

COUNTING MODULAR FORMS BY RATIONALITY FIELD

ALEX COWAN AND KIMBALL MARTIN

ABSTRACT. We investigate the distribution of degrees and rationality fields of weight 2 newforms. In particular, we give heuristic upper bounds on how often degree d rationality fields occur for squarefree levels, and predict finiteness if $d \geq 7$. When $d = 2$, we make predictions about how frequently specific quadratic fields occur, prove lower bounds, and conjecture that $\mathbb{Q}(\sqrt{5})$ is the most common quadratic rationality field.

1. INTRODUCTION

Let $S_k(N) = S_k(\Gamma_0(N))$ be the space of holomorphic cusp forms of weight k and level N with trivial nebentypus. For a newform $f \in S_k(N)$, denote by K_f its rationality field, i.e. the number field generated by its Hecke eigenvalues. Define the (*rationality*) *degree* and *discriminant* of f to be the degree and discriminant of K_f/\mathbb{Q} respectively. In what follows, we always assume trivial nebentypus.

Weight 2 newforms are of special interest as they correspond to modular abelian varieties, i.e., simple factors of the Jacobian $J_0(N)$ of $X_0(N)$. Namely (the Galois orbit of) a degree d newform $f \in S_2(N)$ corresponds to a d -dimensional simple abelian subvariety of $J_0(N)$ which has conductor N^d . The $d = 1$ case is the celebrated bijection between rational newforms in $S_2(N)$ and isogeny classes of elliptic curves of conductor N .

It is expected (e.g., see [35, 23, 28]) that the Galois orbit of a newform is “as large as possible” 100% of the time, so that newforms have small degree rather infrequently. On the other hand, there are a relatively large number of elliptic curves of small conductor. Watkins’ [39] refinement of the Brumer–McGuinness heuristics [7] for counting elliptic curves suggests that the number of weight 2 rational newforms with level $N < X$ grows like $cX^{5/6}$ for some computable constant c . See [12, 36] for some theoretical results towards this growth rate. Note that the total number of weight 2 newforms of level $N < X$ grows roughly like X^2 .

Here we consider the questions: how many weight 2 newforms of level $N < X$ are there with a given degree d or a given rationality field K ? There is no analogue of the Brumer–McGuinness heuristics for $d > 1$, since those rely on having simple equations for elliptic curves. Moreover, as degree $d > 1$ forms are relatively rare, it is difficult to generate enough data to predict precise asymptotics based on calculations.

In fact, even for $d = 1$, it is difficult to make accurate predictions based solely on computations. E.g., as remarked in [39], the growth rate in Cremona’s database of elliptic curves is about $X^{0.98}$; however more recent and very extensive calculations for prime conductors in [2] align closely with the $X^{5/6}$ heuristic.

Date: January 14, 2024.

AC was supported by the Simons Foundation (Collaboration Grant 550031).

KM was supported by the Simons Foundation (Collaboration Grant 512927), the Japan Society for the Promotion of Science (Invitational Fellowship L22540), and the Osaka Central Advanced Mathematical Institute (MEXT Joint Usage/Research Center on Mathematics and Theoretical Physics JPMXP0619217849).

Using a combination of heuristics and data, we predict some bounds on asymptotic orders of growth, and the relative frequency of such forms.

Conjecture 1.1. *Let $\varepsilon > 0$. The number of degree d weight 2 newforms of squarefree level $N \leq X$ is $O(X^{1-d/6+\varepsilon})$ as $X \rightarrow \infty$. In particular, this number is finite if $d \geq 7$.*

Conjecture 1.2. *Among squarefree levels $N \rightarrow \infty$, 100% of degree 2 newforms in $S_2(N)$ have rationality field $\mathbb{Q}(\sqrt{5})$.*

Remark 1.3. The heuristics for these conjectures do not require a restriction to squarefree levels, however there are special considerations for non-squarefree levels. First, one should only count quadratic twist classes for a more general analogue of [Conjecture 1.1](#). Second, CM forms (which do not occur in squarefree level with trivial nebentypus) deserve separate consideration. Third, if $p^r \mid N$ for sufficiently large r , then the rationality field of a newform f of level N must contain a certain cyclotomic subfield (e.g., if $p \geq 5$ and $r \geq 3$, then $K_f \supset \mathbb{Q}(\zeta_p)^+$)—see [\[6, 26\]](#).

It is at least plausible that [Conjecture 1.2](#) holds for general levels, and [Conjecture 1.1](#) holds for general levels if one restricts to counting non-CM newforms up to quadratic twist. However, our data are much more limited for non-squarefree levels.

[Conjecture 1.1](#) is just a conjectural upper bound, and it may not be sharp for $2 \leq d \leq 5$ (see below for more discussion). When $d = 1$, one can prove a lower bound of order $X^{5/6}$ for elliptic curves, but we are not aware of nontrivial analogous lower bounds (or even a proof of infinitude!) for any $d > 1$. Using constructions of genus 2 curves with real multiplication, we obtain the following lower bounds for $d = 2$, without a restriction to squarefree level.

Proposition 1.4. *The number of quadratic twist classes of weight 2 newforms with rationality field $\mathbb{Q}(\sqrt{5})$ (resp. $\mathbb{Q}(\sqrt{2})$) and minimal level $N < X$ is $\gg X^{1/3}$ (resp. $\gg X^{2/7}$).*

The same result for squarefree levels would follow if one knew certain polynomials took on squarefree values sufficiently often.

Remark 1.5. It is not even clear for which $d \leq 6$ there should exist infinitely many weight 2 newforms of squarefree level. Constructions of genus 3 curves with real multiplication suggest it may be infinite for $d = 3$ —see [Section 3.4](#). For $d = 4, 5, 6$, we have little theoretical evidence, but our data suggest these counts are infinite at least for each $d \leq 4$.

We will consider two approaches to predicting counts of newforms with fixed degree or rationality field. First, in [Section 2](#), we present a heuristic using a random model for Hecke polynomials, building off of [\[33, 28\]](#). In fact this random model naively suggests upper and lower bounds for counts of degree d forms on the order of $X^{1-d/6 \pm \varepsilon}$. However, it ignores any geometric considerations for the existence of degree d forms, so it is unclear how accurate this heuristic is. Nevertheless, comparing these predictions with data at least suggests it gives an upper bound, as asserted in [Conjecture 1.1](#).

In [Section 3](#), we suggest an approach to predict counts of weight 2 newforms with a given rationality field K by counting moduli points for suitable abelian varieties. In principle, this would also yield the number of counts of newforms of a fixed degree d , and we expect this approach should give more accurate predictions than the random Hecke polynomial model. However, it requires more knowledge about the moduli spaces and the relation between heights and conductors than we currently possess. We carry out some of this analysis when $d = 2$, namely when $K = \mathbb{Q}(\sqrt{D})$ for $D = 5, 8, 12, 13, 17$. This leads to [Conjecture 1.2](#), and also

suggests $cX^{3/5-\varepsilon}$ may be a lower bound for the total count for degree 2 newforms. However, our analysis is not definitive enough to confidently conjecture this.

A database of all prime-level forms of degree 6 or less and level $2 \cdot 10^6$ or less was computed using an algorithm of the first author [8]. In [Section 4](#), we use this database to investigate [Conjecture 1.1](#) and [Conjecture 1.2](#), and pose some related questions.

Acknowledgements. We are especially grateful to Noam Elkies for many insights and suggestions. We have also benefited from conversations with Eran Assaf, Armand Brumer, Bjorn Poonen, Ari Shnidman, and John Voight. Computations were performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU).

