \qquad (5)$$

$$M_3 = \begin{bmatrix} a & e \\ b & f \end{bmatrix}, \quad N_3 = \begin{bmatrix} c & g \\ d & h \end{bmatrix}.$$

   Now for any such slicing of the cube A into a pair (M_i, N_i) of 2×2 matrices as in (5), we may construct a binary quadratic form $Q_i(x, y)$ as follows:

$$Q_i(x, y) = -\mathrm{Det}(M_i x + N_i y). \qquad (6)$$

Thus any cube A of integers gives rise to three integral binary quadratic forms. A simple computation or elementary argument shows that the discriminants of the three quadratic forms Q_1, Q_2, and Q_3 are the same! And the punchline is:

Theorem 3.1. *If a cube A gives rise to three primitive binary quadratic forms Q_1, Q_2, Q_3 via (4)–(6), then Q_1, Q_2, Q_3 have the same discriminant, and the product of these three forms is the identity in the group defined by Gauss composition.*

Conversely, if Q_1, Q_2, Q_3 are any three primitive binary quadratic forms of the same discriminant whose product is the identity under Gauss composition, then there exists a cube A yielding Q_1, Q_2, Q_3 via (4)–(6).

Thus the cube story gives a very simple and complete description of Gauss composition of binary quadratic forms. In fact, Theorem 3.1 can be used to *define* Gauss composition. The situation is reminiscent of the group law on a plane elliptic curve, where the most elementary way to define the group law is to declare that three points sum to zero if and only if they lie on a common line. In the same way, we may define Gauss composition by declaring that three primitive quadratic forms multiply to the identity if and only if they arise from a common cube. A proof of Theorem 3.1 may be found in [4, Appendix].

Theorem 3.1 is useful not only because it leads to Gauss composition, but also because it leads to various additional laws of composition. First and foremost, it leads to a law of composition on the cubes themselves!

3.1. Composition of cubes. Let us begin by rephrasing Theorem 3.1 as an orbit problem. First, we note that the space of cubes may be identified with the representation $\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$ of the group $G = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$; this representation is a prehomogeneous vector space. The identification is made as follows: if we use $\langle v_1, v_2 \rangle$ to denote the standard basis of $\mathbb{Z}^2$, then the cube described by (4) is simply

$$a\, v_1 \otimes v_1 \otimes v_1 + b\, v_1 \otimes v_2 \otimes v_1 + c\, v_2 \otimes v_1 \otimes v_1 + d\, v_2 \otimes v_2 \otimes v_1$$
$$+\, e\, v_1 \otimes v_1 \otimes v_2 + f\, v_1 \otimes v_2 \otimes v_2 + g\, v_2 \otimes v_1 \otimes v_2 + h\, v_2 \otimes v_2 \otimes v_2$$

as an element of $\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$. In terms of the cubical representation (4), the three factors of $\mathrm{SL}_2(\mathbb{Z})$ in G act by row operations, column operations, and the "other direction" operations respectively.

Theorem 3.1 may be viewed as describing the nondegenerate orbits of $\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$ under the action of G in terms of triples of oriented ideal classes whose "product" is the identity class. To state this description more precisely, we need just two simple definitions. First, we call a triple (I_1, I_2, I_3) of oriented fractional ideals in $S \otimes \mathbb{Q}$ *balanced* if $I_1 I_2 I_3 \subseteq S$ and $\mathrm{N}(I_1)\mathrm{N}(I_2)\mathrm{N}(I_3) = 1$. Also, we define two balanced triples (I_1, I_2, I_3) and (I_1', I_2', I_3') of oriented ideals of S to be *equivalent* if $I_1 = \kappa_1 I_1'$, $I_2 = \kappa_2 I_2'$, and $I_3 = \kappa_3 I_3'$ for some invertible elements $\kappa_1, \kappa_2, \kappa_3 \in S \otimes \mathbb{Q}$. For example, if S is the full ring of integers in a quadratic field, then an equivalence class of balanced triples means simply a triple of oriented ideal classes whose product is the principal class.

Our Theorem 3.1 on cubes may then be stated as the solution to an orbit problem as follows:

Theorem 3.2. *There is a canonical bijection between the set of nondegenerate G-orbits on the space $\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$ of $2 \times 2 \times 2$ integer cubes and the set of isomorphism classes of pairs $(S, (I_1, I_2, I_3))$, where S is a nondegenerate oriented quadratic ring and (I_1, I_2, I_3) is an equivalence class of balanced triples of oriented ideals of S.*

As with Theorem 3.2, we may consider those orbits that correspond solely to *invertible* oriented ideal classes. Let us say a cube A is *projective* if the three oriented ideal classes associated to A in Theorem 3.2 are invertible (i.e., if they are projective as modules). Equivalently, A is projective if the associated three binary quadratic forms Q_i are each primitive.

Let us define the *discriminant* $\mathrm{Disc}(A)$ of a cube A to be the discriminant of any one of the three binary quadratic forms Q_i arising from it. Then Theorem 3.2 is discriminant-preserving: under the bijection, the discriminant of a cube is equal to the discriminant of the corresponding quadratic ring.

We can now describe composition of cubes. It is most easily stated in terms of ideal classes. Recall that Gauss composition can be viewed as multiplication of oriented ideal classes in a fixed quadratic ring S:

$$(S, I) \circ (S, I') = (S, II').$$

When restricted to invertible ideal classes of a fixed quadratic ring $S = S(D)$ (i.e., primitive binary quadratic forms having a fixed discriminant D), this yields the oriented class group $\mathrm{Cl}^+(S(D))$.

Analogously, composition of cubes can be viewed as multiplication of equivalence classes of balanced triples of oriented ideals:

$$(S, (I_1, I_2, I_3)) \circ (S, (I_1', I_2', I_3')) = (S, (I_1 I_1', I_2 I_2', I_3 I_3')).$$

When restricted to invertible ideal classes of a fixed quadratic ring (i.e., projective cubes having a fixed discriminant), this yields the group $\mathrm{Cl}^+(S) \times \mathrm{Cl}^+(S)$, since the last ideal class is determined by the first two. Thus Gauss composition yields $\mathrm{Cl}^+(S)$, while composition of cubes gives $\mathrm{Cl}^+(S) \times \mathrm{Cl}^+(S)$. A surprising consequence of this result is that the number of orbits of projective cubes having a given discriminant D is always a square number.