2. COUNTS BY DEGREE

First we discuss counting newforms of fixed degree. For a newform f , let $\deg f = [K_f : \mathbb{Q}]$ be its rationality degree. Set

$$\mathcal{C}_d(X) = \#\{\text{newforms } f \in S_2(N) : N < X, N \text{ squarefree, } \deg f = d\}.$$

As explained in the introduction, we restrict to squarefree N for simplicity, though our initial discussion applies equally well to counting quadratic twist classes of non-CM weight 2 newforms.

Watkins [39], building on heuristics of Brumer and McGuinness [7], formulates heuristics that suggest

$$(1) \quad \mathcal{C}_1(X) \sim c_1 X^{\frac{5}{6}}$$

for some computable constant c_1 . It is known that $X^{\frac{5}{6}} \ll \mathcal{C}_1(X) \ll X^{1+\varepsilon}$ [12]. Furthermore, Shankar–Shankar–Wang [36] show a growth rate of $b_1 X^{\frac{5}{6}}$ if one restricts to elliptic curves of squarefree conductor coprime to 6 with some restrictions on discriminant-conductor and discriminant-height ratios.

While the exponent $\frac{5}{6}$ is not in clear agreement with databases of elliptic curves in general levels (the Cremona [10] and Stein–Watkins [37] databases), the compatibility with prime-level data is much better. Namely, Watkins’ heuristic suggests a growth rate of $c'_1 \text{li}(X^{\frac{5}{6}})$ for prime levels, and this fits extremely well with the extensive database of elliptic curves of prime conductor in [2]. Thus there is much evidence towards (1).

For $d > 1$, the situation is much more mysterious. In [35], Serre proves a statement which strongly suggests, though does not quite imply, the bound $\mathcal{C}_d(X) = o(X^2)$. Namely, if $N \rightarrow \infty$ along a sequence which is coprime to a fixed prime ℓ , among bases of eigenforms for $S_2(N)$, Serre proves that the number of forms of degree d is $o(\dim(S_2(N)))$ as $N \rightarrow \infty$. Serre’s theorem was made effective by Murty and Sinha [31], and more recently by Sarnak and Zubrilina [34].

Since we do not know a good way to predict precise asymptotics for $\mathcal{C}_d(X)$, we aim to predict weaker estimates of the form

$$(2) \quad X^{\alpha_d} \ll \mathcal{C}_d(X) \ll X^{\beta_d}$$

which are nontrivial, i.e., $\alpha_d > 0$ or $\beta_d < 2$. Computations of modular forms, as well as heuristics in [28], suggest β_d is decreasing in d , and thus we should at least be able to take $\beta_d \leq \frac{5}{6}$ for each $d \geq 1$. In [28, Question 3.1], it was also suggested that one may have $\beta_d = 0$ for $d \gg 0$.

To our knowledge, [Conjecture 1.1](#) is the first prediction of more precise upper bounds (for either squarefree or general levels). In particular, it predicts that one can take $\beta_d = 0$ for

$d \geq 7$, and β_d arbitrarily small for $d = 6$. However, we do not have insight into whether the upper bounds in [Conjecture 1.1](#) should be sharp for $2 \leq d \leq 5$.

Note that [Conjecture 1.1](#) implies that $\alpha_d = 0$ is optimal among lower bounds of the form (2) for $d \geq 6$. In addition, [Proposition 1.4](#) suggests that one may take $\alpha_2 \geq \frac{1}{3}$ for $d = 2$. (Note that [Proposition 1.4](#) does not prove a lower bound for squarefree levels, only for general levels.) This lower bound is almost certainly not sharp. We do not have any predictions for lower bounds when $3 \leq d \leq 5$.

2.1. Random Hecke polynomial model. Here we present a random model to estimate the distribution of degree d newforms that will lead us to [Conjecture 1.1](#). This is based on ideas for heuristics suggested in [\[33\]](#) and [\[28\]](#).

Consider a newspace $S_{2^k}^{\text{new}}(N)$. One can further decompose this space into $2^{\omega(N)}$ joint eigenspaces of the Atkin–Lehner operators W_p for $p \mid N$, which we call the Atkin–Lehner eigenspaces. Each Atkin–Lehner eigenspace is Galois invariant. For non-squarefree levels, one can further decompose each Atkin–Lehner eigenspace into smaller Galois invariant subspaces according to local inertia types of non-CM forms (see [\[11\]](#)) and the subspace of CM forms.

For simplicity, assume N is squarefree. Then there are no CM forms of trivial nebentypus and there is only one local inertial type. Let S be an Atkin–Lehner eigenspace in $S_{2^k}^{\text{new}}(N)$. For a newform $f \in S$, the single Fourier coefficient $a_p(f)$ generates K_f for 100% of p [\[22\]](#), and it is conjectured that this is true for all but finitely many p if $[K_f : \mathbb{Q}] > 4$ [\[32\]](#). Hence, for fixed $p \nmid N$, the factorization type of the characteristic polynomial $c_{T_p}(x) \in \mathbb{Z}[x]$ of the Hecke operator T_p will usually tell us the degrees of the newforms in S . In fact, it will always give us lower bounds.

Let $n = \dim S$. As in [\[33\]](#) and [\[28\]](#), we can model $c_{T_p}(x)$ as a random polynomial in the set $H_n = H_n(k, p)$ of degree n monic integral polynomials whose roots α satisfy $|\alpha| \leq 2p^{k-1/2}$. Alternatively, one can consider the set of Weil q -polynomials of degree $2n$ where $q = p^k$, or the isogeny classes of n -dimensional abelian varieties over \mathbb{F}_{p^k} .

Set $h(n) = \#H_n$. As discussed in [\[28, §2.1\]](#), the number of polynomials in H_n with a degree $d < \frac{n}{2}$ factor is approximately $h(d)h(n-d)$. Thus, if we select polynomials in H_n uniformly at random, then

$$(3) \quad \text{Prob}(p \in H_n \text{ has a degree } d \text{ factor}) \approx \frac{h(d)h(n-d)}{h(n)}.$$

In this section, by approximately (\approx), we mean that for fixed d both sides have the same growth rate in n as $n \rightarrow \infty$.

For fixed q , no good asymptotics are known for $h(n)$ to directly estimate this probability. There is an asymptotic for $h(n)$ when n is fixed and q varies. Instead, [\[33\]](#) and [\[28, §2.1\]](#) analyzed how this probability behaves if one uses a known asymptotic for $\#H_n(k, p)$ in p^k when n is fixed. (The result is certainly too small, as it would predict only finitely many degree 1 forms.)

To circumvent this lack of precise asymptotics for $h(n)$ as $n \rightarrow \infty$, we rewrite the right hand side of (3) as

$$\frac{h(d)h(n-d)}{h(n)} = h(d) \frac{h(n-d)}{h(n-d+1)} \frac{h(n-d+1)}{h(n-d+2)} \cdots \frac{h(n-1)}{h(n)}.$$

One should have that $\frac{h(n-2)}{h(n-1)} \approx \frac{h(n-1)}{h(n)}$, so applying this a small fixed number of times for a given d yields

$$(4) \quad \frac{h(d)h(n-d)}{h(n)} \approx \left(\frac{h(n-1)}{h(n)} \right)^d.$$

Combining (3) and (4) suggests that the probability of a degree d factor of c_{T_p} should approximately be the d -th power of the probability of a degree 1 factor of c_{T_p} . The latter typically corresponds to a degree 1 form, and so we can model it using well-known expectations about counts of elliptic curves.

We will also use the following lemma.

Lemma 2.1. *Let $\nu_{2k}(X)$ be the number of Atkin–Lehner eigenspaces in $\bigcup_N S_{2k}^{\text{new}}(N)$, where N ranges over squarefree levels, having dimension less than X . Then $X \ll \nu_{2k}(X) \ll X^{1+\varepsilon}$, for any $\varepsilon > 0$.*

Proof. Since $\dim S_{2k}^{\text{new}}(N) \ll N$, the lower bound is obvious.

Let us show the upper bound. First, it follows from the dimension formulas for Atkin–Lehner eigenspaces from [27] that any Atkin–Lehner eigenspace in $S_{2k}^{\text{new}}(N)$ has dimension $\frac{(k-1)\phi(N)}{12 \cdot 2^{\omega(N)}} + O(N^{1/2+\varepsilon})$. (The necessary argument, though not the statement, is given in the proof of [29, Proposition 3.10].) This dimension is $\gg N^{1-\varepsilon}$. Hence if an Atkin–Lehner eigenspace has dimension less than X , it occurs in a level $N \ll X^{1+\varepsilon}$. Now the number of Atkin–Lehner eigenspaces in levels less than t is bounded by $\sum_{N \leq t} 2^{\omega(N)}$. It is known that this latter sum is $\frac{6}{\pi^2} t \log t + O(t)$. \square

Lemma 2.1 combines with (3) and (4) to produce the following heuristic.

Heuristic 2.2. *Suppose the number of rational newforms of weight $2k$ and squarefree level $N < X$ is $O(X^{1-\alpha})$ for some $\alpha < 1$. Then, for any $\varepsilon > 0$, the number of degree d weight $2k$ newforms of squarefree level $N \leq X$ is $O(X^{1-\alpha d+\varepsilon})$ as $X \rightarrow \infty$.*

Our reasoning for this heuristic is as follows. Under the hypothetical bound $O(X^{1-\alpha})$, the lemma indicates that the probability of an Atkin–Lehner space of dimension n having a size 1 Galois orbit is approximately $n^{-\alpha}$. Assuming uniform distribution of Hecke polynomials in $H(n)$, the probability of a size d Galois orbit is approximately the probability of a degree d factor of c_{T_p} , which by (3) and (4) is approximately $n^{-d\alpha}$. Applying the lemma again leads to the stated heuristic.

Combining the Brumer–McGuinness and Watkins heuristics for $d = 1$ with Heuristic 2.2 now suggests Conjecture 1.1 from the introduction.

2.2. Assessment of the model. The random model for Hecke polynomials in Section 2.1 uses the counting measure on H_n , i.e., all polynomials in H_n are equally likely. If this were the case, the heuristic reasoning above would suggest both upper and lower bounds: $X^{1-\frac{d}{6}-\varepsilon} \ll \mathcal{C}_d(X) \ll X^{1-\frac{d}{6}+\varepsilon}$, so the upper bound in Conjecture 1.1 would be essentially optimal.

However, there are other factors controlling the distribution of Hecke polynomials in H_n . For instance, trace formulas place arithmetic conditions on the roots of Hecke polynomials. Moreover, there are vertical and horizontal equidistribution results about convergence of the roots to Plancherel and Sato–Tate measures. For this reason, we view our random model as a first approximation to counting degree d forms.

In Section 4.1, we will present data which suggests this heuristic does give an upper bound, but possibly not an optimal one. It is really the data that lends credence to Conjecture 1.1.

An alternative perspective, which we will explore in [Section 3](#), is that it is more natural to model the distribution of degree d forms by modeling d -dimensional modular abelian varieties. The analysis we do there for $d = 2$ is compatible with the notion that the random Hecke polynomial heuristic gives a valid upper bound which might not be optimal.

2.3. Finiteness questions. Related to the question of asymptotics are several questions about finiteness. We do not investigate them here, but suggest them for future consideration.

- (1) We can ask: for what d are there infinitely many weight 2 newforms of squarefree level? [Conjecture 1.1](#) asserts such d must be at most 6, but also suggests the answer could be negative for $d = 6$. We know a positive answer for $d = 1$, and expect a positive answer for $d = 2$. [Section 3.4](#) and our data suggest the answers may be positive for $d = 3, 4$ also.
- (2) More generally, one can ask the same question in weight $2k$. Roberts' conjecture [\[33\]](#) implies that there are only finitely many quadratic twist classes of non-CM rational newforms in weight $2k \geq 6$. So [Heuristic 2.2](#) suggests that there are only finitely many newforms of squarefree level of fixed weight $2k \geq 6$ and any fixed degree $d \geq 1$. One might similarly expect to have finitely many quadratic twist classes of non-CM newforms of fixed degree d and weight $2k \geq 6$.
- (3) One can also ask whether there should be a uniform version of the finiteness part of [Conjecture 1.1](#), i.e., whether for sufficiently large squarefree N and some fixed d_0 (possibly $d_0 = 6$), each Atkin–Lehner eigenspace has a unique Galois orbit of size $d \geq d_0$. This seems plausible based on [Heuristic 2.2](#). See [Question 4.4](#) for a more precise question in prime level.

3. HECKE FIELDS

In [Section 2](#) we considered the question of how often degree d newforms occur and presented a random Hecke polynomial model, which, at least for prime levels, appears to give asymptotic upper bounds. Here we consider the refined question of how often a specific degree d rationality field K should occur, and relate this question to rational points on Hilbert modular varieties. We discuss possible lower bounds for fixed quadratic fields, prove some lower bounds, and predict that $\mathbb{Q}(\sqrt{5})$ is the most common quadratic rationality field.

3.1. Modular varieties. First recall the connection between weight 2 modular forms and abelian varieties.

Let $N \geq 1$. To a newform $f \in S_2(N)$ with $[K_f : \mathbb{Q}] = d$, Shimura constructed a d -dimensional simple abelian variety A_f/\mathbb{Q} satisfying the following properties. First, A_f is a quotient of $J_0(N)$. Moreover A_f is isogenous to A_g if and only if f and g are Galois conjugates. The endomorphism algebra $\text{End}^0(A_f) := \text{End}(A_f) \otimes \mathbb{Q} \simeq K_f$. The conductor of A_f is N^d . Finally, $L(s, A_f) = \prod_{\sigma} L(s, f^{\sigma})$, where f^{σ} ranges over the Galois conjugates of f .

In general, the center of the endomorphism algebra of a d -dimensional abelian variety A has degree $\leq d$. If $\text{End}^0(A)$ contains a totally real field K of degree $d = \dim A$, then we say A has maximal real multiplication (RM). Any A_f as above has maximal RM, and conversely if A/\mathbb{Q} is a simple abelian variety with maximal RM, then it is isogenous to some A_f [\[26, Lemma 3.1\]](#). Hence the correspondence $f \mapsto A_f$ yields a bijection between degree d newforms f of weight 2 and isogeny classes of d -dimensional simple abelian varieties A/\mathbb{Q} with maximal RM.

We propose a heuristic approach to predicting coarse asymptotic counts of such objects. Let K be a totally real number field of degree d , and \mathfrak{a} be an ideal in \mathcal{O}_K . The quotient

$\mathfrak{H}^d/\mathrm{SL}(\mathcal{O}_K \oplus \mathfrak{a})$ parametrizes d -dimensional complex abelian varieties with RM by \mathcal{O}_K together with a polarization structure corresponding to \mathfrak{a} (see [15] for a precise statement). Compactifying this quotient and desingularizing gives a Hilbert modular variety $Y(\mathcal{O}_K \oplus \mathfrak{a})$.