3.2. Composition of binary cubic forms.

The law of composition of cubes now also leads to a number of further composition laws on various other spaces. First, let us consider the space of triply-symmetric cubes, which is equivalent to the space of binary cubic forms $px^3 + 3qx^2y + 3rxy^2 + sy^3$: indeed, just as one often expresses a binary quadratic form $px^2 + 2qxy + ry^2$ as the symmetric 2×2 matrix

$$\begin{bmatrix} p & q \\ q & r \end{bmatrix},$$

one may naturally express a binary cubic form $px^3 + 3qx^2y + 3rxy^2 + sy^3$ via the triply-symmetric $2 \times 2 \times 2$ matrix

$$
\begin{array}{c}
\text{(cube diagram)}
\end{array}
\tag{7}
$$

If we use $\text{Sym}^3\mathbb{Z}^2$ to denote the space of binary cubic forms with triplicate central coefficients, then the above association of $px^3 + 3qx^2y + 3rxy^2 + sy^3$ with the cube (7) corresponds to the natural inclusion

$$
\iota \colon \text{Sym}^3\mathbb{Z}^2 \to \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2
$$

of the space of triply-symmetric cubes into the space of cubes. The space of binary cubic forms under the action of $\text{SL}_2(\mathbb{Z})$ also yields a prehomogeneous vector space.

We call a binary cubic form $C(x, y) = px^3 + 3qx^2y + 3rxy^2 + sy^3$ *projective* if the corresponding triply-symmetric cube $\iota(C)$ given by (7) is projective. It turns out that the $\text{SL}_2(\mathbb{Z})$-orbits on such binary cubic forms having a fixed discriminant D also then inherit a law of composition from the space of cubes, leading to a group structure when restricted to projective forms. It is not hard to guess what this group should be related to. Namely, projective triply-symmetric cubes correspond to a balanced triple of ideals (I, I, I) in $S(D)$, where the three ideals are in fact the same. Thus $I \cdot I \cdot I$ is the identity ideal class, implying that orbits of binary cubic forms essentially correspond to 3-torsion elements in the oriented class group $\text{Cl}^+(S)$. (The precise 3-torsion group one obtains is discussed in [4].) Thus the symmetrization procedure allows us to isolate a certain arithmetic part of the class group.

An interesting consequence of this result is that the number of orbits of projective binary cubic forms having a given discriminant D is always a power of three!

3.3. Composition of pairs of binary quadratic forms. The group law on binary cubic forms of discriminant D was obtained by imposing a triple-symmetry condition on the group of $2 \times 2 \times 2$ cubes of discriminant D. Rather than imposing a threefold symmetry, one may instead impose only a twofold symmetry. This leads to cubes taking the form

$$
\begin{array}{c}
\text{(cube diagram)}
\end{array}
\tag{8}
$$

That is, these cubes can be sliced (along a certain fixed plane) into two 2×2 symmetric matrices and therefore can naturally be viewed as a pair of binary quadratic forms $(ax^2 + 2bxy + cy^2, dx^2 + 2exy + fy^2)$.

If we use $\mathbb{Z}^2 \otimes \text{Sym}^2\mathbb{Z}^2$ to denote the space of pairs of integer-matrix binary quadratic forms, then the above association of $(ax^2 + 2bxy + cy^2, dx^2 + 2exy + fy^2)$ with the cube (8) corresponds to the natural inclusion map

$$\jmath : \mathbb{Z}^2 \otimes \text{Sym}^2\mathbb{Z}^2 \to \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2.$$

The lattice $\mathbb{Z}^2 \otimes \text{Sym}^2\mathbb{Z}^2$ under the action of $\text{SL}_2(\mathbb{Z}) \times \text{SL}_2(\mathbb{Z})$ again yields a prehomogeneous vector space.

As in the case of binary cubic forms, we call a pair of binary quadratic forms *projective* if the corresponding doubly-symmetric cube $\jmath(C)$ given by (8) is projective. Again, the projective pairs of binary quadratic forms having a fixed discriminant D inherit a group structure. Since such elements correspond to balanced triples of ideals (I_1, I_3, I_3) where the last two ideals are the same, one sees that the group thus obtained is again simply the group $\text{Cl}^+(S(D))$ since I_3 determines I_1. That is, not only do binary quadratic forms of a fixed discriminant D give rise to the oriented class group of $S(D)$, but so do *pairs* of binary quadratic forms!

3.4. Further parametrization spaces for quadratic rings. The discussions above illustrate that once we have a law of composition on the space of cubes, then various other of its invariant and covariant spaces also inherit a law of composition; Gauss composition is indeed just one of these.

Symmetrization is one procedure that allows us to generate new prehomogenous vector spaces with composition; this was the subject of Sections 3.2 and 3.3. The determinant trick (6) to produce Gauss composition is another. There are several other operations too that play an important role, such as *skew-symmetrization*, *symplectization*, *hermitianization*, and *dualization*, and each procedure is found to have both invariant-theoretic and number-theoretic meaning, yielding numerous further analogues of Theorem 2.1 involving higher rank rings, higher rank modules, as well as noncommutative rings such as quaternion and octonion algebras. Further details may be found in [4] and [8].

4. Cubic analogues of Gauss composition

In the previous section, we discussed various generalizations of Gauss composition that were found to be closely related to the ideal class groups of quadratic rings. In this section, we show how similar ideas can be used to obtain genuine "cubic analogues" of Gauss composition, i.e., composition laws on appropriate spaces of forms so that the resulting groups are related to the class groups of *cubic rings*.

The fundamental object in our treatment of quadratic composition was the space of $2 \times 2 \times 2$ cubes of integers. It turns out that the fundamental object for cubic

composition is the space of $2 \times 3 \times 3$ boxes of integers, and yields exactly what is needed for a cubic analogue of Gauss's theory. The action of $GL_2(\mathbb{Z}) \times SL_3(\mathbb{Z}) \times SL_3(\mathbb{Z})$ on $2 \times 3 \times 3$ integer boxes is again a prehomogeneous vector space, and the orbits correspond in a natural way to cubic rings and ideal classes in those rings. Before the resulting cubic analogues of Gauss composition can be described, it is necessary first to understand how cubic rings are parametrized.

4.1. The parametrization of cubic rings. In Section 2.1, we saw that quadratic rings are parametrized up to isomorphism by their discriminants. This is not so for cubic rings; indeed, there may sometimes be several nonisomorphic cubic rings having the same discriminant. The correct object parametrizing cubic rings – i.e., rings free of rank 3 as $\mathbb{Z}$-modules – was first determined by Delone–Faddeev in their beautiful 1964 treatise on cubic irrationalites [21]. They showed that cubic rings are in bijective correspondence with $GL_2(\mathbb{Z})$-equivalence classes of integral binary cubic forms $ax^3 + bx^2y + cxy^2 + dy^3$, as follows.