Now consider a newform $f \in S_2(N)$ with $K_f = K$. The abelian variety A_f has endomorphism ring an order in \mathcal{O}_K . Typically we expect it is all of \mathcal{O}_K , but if not, one can replace A_f by an isogenous variety with RM by \mathcal{O}_K . Thus f corresponds to a rational point y on $Y(\mathcal{O}_K \oplus \mathfrak{a})$ for some \mathfrak{a} , which we can take to be in a given set of representatives for $\mathrm{Cl}^+(K)$.

This correspondence is far from one-to-one. First, replacing A_f by an isogenous variety, or modifying the polarization structure, may give a different point y on $Y(\mathcal{O}_K \oplus \mathfrak{a})$. Second, if g is another weight 2 newform and A_g is \mathbb{C} -isogenous to A_f , then both f and g correspond to the same rational points. Third, this is not a fine moduli space, so not all rational points on $Y(\mathcal{O}_K \oplus \mathfrak{a})$ will correspond to abelian varieties defined over \mathbb{Q} , and of those that do, some will correspond to non-simple abelian varieties.

That said, it seems reasonable to expect that, generically, quadratic twist classes of Galois orbits of weight 2 newforms correspond to finite sets of rational points on $Y = \bigcup_{\mathfrak{a} \in \mathrm{Cl}^+(K)} Y(\mathcal{O}_K \oplus \mathfrak{a})$. Thus one can attempt estimate the number of quadratic twist classes by estimating counts of rational points on Y . A priori, it is not clear how different orderings of classes of newforms (e.g., by minimal level) will correlate with different orderings of sets of rational points (e.g., by minimal height, for some choice of height function), and we will speculate more on this for $d = 2$ anon.

Note that typically (each component of) Y will be of general type, and one might expect that it has finitely many (and often no) rational points. Hence, for a given d , to estimate counts of degree d newforms, it should in principle suffice to consider finitely many Y . Moreover, this philosophy suggests that some totally real degree d rationality fields will be more common than others, roughly according to whether the moduli spaces Y have many or few rational points. Of course this is not the only consideration, due to various complications of the correspondence between newforms and rational points mentioned above.

This philosophy is in line with Coleman's conjecture (e.g., see [5]), which predicts there are only finitely many isomorphism classes of endomorphism algebras for d -dimensional abelian varieties over \mathbb{Q} . Hence Coleman's conjecture implies that, for a fixed d , only finitely many degree d rationality fields K_f occur as f varies over weight 2 newforms.

3.2. Rational points on Hilbert modular surfaces. Now we estimate point counts on certain Hilbert modular surfaces, and pursue the ideas of the previous section for $d = 2$.

Let $D > 0$ be a fundamental discriminant and \mathcal{O}_D be the ring of integers of $\mathbb{Q}(\sqrt{D})$. Let $Y_-(D)$ be the Hilbert modular surface constructed from the quotient $\mathfrak{H}^2/\mathrm{SL}(\mathcal{O}_D \oplus \sqrt{D}\mathcal{O}_D)$. This parametrizes principally polarized abelian surfaces with RM by \mathcal{O}_D (together with a polarization structure). See [38], [15] for details. For brevity, we will write RM D for RM by \mathcal{O}_D .

For our heuristic point counts, we will use explicit models for Hilbert modular surfaces. For $D < 100$, Elkies and Kumar [14] computed models for $Y_-(D)$. By work of Hirzebruch and Zagier [19], $Y_-(D)$ is rational (i.e., birational to \mathbb{P}^2) if and only if $D \in \{5, 8, 12, 13, 17\}$.

We expect that 100% of degree 2 weight 2 newforms correspond to rational points on Hilbert modular surfaces with the most rational points, i.e., the rational surfaces. While the Hilbert modular surfaces parametrizing non-principally polarizable surfaces with RM D are rational over \mathbb{C} for $D = 12, 21, 24, 28, 33, 60$ (see [38, Theorem VII.3.3]), we at least expect that the 5 rational $Y_-(D)$'s should account for a positive proportion of degree 2 weight 2 newforms, and this is supported by data.

To be more precise, the polarization classes of abelian surfaces with RM D are in bijection with the narrow ideal classes $\text{Cl}^+(\mathbb{Q}(\sqrt{D}))$. In particular, if $D = 5, 8, 13, 17$, the abelian surface is automatically principally polarizable. For $D = 12$, the narrow class number is 2, and the moduli spaces for each polarization type are rational (at least over \mathbb{C}), so it is not clear whether a positive proportion of abelian surfaces with RM 12 should be principally polarizable. We cannot yet analyze counts for the non-principal polarization types as we do not know models for the corresponding moduli spaces together with appropriate invariants. However, our data suggest that, at least for prime level, most degree 2 weight 2 newforms have rationality field $\mathbb{Q}(\sqrt{D})$ with $D = 5$ or 8, and thus correspond to points on $Y_-(5)$ and $Y_-(8)$.

For the remainder of the section, assume $D \in \{5, 8, 12, 13, 17\}$. Then $Y_-(D)$ is birational to $\mathbb{P}_{m,n}^2$. Let \mathcal{A}_2 be the moduli space for principally polarized abelian surfaces. Forgetting the RM action yields a map $Y_-(D) \rightarrow \mathcal{A}_2$.

Let \mathcal{M}_2 be the moduli space of genus 2 curves. To a genus 2 curve $C : y^2 = h(x)$, one associates Igusa–Clebsch invariants $I_{2j}(C)$ for $j = 1, 2, 3, 5$. Here $I_{2j}(C)$ can be regarded as a degree $2j$ -polynomial in the coefficients of $h(x)$, and $I_{10}(C)$ is the discriminant of h . One can realize \mathcal{A}_2 as weighted projective space $\mathbb{P}_{1,2,3,5}^3$ with coordinates $(I_2 : I_4 : I_6 : I_{10})$. The Torelli map $\mathcal{M}_2 \rightarrow \mathcal{A}_2$ sends the moduli of C to $(I_2(C) : I_4(C) : I_6(C) : I_{10}(C))$, and the image is the complement of the hyperplane $I_{10} = 0$. We note that Igusa–Clebsch invariants are only isomorphism invariants of C up to weighted projective scaling.

Elkies and Kumar [14] gave a birational model for $Y_-(D)$. In particular, for generic affine coordinates $(m, n) \in \mathbb{A}^2$, one has an associated point on $Y_-(D)$ and thus weighted projective coordinates $(I_2(m, n) : I_4(m, n) : I_6(m, n) : I_{10}(m, n)) \in \mathcal{A}_2$, where the $I_{2j}(m, n)$'s are explicit rational functions in m, n .

Now we will attempt to estimate the number of rational points (m, n) with bounded Igusa–Clebsch invariants. First we want to scale Igusa–Clebsch invariants (in $\mathbb{P}_{1,2,3,5}^3$) to be integral, as will be the case for the $I_{2j}(C)$'s given a curve C over \mathbb{Z} . Let us write $(m, n) = (a/c, b/c)$ for $a, b, c \in \mathbb{Z}$ with $\gcd(a, b, c) = 1$. Regarding a, b, c as variables, we scale the $I_{2j}(m, n)$'s to get polynomials $I_{2j}(a, b, c) \in \mathbb{Z}[a, b, c]$'s which are minimal integral over $\mathbb{Z}[a, b, c]$. That is, we scale out denominators, and also any factors of the numerators π so that $\pi^j \mid I_{2j}(a, b, c)$ for all $j \in \{1, 2, 3, 5\}$ implies π is a unit in $\mathbb{Z}[a, b, c]$. The resulting $I_{10}(a, b, c)$'s (which are uniquely determined up to ± 1) are given in [Table 3.1](#).

D	$I_{10}(a, b, c)$
5	$8(a^5 - 10a^3b^2 + 25ab^4 + 5a^4c - 50a^2b^2c + 125b^4c - 5a^3c^2 + 25ab^2c^2 - 45a^2c^3 + 225b^2c^3 + 108c^5)^2$
8	$8c^3(a - c)^3(a + c)^6(-16a^2b^2 + 32b^4 + a^3c - 56ab^2c + 9a^2c^2 - 72b^2c^2 + 27ac^3 + 27c^4)^2$
12	$(a + c)^3(a - c)^9(-27a^2 + b^2 + 27c^2)^2(a^2b + 9a^2c - 8c^3)^3$
13	$2^3 \cdot 3^{11} \cdot (-267a^3 + 72a^2b - ab^2 - 3552a^2c + 1440abc - 128b^2c + 768ac^2)^2 \cdot (-12a^3 + 3a^2c + b^2c)^4(-a^3 - 150a^2c + 6abc - 264ac^2 + 120bc^2 + 64c^3)^4$
17	$2^{15} \cdot 3^{11} \cdot (-132a + b + 3c)^3(-256a^3 - 1200a^2c + 18abc - 6006ac^2 + 99bc^2 + 41c^3)^5 \cdot (456a^2 + ab + 723ac - 8bc + 24c^2)^3(4608a^3 - 1728a^2c + b^2c + 216ac^2 - 9c^3)^2$

Table 3.1. I_{10} polynomials for $Y_-(D)$

Specializing a, b, c to integers, we denote by $I_{2j}^{\min}(a, b, c) \in \mathbb{Z}$ scalings which are minimal integral over \mathbb{Z} . Note that $I_{2j}^{\min}(a, b, c) \mid I_{2j}(a, b, c)$ but they are often not equal. E.g., when

$D = 5$, then the invariants $I_{2j}(1, 3, 2)$'s are $(-2^4 \cdot 5^3, 2^8 \cdot 5^4, -2^{15} \cdot 5 \cdot 599, 2^{21} \cdot 3^8)$, whereas the \mathbb{Z} -minimal invariants $I_{2j}^{\min}(a, b, c)$ are obtained by scaling out a factor of 2^4 , i.e., they are $(-5^3, 5^4, -2^3 \cdot 5 \cdot 599, 2 \cdot 3^8)$.

First we want to estimate, in terms of a real parameter T , the growth of the cardinality of

$$Z_D(T) := \{(a, b, c) \in \mathbb{Z}^3 - U_D : \gcd(a, b, c) = 1 \text{ and } |I_{2j}^{\min}(a, b, c)| < T^{2j} \text{ for } j \in 1, 2, 3, 5\},$$

where U_D consists of (a, b, c) such that the map $(a/c, b/c) \rightarrow \mathcal{A}_2$ is either undefined (e.g., $c = 0$) or is not finite-to-one (e.g., for $D = 8$, all points with $m = a/c = -1$ map to $(1 : 0 : 0 : 0) \in \mathcal{A}_2$). Really our interest is just in bounding I_{10}^{\min} , but we impose bounds on the other I_{2j}^{\min} 's to guarantee finiteness of $Z_D(T)$.

Precise estimates are difficult, so we make two simplifications which are sufficient to get lower bounds: (1) We impose the stronger bound $|I_{2j}^{\min}(a, b, c)| \leq |I_{2j}(a, b, c)| < T^{2j}$. (2) We will suppose each monomial in $I_{2j}(a, b, c)$ is bounded by T^{2j} . Note that (1) and (2) can respectively be thought of as non-archimedean and archimedean simplifications to monomials.

Proposition 3.2. *We have $\#Z_D(T) \gg T^{r_D}$ as $T \rightarrow \infty$, where respectively $r_D = 3, \frac{3}{2}, 2, 1, 1$ for $D = 5, 8, 12, 13, 17$.*

Proof. Each $I_{2j}(a, b, c)$ is a homogeneous polynomial in a, b, c , say of degree d_j . Taking a, b, c independently up to size T^{2j/d_j} shows there are $\gg T^{6j/d_j}$ tuples (a, b, c) with $|I_{2j}(a, b, c)| < T^{2j}$. Moreover, one checks the ratio j/d_j is independent of the choice of j . Since the conditions $\gcd(a, b, c) = 1$ and $(a, b, c) \notin U_D$ are satisfied for a positive proportion of (a, b, c) , we get the asymptotic lower bound $\#Z_D(T) \gg T^{30/d}$, where $d = \deg I_{10}(a, b, c)$. We respectively have $d = 10, 20, 25, 30, 30$ for $D = 5, 8, 12, 13, 17$, which gives the asserted lower bounds for $D = 5, 8, 13, 17$.

For $D = 12$, one can get better lower bounds using a $\mathbb{P}^1 \times \mathbb{P}^1$ parametrization for (m, n) . Namely, write $(m, n) = (r/s, t/u)$ for $r, s, t, u \in \mathbb{Z}$. Let $I_{2j}(r, s, t, u)$ be the minimal invariants over $\mathbb{Z}[r, s, t, u]$. These have degrees 6, 12, 18, 30 for $j = 1, 2, 3, 5$. Using same argument as above gives a lower bound of $\#Z_{12}(T) \gg T^{4/3}$. However, if we regard $I_{2j}(r, s, t, u)$ as polynomials only in t and u , the respective degrees are 2, 4, 6, 10 for $j = 1, 2, 3, 5$. Thus by taking r, s uniformly bounded and $|t|, |u| \ll T$ yields $\#Z_{12}(T) \gg T^2$ as claimed. \square

The lower bounds in the proposition are the optimal ones we could find using the \mathbb{P}^2 or $\mathbb{P}^1 \times \mathbb{P}^1$ parametrizations for (m, n) by allowing either each of a, b, c or r, s, t, u to vary independently up to some power of T (not necessarily the same power for each variable). We remark that the optimal exponents for lower bounds using the $\mathbb{P}^1 \times \mathbb{P}^1$ parametrization for $D = 5, 8, 13, 17$ are 2, 4/3, 1, 1, respectively. To get these exponents, for $D = 5$ one can take each of $|r|, |s|, |t|, |u| \ll T$. For $D = 8$, one takes $|r|, |s| \ll T^{2/3}$ and $|t|, |u| \ll 1$. For both $D = 13$ and $D = 17$, one takes $|r|, |s| \ll 1$ and $|t|, |u| \ll T^{1/2}$.

Question 3.3. *For $D \in \{5, 8, 12, 13, 17\}$, is $\#Z_D(T) \ll T^{r_D + \varepsilon}$ for any $\varepsilon > 0$?*

It is not clear if the simplifications to monomials affect the exponents in our estimates, but if the I_{2j} polynomials are sufficiently general type, one might expect they only account for a multiplicative factor of size $O(1 + T^\varepsilon)$. Note that in the Brumer–McGuinness heuristics, it is believed the analogous archimedean simplification (2) only affects counts by an $O(1)$ factor.