Given a binary cubic form $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ with $a, b, c, d \in \mathbb{Z}$, we associate to f the ring $R(f)$ having $\mathbb{Z}$-basis $\langle 1, \omega, \theta \rangle$ and multiplication table

$$
\begin{aligned}
\omega\theta &= -ad \\
\omega^2 &= -ac + b\omega - a\theta \\
\theta^2 &= -bd + d\omega - c\theta.
\end{aligned}
\tag{9}
$$

One easily verifies that $GL_2(\mathbb{Z})$-equivalent binary cubic forms then yield isomorphic rings, and conversely, that every isomorphism class of ring R can be represented in the form $R(f)$ for a unique binary cubic form f, up to such equivalence. Thus we may say that isomorphism classes of cubic rings are parametrized by $GL_2(\mathbb{Z})$-equivalence classes of integral binary cubic forms. An easy calculation using (9) shows that the discriminant $Disc(R(f))$ is equal to the discriminant $Disc(f)$ of the binary cubic form f, where $Disc(f) = b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2$ is the unique polynomial invariant for the action of $GL_2(\mathbb{Z})$ on binary cubic forms. We thus obtain:

Theorem 4.1 ([21]). *There is a canonical bijection between the set of $GL_2(\mathbb{Z})$-equivalence classes of integral binary cubic forms and the set of isomorphism classes of cubic rings, by the association*

$$
f \leftrightarrow R(f).
$$

Moreover, $Disc(f) = Disc(R(f))$.

We say a cubic ring is *nondegenerate* if it has nonzero discriminant (equivalently, if it is an order in an étale cubic algebra over $\mathbb{Q}$). Similarly, a binary cubic form is *nondegenerate* if it has nonzero discriminant (equivalently, if it has distinct roots in $\mathbb{P}^1(\mathbb{Q})$). The discriminant equality in Theorem 4.1 implies, in particular, that isomorphism classes of nondegenerate cubic rings correspond bijectively with equivalence classes of nondegenerate integral binary cubic forms.

4.2. Cubic composition and 2 × 3 × 3 boxes. Imitating Section 3.1, for a cubic ring R let us say a pair (I, I') of fractional R-ideals in $K = R \otimes \mathbb{Q}$ is *balanced* if $II' \subseteq R$ and $N(I)N(I') = 1$. Furthermore, we say two such balanced pairs (I_1, I_1') and (I_2, I_2') are *equivalent* if there exist invertible elements $\kappa, \kappa' \in K$ such that $I_1 = \kappa I_2$ and $I_1' = \kappa' I_2'$. For example, if R is the full ring of integers in a cubic field then an equivalence class of balanced pairs of ideals is simply a pair of ideal classes that are inverse to each other in the ideal class group.

The analogue of Theorem 3.2 in the theory of cubic composition states that $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$-classes of $2 \times 3 \times 3$ integer boxes correspond to equivalence classes of balanced pairs of ideals in cubic rings.

Theorem 4.2. *There is a canonical bijection between the set of nondegenerate* $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$-*orbits on the space* $\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$ *and the set of isomorphism classes of pairs* $(R, (I, I'))$, *where R is a nondegenerate cubic ring and (I, I') is an equivalence class of balanced pairs of ideals of R.*

How does one recover the cubic ring R from a $2 \times 3 \times 3$ box of integers? A $2 \times 3 \times 3$ box may be viewed (by an appropriate slicing) as a pair (A, B) of 3×3 matrices. Then $f(x, y) = \mathrm{Det}(Ax - By)$ is a binary cubic form. The ring R is simply the cubic ring $R(f)$ associated to f via Theorem 4.1.

If we define the *discriminant* $\mathrm{Disc}(A, B)$ of (A, B) as $\mathrm{Disc}(\mathrm{Det}(Ax - By))$, then one shows again that this discriminant is the unique polynomial invariant for the action of $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ on $2 \times 3 \times 3$ boxes. By the method of recovering R from (A, B) above, we see again that the bijection of Theorem 4.2 preserves discriminants.

We may now describe composition of $2 \times 3 \times 3$ boxes. Given a binary cubic form f, let $(\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3)(f)$ denote the set of all elements $(A, B) \in \mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$ such that $\mathrm{Det}(Ax - By) = f(x, y)$; all such elements correspond to the same cubic ring in Theorem 4.2. The group $\mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ is seen to act naturally on the set $(\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3)(f)$ via simultaneous row and column operations on A and B; this action evidently does not change the value of $\mathrm{Det}(Ax - By)$. Moreover, one finds that two elements of $(\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3)(f)$ yield equivalent balanced pairs of ideals in $R(f)$ if and only if they are equivalent under $\mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$.

As with the quadratic cases of composition in Section 3, composition of $2 \times 3 \times 3$ boxes having a fixed f can now be viewed as multiplication of equivalence classes of balanced pairs of ideals in the corresponding cubic ring $R = R(f)$:

$$(R, (I, I')) \circ (R, (J, J')) = (R, (IJ, I'J')).$$

When restricted to invertible ideal classes (i.e., *projective* $2 \times 3 \times 3$ boxes), this yields the ideal class group $\mathrm{Cl}(R)$ of R, since the second ideal class is determined by the first (as they are inverses to each other). Thus composition of $\mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$-equivalence classes of projective $2 \times 3 \times 3$ boxes yields the class groups of cubic rings, in complete analogy with Gauss composition in the quadratic case.

To summarize:

- In the case of binary quadratic forms, the unique SL_2-invariant is the discriminant D, which classifies orders in quadratic fields. The primitive classes having a fixed value of D form a group under a certain natural composition law. This group is naturally isomorphic to the narrow class group of the corresponding quadratic order.

- In the case of $2 \times 3 \times 3$ integer boxes, the unique $SL_3 \times SL_3$-invariant is the binary cubic form f, which classifies orders in cubic fields. The projective classes having a fixed value of f form a group under a certain natural composition law. This group is naturally isomorphic to the ideal class group of the corresponding cubic order.

Thus the composition law on the space of $2 \times 3 \times 3$ integer cubes is really the cubic analogue of Gauss composition.

4.3. Cubic composition and pairs of ternary quadratic forms.

Just as we were able to impose a symmetry condition on $2 \times 2 \times 2$ matrices to obtain information on the exponent 3-parts of class groups of quadratic rings, we can impose a symmetry condition on $2 \times 3 \times 3$ matrices to obtain information on the exponent 2-parts of class groups of cubic rings. The "symmetric" elements in $\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$ are the elements of $\mathbb{Z}^2 \otimes \mathrm{Sym}^2\mathbb{Z}^3$, i.e., pairs (A, B) of symmetric 3×3 integer matrices, which may be viewed as pairs (A, B) of integral ternary quadratic forms. The action of $GL_2(\mathbb{Z}) \times SL_3(\mathbb{Z})$ on pairs of ternary quadratic forms is again a prehomogeneous vector space.