A more serious reason to doubt the exponents in these lower bounds are optimal is that there may exist (i) other rational parametrizations of $Y_-(D)$ where the I_{2j} degrees are smaller, or (ii) special curves on $Y_-(D)$ which intersect Z_D in an especially large number of rational

points. Indeed, the proof makes clear that different parametrizations may yield different counts for a given surface.

Now we explain how these estimates for $\#Z_D(T)$ are related to counting quadratic twist classes of weight 2 newforms f with rationality field $\mathbb{Q}(\sqrt{D})$. As explained above, a positive proportion of f (at least if $D \neq 12$) should correspond to rational points on $Y_-(D)$. Conversely, a rational point on $Y_-(D)$ may not come from a simple abelian surface with RM defined over \mathbb{Q} —one needs that the Mestre conic has a point, the RM is defined over \mathbb{Q} , and the Jacobian is nonsplit. However, we expect that these obstructions will only contribute logarithmic factors to asymptotics. (See [9] for details about when the Mestre obstruction vanishes and when the RM is defined over \mathbb{Q} .)

Consider a point $(a, b, c) \in Z_D(T)$ which corresponds to a $\bar{\mathbb{Q}}$ -isomorphism class of simple abelian surfaces A/\mathbb{Q} . Generically this $\bar{\mathbb{Q}}$ -isomorphism class should be the family of Jacobians of quadratic twists of a genus 2 curve C/\mathbb{Q} with RM D . Suppose this, and assume C is a minimal quadratic twist of conductor N_C . Then (a, b, c) corresponds to the quadratic twist class of some minimal newform f of level N_f where $N_C = N_f^2$. One can write down a minimal integral model for C , and the polynomially-defined Igusa–Clebsch invariants $I_{2j}(C)$ are necessarily divisible by $I_{2j}^{\min}(a, b, c)$. Thus $I_{10}^{\min}(a, b, c)$ divides the minimal discriminant $\Delta_C = 2^{-12}I_{10}(C)$ of C . One also knows that the conductor $N_C \mid \Delta_C$.

Now we would like to understand how N_C relates to $I_{10}^{\min}(a, b, c)$ or $I_{10}(a, b, c)$. There are no general upper or lower bounds, and in fact there are competing issues in opposite directions. One is that Δ_C may be much larger than either $I_{10}^{\min}(a, b, c)$ or $I_{10}(a, b, c)$, and numerically this is quite typical. E.g., if $p^m \parallel I_{10}^{\min}(a, b, c)$ it often happens that $p^{m+10} \mid \Delta_C$ (see [24] for local results). On the other hand, the prime powers occurring in N_C are often smaller than those in Δ_C . E.g., if $p^3 \parallel \Delta_C$ then necessarily $p^2 \parallel N_C$ or $p \nmid N_C$. Some preliminary investigations suggest that the latter issue has more impact, and asymptotic counts of curves by I_{10}^{\min} may essentially be lower bounds (up to logarithmic factors) of counts by conductor. This leads us to ask:

Question 3.4. *Let $C_2^{\text{tw}}(D; X)$ be the number of quadratic twist classes of non-CM weight 2 newforms of minimal level $N < X$ with rationality field $\mathbb{Q}(\sqrt{D})$. Is $C_2^{\text{tw}}(D; X) \gg X^{\alpha-\varepsilon}$ for α such that $\#Z_D(T) \gg T^{\alpha/5}$? Note that one can take $\alpha = \frac{3}{5}, \frac{3}{10}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5}$ for $D = 5, 8, 12, 13, 17$.*

Observe that all of these exponents are less than the exponent of $2/3$ from the $d = 2$ case of [Conjecture 1.1](#). We will compare these speculative lower bounds with our prime level data below.

Even if these lower bounds hold, there are several reasons why they may not be sharp. For one, there is the issue of [Question 3.3](#). Perhaps most serious is the issue of how prime powers in N_C relate to prime powers in Δ_C mentioned above.

Another potential issue comes from the way we defined $Z_D(T)$: for the comparison with conductors, we are only interested in bounds on I_{10}^{\min} , and there may be many points with I_{10}^{\min} small relative to $I_2^{\min}, I_4^{\min}, I_6^{\min}$. In particular, for $D = 12$, if one views $I_{2j}(a, b, c)$ as a polynomial in b with $a = 0$ and c fixed, then the degrees are 2, 2, 4, 4, so one gets at least $T^{5/2}$ points with $I_{10}^{\min} \ll T^{10}$, which is a better than the lower bound $\gg T^2$ in [Proposition 3.2](#). These rational points correspond to the curve $m = 0$ on $Y_-(12)$, and numerical investigations suggests this is a Shimura curve parametrizing abelian surfaces with geometric endomorphism algebra the quaternion algebra of discriminant 6. Consequently, one might be able to take $\alpha = \frac{1}{2}$ in [Question 3.4](#) when $D = 12$. However, it is not clear that there is a family of genus 2 curves with RM 12 over \mathbb{Q} that would achieve $\alpha = \frac{1}{2}$.

3.3. Lower bounds for quadratic fields. There are several known families of genus 2 curves with RM 5 and RM 8. These can be used to give lower bounds on counting such curves with bounded discriminant, and therefore conductor. For a genus 2 curve C/\mathbb{Q} , let Δ_C denote the minimal integral discriminant.

Proposition 3.5. *The number of $\bar{\mathbb{Q}}$ -isomorphism classes of genus 2 curves C/\mathbb{Q} with RM 5 (resp. RM 8) with $|\Delta_C| < X$ is $\gg X^{1/6}$ (resp. $\gg X^{1/7}$).*

Proof. First consider RM 5. Brumer exhibited a 3-parameter family of curves $C_{b,c,d}$ with RM 5 over \mathbb{Q} (see [6] for an announcement and [18] for a proof), however it is not clear when two such curves are isomorphic, either over \mathbb{Q} or $\bar{\mathbb{Q}}$. We consider the 1-parameter subfamily C_d with $b = c = 0$, which is given by

$$C_d : y^2 + (x^3 + x + 1)y = -dx^3 + x^2 + x.$$

This defines a genus 2 curve with RM 5 for all $d \in \mathbb{Z}$ (I_{10} is never 0 for $d \in \mathbb{Z}$), and the discriminant of this model, $(27d^3 - 81d^2 - 34d - 103)^2$, is degree 6 in d . Thus to complete the RM 5 case of the proposition, it suffices to show that the number of $C_{d'}$ isomorphic to a given C_d over $\bar{\mathbb{Q}}$ is finite and uniformly bounded.

Let I_{2j} (resp. I'_{2j}) denote the polynomial Igusa–Clebsch invariants for C_d (resp. $C_{d'}$) for $j = 1, 2, 3, 5$. Then I_4/I_2^2 and is an absolute invariant for C_d . Thus if C_d and $C_{d'}$ are $\bar{\mathbb{Q}}$ -isomorphic, one must have $F := (I'_2)^2 I_4 - I_2^2 I'_4 = 0$. Now $F = 0$ defines a union of 3 curves in the (d, d') -plane, none of which are of the form $d = d_0$. So the number of $C_{d'}$ which are $\bar{\mathbb{Q}}$ -isomorphic to a fixed C_d is bounded by $\deg F$.