The cubic form invariant f and the *discriminant* $\mathrm{Disc}(A, B)$ of (A, B) may be defined in the identical manner; we have $f(x, y) = \mathrm{Det}(Ax - By)$ and $\mathrm{Disc}(A, B) = \mathrm{Disc}(\mathrm{Det}(Ax - By))$. We say an element $(A, B) \in \mathbb{Z}^2 \otimes \mathrm{Sym}^2\mathbb{Z}^3$ is *projective* if it is projective as a $2 \times 3 \times 3$ box.

As in the case of binary cubic forms and symmetric cubes (see Section 3.2), the space of pairs of ternary quadratic forms also inherits a law of composition from the space of $2 \times 3 \times 3$ boxes. Again, the restriction to symmetric classes isolates a certain arithmetic part of the class group. Namely, symmetric projective $2 \times 3 \times 3$ boxes yield pairs of the form $(R, (I, I))$ where the two ideals are in fact the same. Thus $I \cdot I$ is the identity ideal class of R, so we see that $GL_2(\mathbb{Z}) \times SL_3(\mathbb{Z})$-orbits of pairs of integer-matrix ternary quadratic forms essentially parametrize 2-torsion elements in the class groups of cubic rings (see [5] for further details).

This parametrization has several interesting consequences. For example, it implies that the number of equivalence classes of projective pairs (A, B) of ternary quadratic forms having a given binary cubic $\mathrm{Det}(Ax - By)$ is always a power of 2!

The parametrization also enables one to prove the first known case of the Cohen–Martinet heuristics for class groups, namely for the average size of the 2-torsion subgroup in the class groups of cubic fields. This average number of 2-torsion elements turns out to be 5/4 for real cubic fields and 3/2 for complex cubic fields. In particular,

this implies that at least 75% of totally real cubic fields, and at least 50% of complex cubic fields, have odd class number. Further details may be found in [9]. The case of narrow class groups can also be handled by generalizations of these arguments (to appear in future work).

5. The parametrization of quartic and quintic rings

The composition laws and results of the previous two sections depended heavily on the simple but beautiful parametrizations of quadratic and cubic rings given by (2) and (9) respectively. Namely, we saw that quadratic rings are parametrized by integers congruent to 0 or 1 (mod 4), while cubic rings are parametrized by $GL_2(\mathbb{Z})$-equivalence classes of binary cubic forms.

It has been a long-time open question to determine whether analogous parametrizations exist for rings of rank 4. The ideas of the previous sections, together with a theory of *resolvent rings*, lead to a parametrization of quartic rings that is just as complete as in the quadratic and cubic cases. These "resolvent rings" are so named because they form natural integral models of the resolvent fields occurring in the classical literature; see [6] for further details.

This perspective leads one to show that the analogous objects parametrizing quartic rings are essentially *pairs of integer-valued ternary quadratic forms*, up to integer equivalence. To make a precise statement, let $(\mathbb{Z}^2 \otimes \mathrm{Sym}^2\mathbb{Z}^3)^*$ denote the space of pairs of ternary quadratic forms having integer coefficients. Then we have:

Theorem 5.1. *There is a canonical bijection between the set of $GL_2(\mathbb{Z}) \times SL_3(\mathbb{Z})$-orbits on the space $(\mathbb{Z}^2 \otimes \mathrm{Sym}^2\mathbb{Z}^3)^*$ of pairs of integer-valued ternary quadratic forms and the set of isomorphism classes of pairs (Q, R), where Q is a quartic ring and R is a cubic resolvent ring of Q.*

A cubic resolvent ring of a quartic ring Q is a cubic ring R equipped with a certain natural quadratic mapping $Q \to R$. It turns out that all quartic rings have at least one cubic resolvent ring; moreover, for "most" quartic rings (e.g., for maximal quartic rings) this cubic resolvent ring is in fact unique. Thus every quartic ring arises in Theorem 5.1, and the theorem yields a bijection on the quartic rings of primary interest to algebraic number theorists, namely the maximal orders in quartic fields.

The theory of resolvent rings used in [6] to prove Theorem 5.1 makes heavy use of many of the formulae arising in the solution to the quartic equation. The same ideas also yield a purely ring-theoretic interpretation of the Delone–Faddeev parametrization of cubic rings, using formulae arising in the classical solution to the cubic equation.

Since the quintic equation is known to be "unsolvable", it may then seem that such parametrization methods could not extend beyond the quartic. However, there has been a lot of literature on the quintic equation, and some of the formulae that arise

in these works – although they fail to "solve" the quintic equation – can nevertheless be adapted to develop a completely analogous theory for parametrizing quintic rings! It turns out that quintic rings are essentially parametrized by *quadruples of quinary alternating bilinear forms*, i.e., quadruples of 5×5 skew-symmetric integer matrices.

Let $\mathbb{Z}^4 \otimes \wedge^2 \mathbb{Z}^5$ denote the space of quadruples of 5×5 skew-symmetric integer matrices. Then the parametrization result for quintic rings is as follows.

Theorem 5.2. *There is a canonical bijection between the set of* $\mathrm{GL}_4(\mathbb{Z}) \times \mathrm{SL}_5(\mathbb{Z})$-*orbits on the space* $\mathbb{Z}^4 \otimes \wedge^2 \mathbb{Z}^5$ *of quadruples of* 5×5 *skew-symmetric integer matrices and the set of isomorphism classes of pairs* (R, S), *where R is a quintic ring and S is a sextic resolvent ring of R.*

A *sextic resolvent ring* of a quintic ring R is a sextic ring S equipped with a certain natural mapping $R \to \wedge^2 S$ which seems to have been missed in the classical literature on the quintic equation. The notion of sextic resolvent ring yields a natural integral model for the sextic resolvent fields studied by Cayley and Klein. As in the quartic case, one finds that all quintic rings have a sextic resolvent, and maximal quintic rings have exactly one sextic resolvent ring. Thus Theorem 5.2 yields a bijection on maximal orders in quintic fields!

These parametrization results have an important application to determining the density of discriminants of number fields of degree less than or equal to five, which we discuss in the next section.

Because of the classification of prehomogeneous vector spaces, one can show that parametrizations of the *same type* cannot exist for rings of rank $n > 5$. This is in agreement with the classification of group stabilizers by Sato–Kimura [33], and with the classification of orbits over fields by Wright–Yukie [41]. Thus parametrizations of this type end with the quintic.