Now consider RM 8. Here we use Mestre's 2-parameter family $C'_{a,b}$ of genus 2 RM 8 curves over \mathbb{Q} from [30]. Consider the subfamily $C'_b = C'_{2,b}$, which is given by

$$C'_b : y^2 = 7500x^5 + (-75b + 3400)x^4 + (-34b + 2283)x^3 + (-3b + 1111)x^2 + 177x + 9.$$

This is a genus 2 curve with RM 8 for $b \in \mathbb{Z} - \{-88, 112\}$, and the discriminant has degree 7 in b . One can complete the argument just as in the RM 5 case. \square

Corollary 3.6. *The number of quadratic twist classes of weight 2 newforms with rationality field $\mathbb{Q}(\sqrt{5})$ (resp. $\mathbb{Q}(\sqrt{2})$) and minimal level $N < X$ is $\gg X^{1/3}$ (resp. $\gg X^{2/7}$).*

Proof. If C is a genus 2 curve with RM D over \mathbb{Q} with nonsplit Jacobian, then modularity tells us that C corresponds to a weight 2 newform f of level $N \leq \sqrt{|D_C|}$ and rationality field $\mathbb{Q}(\sqrt{D})$, so it suffices to show that a positive proportion of the curves in the proofs have nonsplit Jacobian. For the RM 5 C'_d family above, one can check that the above model has discriminant coprime to 5 when $d \equiv 1 \pmod{5}$. Moreover computing the L -polynomial shows the mod 5 Jacobian is nonsplit, whence the Jacobian of C_d over \mathbb{Q} is nonsplit. Similarly, for the RM 8 family, the curve C'_b has nonsplit Jacobian when $b \equiv 1 \pmod{7}$. \square

Note that these lower bounds are significantly smaller than those in [Question 3.4](#).

Remark 3.7.

- (1) Under the Bateman–Horn conjecture [1], $27d^3 - 81d^2 - 34d - 103$ is prime for $\gg X/\log X$ integers $d \equiv 1 \pmod{5}$. For such d , C_d has prime-squared discriminant and the associated newform f has prime conductor. Consequently, subject to this conjecture, the above argument shows there are $\gg X^{1/3}/\log X$ weight 2 newforms f with rationality field $\mathbb{Q}(\sqrt{5})$ and prime level $< X$. The analogous argument does not work for RM 8 as the discriminant of C'_b splits into linear factors over \mathbb{Z} .

- (2) Elkies [13] recently gave similar lower bounds for genus 2 curves with RM 5 which satisfy an Eisenstein congruence, but with the exponent $\frac{1}{3}$ replaced by $\frac{1}{4}$.
- (3) One should similarly be able to give explicit lower bounds for $D = 12, 13, 17$ (and some other D). Namely, [14] gives infinite families of RM D genus 2 curves for various D , though without explicit models. One can use Mestre's algorithm to construct models, and then follow the above argument. However, this typically yields families with discriminants of very large degree, and so one would need to do more work to get decent lower bounds.

3.4. Remarks for higher degree. Much less is known about Hilbert modular n -folds for $n > 2$ than for $n = 2$. Grundman and Lippincott ([16, 17]) have done work towards classifying such spaces by arithmetic genus for $n = 3, 4$. Then Borisov and Gunnells [4] studied the geometry of a Hilbert modular 3-fold attached to $\mathbb{Q}(\zeta_7^+)$. However, to our knowledge, not much is known about the explicit geometry of higher-dimensional Hilbert modular varieties beyond these works and their references.

One case where we do know of geometric constructions leading to weight 2 modular forms of degree $d > 2$ is the following. In [30], Mestre constructs (among other things) families of genus $d = \frac{p-1}{2}$ hyperelliptic curves with potential RM by $\mathbb{Z}[\zeta_p^+]$. When $p = 5, 7$, the RM is actually defined over \mathbb{Q} , and generically the curves should correspond to degree 2 and 3 modular forms with rationality fields $\mathbb{Q}(\sqrt{5})$ and $\mathbb{Q}(\zeta_7^+)$.

Mestre's constructions have been extended by various authors. For instance, [21] and [20] construct genus 3 curves C with RM by an order in $\mathbb{Q}(\zeta_7^+)$ such that the RM is generically defined over the base field. This at least suggests there may be infinitely many weight 2 newforms of squarefree level with rationality field $\mathbb{Q}(\zeta_7^+)$.

4. DATA

The LMFDB [25] currently contains all weight 2 newforms of level $N \leq 10^4$ [3]. However, this range is not nearly sufficient to study asymptotic behavior of the distribution of newform degrees in squarefree level. For instance, one is still forced to have many small degree forms of level close to 10000 as the size of Atkin–Lehner spaces can still be small. For instance, for $N = 9870 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 47$, there are 183 newforms in $S_2(N)$ divided among 32 Atkin–Lehner eigenspaces, and each Atkin–Lehner eigenspace has dimension between 3 and 8.

One can mitigate this effect by restricting to levels with at most 2 or 3 prime factors, or ordering counts by dimensions of Atkin–Lehner eigenspaces rather than level. However, even with such considerations, the range is still not large enough to say much about asymptotic counts.

Instead, we analyze data we computed in prime level using the algorithms from [8]. Namely, we computed all weight 2 newforms of prime level less than $2 \cdot 10^6$ and degree at most 6, as well as the degrees of all newforms of prime level less than 10^6 . Our data is available on the first author's personal webpage, and is currently in the process of being added to the LMFDB.

4.1. Data for counts by degree. In Figure 4.1, we plot counts $\mathcal{C}'_d(X)$ of degree d Galois orbits in weight 2 and prime level at most X for $1 \leq d \leq 4$. The plot uses a log-log scale, and because the number of primes up to X is approximately $\text{li}(X)$, we do a least-squares fit of functions of the form $y = \log(\text{ali}(\exp(x)^b))$ to the data

$$\{(\log X, \log \mathcal{C}'_d(X)) : X < 2 \cdot 10^6, X \text{ prime}, \mathcal{C}'_d(X) \geq 1\}.$$

For degree 1, our best fit values $a \approx 0.97$ and $b \approx 0.832$ are in agreement with the elliptic curve database [2], which, for X up to $2 \cdot 10^9$, finds that $a \approx 0.97$ and $b \approx 0.833$. This agrees

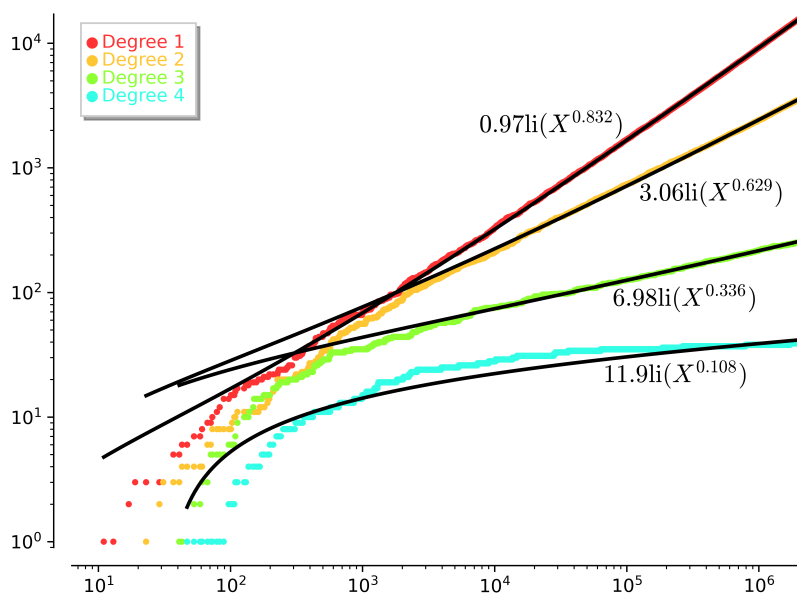


Figure 4.1. Number of forms with prime level less than X by degree, with least-squares fits to the log-log data.

very well with the Brumer–McGuinness–Watkins heuristic (1). For $d = 2, 3, 4$, the best fit exponents are approximately 0.629, 0.336, 0.108.