6. Counting number fields of low degree

Number fields – i.e., field extensions of the rational numbers of finite degree – are perhaps the most fundamental objects in algebraic number theory, yet very little is known about their distribution with respect to basic invariants.

The most fundamental numerical invariant of a degree n number field K is its *discriminant* $\mathrm{Disc}(K)$. The quantity $\mathrm{Disc}(K)$ is defined as $\mathrm{Disc}(O_K)$, where O_K denotes the unique maximal ring of rank n contained in K (equivalently, O_K is the ring of algebraic integers in K). A fundamental theorem of Minkowski states that, up to isomorphism, there can be only finitely many number fields having any given discriminant D. The question thus arises: how many? The number of number fields of discriminant D fluctuates with D in a seemingly random manner, so that obtaining an exact answer would be rather unwieldy. Nevertheless, it is still natural to ask whether one can understand the answer on average. That is, how many number fields do we expect having discriminant D?

It is natural to refine the latter question by considering each degree and each associated Galois group separately. For the remainder of this section, we fix the degree to be n and consider the degree n number fields whose Galois closures have Galois group S_n, which is in some sense the "generic" case. We now consider successive cases of n, starting with the simplest case, namely

$n = 1$. There is only one degree 1 number field, namely the field $\mathbb{Q}$ of rational numbers, and its discriminant is 1. Thus we expect zero degree 1 number fields per discriminant as the discriminant tends to infinity. $\square$

$n = 2$. The case $n = 2$ is also not difficult to handle. Recall that, for each nonsquare discriminant D, there is a unique quadratic order having discriminant D. Maximal orders correspond to discriminants that are not square multiples of other discriminants, so that maximality essentially amounts to a squarefree condition on D.[5] It is known that the probability that a number is squarefree is $6/\pi^2$; it follows that we expect about $6/\pi^2 \approx .607\ldots$ quadratic fields per discriminant. $\square$

In the 1960s, the cases $n = 1$ and $n = 2$ apparently provided enough evidence for the following bold folk conjecture to come into existence. The origin of this conjecture seems to be unknown.

Conjecture 6.1. Let $N_n(X)$ denote the number of S_n-number fields of degree n having absolute discriminant at most X. Then

$$c_n = \lim_{X \to \infty} \frac{N_n(X)}{X}$$

exists, and is positive for $n \geq 2$.

That is, we expect about c_n S_n-number fields of degree n per discriminant, where c_n is some positive constant when $n > 1$. One question that immediately arose from the circulation of this conjecture was: what should the value of c_n be? Evidently $c_1 = 0$ and $c_2 = 6/\pi^2$, but no general formula for the value of c_n was known.

$n = 3$. Some further data as to the nature of c_n was provided in the seminal 1970 work of Davenport and Heilbronn [20], who explicitly determined the value of c_3. A key ingredient in their work was the parametrization of cubic orders by $\mathrm{GL}_2(\mathbb{Z})$-equivalence classes of binary cubic forms (see Section 4.1).

To count the number of $\mathrm{GL}_2(\mathbb{Z})$-equivalence classes of binary cubic forms having absolute discriminant less than X, Davenport constructed a fundamental domain $\mathcal{F}$ for the action of $\mathrm{GL}_2(\mathbb{Z})$ on the four-dimensional real vector space V of binary cubic

---

[5]The precise condition is that the number be squarefree and 1 (mod 4) or be four times an integer that is 2 or 3 (mod 4). So it is not quite a squarefree condition at 2. Nevertheless, the density of such numbers is still $6/\pi^2$! This is *not* a coincidence, but is part of a general phenomenon occurring in all degrees and at all primes dividing the degree, which may be explained via certain "mass formulae" arising in work of Serre. More details may be found in [11].

forms over $\mathbb{R}$. The number of cubic orders having absolute discriminant at most X is then simply the number of integer points in the region $\mathcal{F}_X$, where

$$\mathcal{F}_X = \mathcal{F} \cap \{v \in V : |\mathrm{Disc}(v)| \leq X\}.$$

This region is seen to have finite volume, namely $(\pi^2/18)X$.

Now given any region $\mathcal{R}$ in n-dimensional Euclidean space, it is very natural to approximate the number of integer lattice points in $\mathcal{R}$ by the Euclidean volume $\mathrm{Vol}(\mathcal{R})$. Such an approximation will be particularly good if the region is compact and somewhat "round-looking" in the sense that its boundaries are smooth, and there are no serious "spikes" or "tentacles" jutting out of the region.

However, if the region is noncompact or it possesses thin, long tentacles or spikes, then all bets are off. For example, one may have a region with a tentacle thinning as it runs off to infinity, which has finite volume yet contains an infinite number of lattice points. Or one could have a region with one infinitely long tentacle, which has arbitrarily large (finite or infinite) volume yet contains *no* lattice points! It is easy to draw pictures even in two-dimensional space that illustrate each of these unruly scenarios. For such "bad" regions, there may be little correlation between the volume and the number of lattice points lying within.

Let us consider again Davenport's domain $\mathcal{F}_X$. If this subset of V were compact and round, we could then conclude that the number of lattice points within is essentially $(\pi^2/18)X$. However, the region $\mathcal{F}_X$ is not compact. Although we do not attempt to draw this region here – as it is four-dimensional – it is nevertheless easy to visualize roughly what this region looks like. Namely, $\mathcal{F}_X$ is relatively round-looking, but there is a single problematic tentacle going off to infinity (arising from the fact that $\mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{SL}_2(\mathbb{R})$ is noncompact). Thus, to make any conclusions regarding the number of lattice points in $\mathcal{F}_X$, it is necessary to deal with this tentacle.

What Davenport shows is that although this tentacle (or *cusp*) contains a very large number of lattice points, nearly all of these lattice points are *reducible* cubic forms; i.e., they correspond to cubic rings sitting not in a cubic field, but in the direct sum of $\mathbb{Q}$ and a quadratic field. Only a negligible number of irreducible points are found to lie in the cusp. Meanwhile, the lattice points in the main body of the region and away from the cusps correspond almost entirely to irreducible points, i.e., orders in cubic fields; only a negligible number of points in this main body are reducible.[6]

It follows that, as $X \to \infty$, one may approximate the number of irreducible points in $\mathcal{F}_X$ by the volume of the main body of the region. As the above cusps are found to have negligible volume, we conclude that the number of irreducible points in $\mathcal{F}_X$ is $(\pi^2/18)X$, where the error is $o(X)$. We therefore obtain

Theorem 6.2. *The number of cubic orders* (*in cubic fields*) *having absolute discriminant at most X is asymptotic to $(\pi^2/18)X$ as $X \to \infty$.*

---

[6]Here, by negligible, we mean "$o(X)$".

To pass from such cubic orders to maximal cubic orders (and thus to cubic fields) requires a delicate sieve, which was carried out in the remarkable work of Davenport–Heilbronn. The result of this sieve is:

Theorem 6.3 (Davenport–Heilbronn [20]). *The number of cubic fields having absolute discriminant at most X is asymptotic to $(1/3\zeta(3))X$ as $X \to \infty$, where $\zeta(s)$ denotes the Riemann zeta function.*

Thus Davenport and Heilbronn showed, in sum, that $c_3 = \frac{1}{3\zeta(3)} \approx .277\ldots$. That is, we expect approximately .277 cubic fields per discriminant.   □

It has been a long-time open problem to extend Davenport–Heilbronn's theorem to $n = 4$, i.e., to gain an understanding of quartic number fields in the same way. Having now obtained a parametrization of quartic orders, it is natural to try and proceed in a manner analogous to Davenport–Heilbronn.

$n = 4$. As discussed in Section 5, quartic orders are parametrized by pairs of integer-coefficient ternary quadratic forms, modulo the action of the group $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$. In analogy with Davenport–Heilbronn's work, we construct a fundamental domain $\mathcal{F}$ for the action of $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ on the space V of pairs of real-coefficient ternary quadratic forms. Then $\mathcal{F}$ is a certain region in the 12-dimensional real vector space V, and quartic rings are seen to correspond to lattice points inside the fundamental domain $\mathcal{F}$.

In order to understand the number of quartic orders (and eventually quartic fields) having absolute discriminant at most X, we therefore wish to count the number of integer points inside the region $\mathcal{F}_X$, where the region $\mathcal{F}_X$ is as usual defined by $\mathcal{F} \cap \{v \in V : |\mathrm{Disc}(v)| \leq X\}$. However, the region $\mathcal{F}_X$ is not so simple; indeed, the geometry of this fundamental domain is significantly more complicated than the analogous region considered by Davenport and Heilbronn. For one thing, the dimension is now twelve instead of four! Moreover, there are now three major cusps or tentacles rather than one, and the cross sections of these cusps lie in several dimensions. If these cusps did not exist, and $\mathcal{F}_X$ were compact, it would be an easy matter to estimate the number of integer points in $\mathcal{F}_X$.

It takes quite a bit of hard work to deal with the cusps, but in the end, what happens with these cusps is quite beautiful. All three cusps contain *many* points (i.e., at least on the order of X in number). However, essentially all the lattice points in the first cusp are found to be "reducible": they correspond to quartic rings that lie in the direct sum of two quadratic fields instead of a single quartic field. The second cusp also consists almost entirely of "reducible" points! These points correspond to orders lying in the direct sum of $\mathbb{Q}$ and a cubic field (or some other étale cubic algebra) rather than a quartic field. In the third cusp, another very interesting phenomenon occurs, namely the lattice points inside almost entirely correspond to orders in quartic fields whose Galois closure has Galois group D_4 (the dihedral group of order 8) rather than S_4! Meanwhile, the main body of the region away from the cusps is shown to consist

almost entirely (i.e., up to $o(X)$) of the lattice points that correspond to orders in S_4-quartic fields.[7]

As a result, to count orders in S_4-quartic fields (i.e., "S_4-quartic orders"), one may simply count lattice points in the region $\mathcal{F}_X$ with its tentacles cut off. This region is then compact, and is sufficiently round for one to deduce that the number of lattice points inside this region is essentially its volume, which is computed to be $(5\zeta(2)^2\zeta(3)/24)X$. It follows (in conjunction with Theorem 5.1) that the number of pairs (Q, R), where Q is an S_4-quartic order of discriminant at most X and R is a cubic resolvent ring of Q, is asymptotic to $(5\zeta(2)^2\zeta(3)/24)X$ as $X \to \infty$.

Finally, one shows that counting quartic rings just once each – i.e., without weighting by the number of cubic resolvents – affects this final answer simply by a factor of $\zeta(5)$. We obtain:

Theorem 6.4. *The number of S_4-quartic orders having absolute discriminant at most X is asymptotic to* $\frac{5\zeta(2)^2\zeta(3)}{24\,\zeta(5)} X$ *as $X \to \infty$.*

To count only the maximal orders in S_4-quartic fields again requires a fairly delicate sieve. The end result of this sieve is:

Theorem 6.5. *The number of S_4-quartic fields having absolute discriminant at most X is asymptotic to* $\frac{5}{24} \prod_p (1 + p^{-2} - p^{-3} - p^{-4}) \cdot X$ *as $X \to \infty$.*

Thus $c_4 = \frac{5}{24} \prod_p (1 + p^{-2} - p^{-3} - p^{-4}) \approx .253\ldots$; that is, we expect about $.253$ S_4-quartic fields per discriminant. $\square$

Theorem 6.5 has a number of interesting consequences. First, it is related to the proof of the case of the Cohen–Lenstra–Martinet class group heuristics mentioned at the end of Section 4.3. Second, in conjunction with the work of Baily [1] and Cohen–Diaz–Olivier [13] on D_4-fields, Theorem 6.5 implies that when all quartic fields are ordered by the size of their discriminants, a positive proportion of them *do not* have Galois group S_4! In fact, "only" about 90.644% have Galois group S_4, while the remaining correspond to the Galois group D_4 (0% have any of the other possible Galois groups). This is interesting because it is in direct opposition to the situation for polynomials, where Hilbert's irreducibility theorem implies that if integer polynomials of degree n are ordered by the size of their coefficients, then 100% will have Galois group S_n.

$n = 5$. Last but not least, the parametrization results described in the previous section also allow one to asymptotically determine the number of quintic fields of bounded discriminant. This represents the first instance where one can count *unsolvable* extensions.

---

[7]In practice, to simplify the details, we perform all these computations using not one but several fundamental domains $\mathcal{F}_X \subset \mathbb{R}^{12}$, but the spirit of the argument remains unchanged. The details of this "averaging" method may be found in [9] and [10].

We have shown that quintic rings correspond to the $GL_4(\mathbb{Z}) \times SL_5(\mathbb{Z})$-orbits on quadruples of 5×5 skew-symmetric integer matrices. Following the cases $n = 3$ and $n = 4$, we begin by constructing a fundamental domain for the action of $GL_4(\mathbb{Z}) \times SL_5(\mathbb{Z})$ on the corresponding forty-dimensional real vector space V. We wish to understand the number of integer points in $\mathcal{F}_X$, where as before $\mathcal{F}_X = \mathcal{F} \cap \{v \in V : |\mathrm{Disc}(v)| \leq X\}$. This turns out to be significantly more difficult than the corresponding problem in the cubic and quartic cases. Besides being highly non-compact, the forty-dimensional fundamental domain $\mathcal{F}_X$ has an intriguingly complex system of numerous high-dimensional tentacles and cusps!

But in the end these cusps too exhibit a surprisingly beautiful structure, and can be handled in much the same way as in the cubic and quartic cases. Namely, we cut up this system of cusps into a finite number (approximately 160) of sub-cusps, all of which run off to infinity and thus present a problematic tentacle-type scenario. In each one of these 160 sub-cusps, one shows either that there are a negligible number of points within, *or* that essentially all points in that tentacle are reducible in a certain way. In this aspect, these cusps are very similar to those occurring in the cases $n = 3$ and $n = 4$ – the difference being only that there are *many* more of them this time!

Lastly, one shows that 100% of the points in the main body of the region correspond to orders in S_5-quintic fields. By computing the volume of this main body, and then sieving down to the maximal quintic orders (for details on these tasks, see [10]), one finally obtains the following theorem.

Theorem 6.6. *The number of quintic fields having absolute discriminant at most X is asymptotic to* $\frac{13}{120} \prod_p (1 + p^{-2} - p^{-4} - p^{-5}) \cdot X$ *as $X \to \infty$.*

Therefore $c_5 = \frac{13}{120} \prod_p (1 + p^{-2} - p^{-4} - p^{-5}) \approx .149\dots$, and so there are about .149 quintic fields per discriminant on average.   □

$n \geq 6$? Given the success in the cases $n \leq 5$, the question is only too tempting: what happens for $n \geq 6$? We put forth the following conjecture:

Conjecture 6.7. We have

$$c_n = \frac{r_2(S_n)}{2 \cdot n!} \prod_p \left(\sum_{k=0}^{n} \frac{q(k, n - k) - q(k - 1, n - k + 1)}{p^k} \right) \qquad (10)$$

where $q(i, j)$ denotes the number of partitions of i into at most j parts, and $r_2(S_n)$ denotes the number of 2-torsion elements in S_n.

That is, we conjecture that the number of S_n-number fields per discriminant will be c_n on average, where c_n is given as in (10).

Conjecture 6.7 was obtained by combining global heuristics with new mass formulae for étale extensions of local fields inspired by work of Serre [36]. It is readily checked that Conjecture 6.7 agrees with the values of c_n now proven for $n = 1$ through 5. For further details on this conjecture and the related mass formulae, see [11]. The proofs of the conjecture for $n = 3$, 4, and 5 may be found in [20], [9], and [10] respectively.

7. Related and future work

The composition and parametrization laws described in Sections 3–5 all turn out to be closely related to certain exceptional Lie groups. More precisely, let E be an exceptional Lie group and let P be a maximal parabolic of E. If we write $E = GU$, where G is the Levi factor and U is the unipotent radical at P, then the group G acts naturally (by conjugation) on the abelianized unipotent radical $V = U/[U, U]$. For appropriate choices of E and P, we find that we obtain precisely the prehomogeneous vector spaces (G, V) underlying the composition laws and parametrizations described in Sections 3–5. For example, the first case we considered in Section 3 was the space of $2 \times 2 \times 2$ cubes, and this representation of $SL_2(\mathbb{Z}) \times SL_2(\mathbb{Z}) \times SL_2(\mathbb{Z})$ arises in this way when E is the exceptional Lie group of type D_4 and P is the Heisenberg parabolic.

This remarkable connection with Lie groups in fact appears to run much further – see [4, §4] and [5, §4] – and needs exploration, perhaps in connection with automorphic forms on these groups in the sense of Gan–Gross–Savin [26] and in the subsequent work of Lucianovic [31] and Weissman [38]. The reduction-theoretic aspect of some of the exceptional representations that arise in this way and their relation to noncommutative rings has been the subject of study in the recent work of Krutelevich.

The parametrization results described in Sections 3–5 for commutative rings extend to a large extent also to many noncommutative rings such as quaternions, octonions, and higher rank division algebras (see [8]). Many of these parametrizations of noncommutative rings and modules were discovered by applying certain number-theoretic operations (mentioned in Section 3.4) to parametrizations involving quadratic and cubic rings, indicating that there is a great deal of number theory lurking behind noncommutative – and even nonassociative! – rings such as the octonions. These number-theoretic connections beg for further investigation.

We note that the spaces underlying these various parametrizations also come equipped with a theory of zeta functions. Zeta functions associated to prehomogeneous vector spaces were first introduced by Sato and Shintani [34], and were further developed by Datskovsky, Wright, Yukie, and others. In particular, Datskovsky and Wright [16] used such zeta functions to give an alternative proof of Davenport–Heilbronn's theorem, which applies over an arbitrary number field or function field. The more difficult quartic analogue of their work was initiated by Yukie [42], and could eventually lead to an alternative method for counting quartic fields. The problem of understanding the relationship between the various parametrizations discussed here and the associated zeta functions is intriguing and deserves further investigation, both in the commutative and noncommutative cases.

Regarding commutative rings, the problem of finding parametrizations for rings of rank > 5 is a very interesting one. Although we have already noted that commutative rings cannot be parametrized by prehomogeneous vector spaces beyond the quintic case, there may be other ways to accomplish the task, such as through the study of

integer points on certain special varieties. This is a central problem in the theory, and of importance not only algebraically but also with respect to understanding the distribution of algebraic number rings and fields of higher degree.

As to our Conjecture 6.7 on counting S_n-number fields having fixed degree n and absolute discriminant less than X, even the correct order of growth (i.e., $O(X)$) is not known. The best general bounds known for $n \geq 6$ are due to Ellenberg and Venkatesh [24], who prove a bound of $O(X^{n^{\varepsilon}})$. Conjectures for the density of discriminants of degree n number fields having a specified Galois group G (yielding the expected orders of growth but not the constants) have been suggested by Malle [32]. These conjectures have been proven for many specific cases of G, including S_n for $n \leq 5$ (see Section 6), abelian groups (Wright [40]), D_4 (Cohen–Diaz–Olivier [13]), and certain nilpotent groups (Klüners–Malle [30]). The constants $c(G)$ in these conjectures for $G \neq S_n$ are unknown in general, even conjecturally, although there has been some recent progress. If the case $G = S_n$ is any indication, the constants $c(G)$ likely contain a great deal of arithmetic information.

One important ingredient in the case $G = S_n$ in determining the corresponding constants $c_n = c(S_n)$ was the development of mass formulae that count étale extensions of local fields by appropriate weights (see [11]). How these mass formulae change with G is an intriguing question, and several interesting cases and families of finite groups G have been treated by Kedlaya [29] and more recently by Wood. The manner in which these various local mass formulae glue together globally to give the global constants $c(G)$ is still an open question.

The counting arguments described in Section 6 can be taken much further, leading e.g. to further information on the distribution of class numbers, narrow class numbers, units, and regulators of cubic rings and fields. They can also be used to obtain information on the discriminant density of noncommutative rings such as quaternion and octonion rings, and modules over these rings. Finally, results analogous to those described in this survey can be obtained for ring and field extensions not just over $\mathbb{Z}$ and $\mathbb{Q}$ but over more general base rings. These directions too must be investigated, and we hope to treat them further in future work.

References

[1] Baily, A. M., On the density of discriminants of quartic fields. *J. Reine Angew. Math.* **315** (1980), 190–210.

[2] Belabas, K., Bhargava, M., Pomerance, C., Error terms for the Davenport-Heilbronn theorems. Preprint.

[3] Bhargava, M., Higher Composition Laws. Ph.D. Thesis, Princeton University, 2001.

[4] Bhargava, M., Higher composition laws I: A new view on Gauss composition. and quadratic generalizations. *Ann. of Math.* **159** (1) (2004), 217–250.

[5] Bhargava, M., Higher composition laws II: On cubic analogues of Gauss composition. *Ann. of Math.* **159** (2) (2004), 865–886.

[6] Bhargava, M., Higher composition laws III: The parametrization of quartic rings. *Ann. of Math.* **159** (3) (2004), 1329–1360.

[7] Bhargava, M., Higher composition laws IV: The parametrization of quintic rings. *Ann. of Math.*, to appear.

[8] Bhargava, M., Higher composition laws V: The parametrization of quaternionic and octonionic rings and modules. In preparation.

[9] Bhargava, M., The density of discriminants of quartic rings and fields. *Ann. of Math.* **162** (2005), 1031–1063.

[10] Bhargava, M., The density of discriminants of quintic rings and fields. *Ann. of Math.*, to appear.

[11] Bhargava, M., Mass formulae for local extensions and conjectures on the density of number field discriminants. Preprint.

[12] Brahmagupta, *Brahma-sphuṭa-siddhānta*. 628.

[13] Cohen, H., Diaz y Diaz, F., Olivier, M., Counting discriminants of number fields of degree up to four. In *Algorithmic Number Theory* (Leiden, 2000), Lecture Notes in Comput. Sci. 1838, Springer-Verlag, New York 2000, 269–283.

[14] Cohen, H., Martinet J., Étude heuristique des groupes de classes des corps de nombres. *J. Reine Angew. Math.* **404** (1990), 39–76.

[15] Cohen, H., Martinet J., Heuristics on class groups: some good primes are not too good. *Math. Comp.* **63** (1994), 329–334.

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

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

[18] Davenport, H., On the class-number of binary cubic forms I and II. *J. London Math. Soc.* **26** (1951), 183–198.

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

[20] Davenport, H., Heilbronn, H., On the density of discriminants of cubic fields II. *Proc. Roy. Soc. London Ser. A* **322** (1551) (1971), 405–420.

[21] Delone, B. N., Faddeev, D. K., *The theory of irrationalities of the third degree*. Transl. Math. Monographs 10, Amer. Math. Soc., Providence, R.I., 1964

[22] Dirichlet, P. G. L., *Zahlentheorie*. 4th. edition, Vieweg, Braunschweig 1894.

[23] Eisenstein, G., Théorèmes sur les formes cubiques et solution d'une équation du quatrième degré indéterminées. *J. Reine Angew. Math.* **27** (1844), 75–79.

[24] Ellenberg, J., Venkatesh, A., The number of extensions of a number field with fixed degree and bounded discriminant. *Ann. of Math.* **163** (2) (2006), 723–741.

[25] Ennola, V., Turunen, R., On totally real cubic fields. *Math. Comp.* **44** (1985), 495–518.

[26] Gan, W.-T., Gross, B. H., Savin, G., Fourier coefficients of modular forms on G_2. *Duke Math. J.* **115** (1) (2002), 105–169.

[27] Gauss, C. F., *Disquisitiones Arithmeticae*. Leipzig, 1801.

[28] Hilbert, D., *Theory of Algebraic Invariants*. Engl. trans. by R. C. Laubenbacher, Cambridge University Press, Cambridge 1993.

[29] Kedlaya, K. S., Mass formulas for local Galois representations (after Serre, Bhargava). Preprint.

[30] Klüners, J., Malle, G., Counting nilpotent Galois extensions. *J. Reine Angew. Math.* **572** (2004), 1–26.

[31] Lucianovic, M., Quaternion Rings, Ternary Quadratic Forms, and Fourier Coefficients of Modular Forms on PGSp(6). Ph.D. Thesis, Harvard University, 2003.

[32] Malle, G., On the distribution of Galois groups. *J. Number Theory* **92** (2002), 315–329.

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

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

[35] Serre, J-P., Modules projectifs et espaces fibrés à fibre vectorielle. *Séminaire Dubreil-Pisot* 1957/58, no. 23.

[36] Serre, J-P., Une "formule de masse" pour les extensions totalement ramifiées de degré donné d'un corps local. *C. R. Acad. Sci. Paris Sér. A-B* **286** (22) (1978), A1031–A1036.

[37] Vinberg, E. B., On the classification of the nilpotent elements of graded Lie algebras. *Soviet Math. Dokl.* **16** (1975), 1517–1520.

[38] Weissman, M., D_4 Modular Forms. *Amer. J. Math*, to appear.

[39] Wong, S., Automorphic forms on GL(2) and the rank of class groups. *J. Reine Angew. Math.* **515** (1999), 125–153.

[40] Wright, D. J., Distribution of discriminants of abelian extensions. *Proc. London Math. Soc.* (3) **58** (1989), 17–50.

[41] Wright, D. J., Yukie, A., Prehomogeneous vector spaces and field extensions. *Invent. Math.* **110** (1992), 283–314.

[42] Yukie, A., *Shintani Zeta Functions*. London Math. Soc. Lecture Note Ser. 183, Cambridge University Press, Cambridge 1993.

[43] Yukie, A., Density theorems for prehomogeneous vector spaces. Preprint.

Department of Mathematics, Princeton University, Princeton, NJ 08544, U.S.A.

E-mail: bhargava@math.princeton.edu