A least-squares fit of $y = ali(x^b)$ to the data $(X, C'_d(X))$ yields similar values; the exponents are 0.841, 0.622, 0.330, 0.107 for $d = 1, 2, 3, 4$, respectively.

When $1 \leq d \leq 4$, [Conjecture 1.1](#) is consistent with our prime level data. In [Figure 4.1](#), the growth rate $O(X^{1-d/6})$ appears to be an upper bound. In this range the best fit has a notably lower exponent for $d = 3, 4$, but there may exist logarithmic factors in the main asymptotics for $d \geq 2$. For example, for $d = 2$, the Mestre obstruction to rationality of genus 2 curves (see [Section 3.2](#)) may introduce a logarithmic factor in the denominator. Note that in this range, $\log X > X^{\frac{1}{6}}$, so it is difficult to distinguish between logarithmic factors and small powers of X .

[Table 4.2](#) shows the number of newform orbits of degree at most 6 with prime level between 1 and 10^4 , between 10^4 and 10^6 , and between 10^6 and $2 \cdot 10^6$. Note that the first data column counts prime level forms which were already in the LMFDB.

Degree	Level range			Total
	$1 - 10^4$	$10^4 - 10^6$	$10^6 - 2 \cdot 10^6$	
1	329	8843	6406	15578
2	212	2200	1096	3508
3	76	142	35	253
4	28	10	1	39
5	20	2		22
6	11	1		12

Table 4.2. Number of prime-level newform orbits by degree and level. Blank entries are 0.

Many of the forms of degree 5 and 6 in this dataset have very small levels. For instance, 12 of the 22 degree 5 forms are the largest-degree forms in their Atkin–Lehner eigenspaces,

and similarly for 7 of the 12 degree 6 forms. Only 5 of the degree 5 forms and 3 of the degree 6 forms have levels greater than 1000.

Given this paucity of data, we refrain from any quantitative analysis of these forms, but remark that the counts for $d = 5, 6$ appear to be consistent with [Conjecture 1.1](#).

[Table 4.3](#) gives the decomposition type, i.e., sizes of Galois orbits, of all prime levels up to 1 million which have two or more newform orbits of degree at least 7 in the same Atkin–Lehner eigenspace. Here $S_2^\pm(p)$ denotes the subspace of $S_2(p)$ with Atkin–Lehner eigenvalue ± 1 at p .

Level	$S_2^+(p)$	$S_2^-(p)$
607	5 + 7 + 7	31
911	9 + 14	53
1223	34	9 + 59
1249	7 + 37	59
4751	153	18 + 225

Table 4.3. Atkin–Lehner eigenspaces with multiple orbits of size at least 7

[Table 4.3](#) shows that there are only 6 newforms orbits of prime level less than 10^6 with degree 7 or more that are not the unique largest in their Atkin–Lehner eigenspaces (2 of which are tied for the largest), with degrees 7, 7, 7, 9, 9, and 18. This data for $d \geq 7$ appears to be consistent with [Conjecture 1.1](#).

Moreover, [Table 4.3](#) shows that, for each prime level between 4751 and 10^6 , the newforms in each Atkin–Lehner eigenspace consist of a single large Galois orbit together with orbits of size ≤ 6 . This prompts us to ask the following:

Question 4.4. *Is there a prime $p > 4751$ and a sign \pm such that $S_2^\pm(p)$ contains two or more newform orbits each of degree 7 or more?*

This question can be viewed as a uniform version of [Conjecture 1.1](#) in prime level. It seems plausible to us that in fact [Table 4.3](#) is a complete list of all “mid-sized forms” of prime level.

4.2. Data counts by quadratic field. In this section we investigate [Conjecture 1.2](#) and [Question 3.4](#) empirically. [Table 4.5](#) and [Figure 4.6](#) present the relevant contents of our dataset.

Disc	Level range			Total
	$1 - 10^4$	$10^4 - 10^6$	$10^6 - 2 \cdot 10^6$	
5	158	1900	986	3044
8	37	242	100	379
12	1	14	3	18
13	13	40	6	59
17		1		1

Table 4.5. Number of prime-level degree 2 newform orbits by discriminant and level, for discriminants D such that $Y_-(D)$ is rational. Blank entries are 0.

The growth rate of the counts plotted in [Figure 4.6](#) appears to be the largest for $D = 5$, and $D = 8$ appears to be the next largest. This is consistent with [Conjecture 1.2](#).

Let $\mathcal{C}'_{d,D}(X)$ denote the number of newform orbits of degree d , discriminant D , and prime level at most X . For each of $D = 5, 8, 12$, and 13 , we compute least-squares fits to the following four sets of data points:

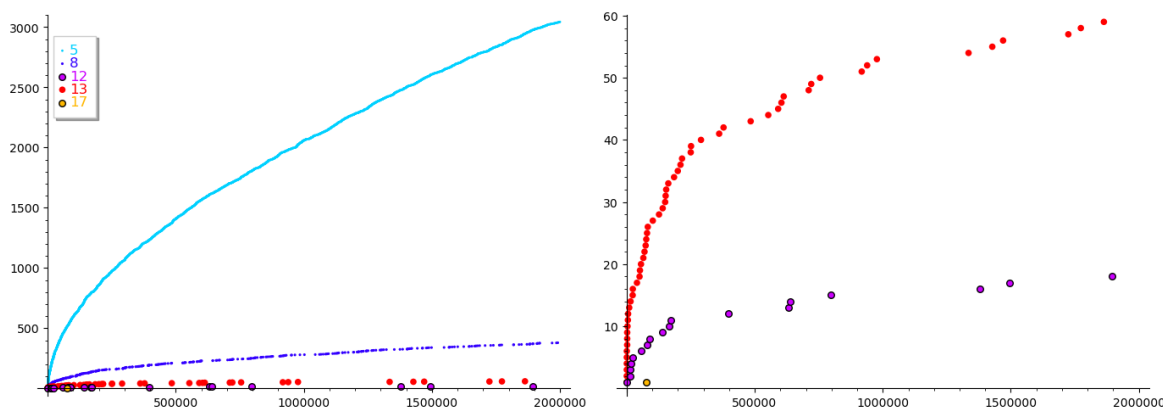


Figure 4.6. Counts of degree 2 forms by discriminant D for rational surfaces $Y_-(D)$. The plot on the right excludes discriminants 5 and 8.

- (1) $\left\{ (\log X, \log \mathcal{C}'_{d,D}(X)) : 1 < X < 2 \cdot 10^6, X \text{ prime}, \mathcal{C}'_{d,D}(X) \geq 1 \right\}$
- (2) $\left\{ (\log X, \log \mathcal{C}'_{d,D}(X)) : 10^4 < X < 2 \cdot 10^6, X \text{ prime}, \mathcal{C}'_{d,D}(X) \geq 1 \right\}$
- (3) $\left\{ (X, \mathcal{C}'_{d,D}(X)) : 1 < X < 2 \cdot 10^6, X \text{ prime}, \mathcal{C}'_{d,D}(X) \geq 1 \right\}$
- (4) $\left\{ (X, \mathcal{C}'_{d,D}(X)) : 10^4 < X < 2 \cdot 10^6, X \text{ prime}, \mathcal{C}'_{d,D}(X) \geq 1 \right\}$

In the first two cases we fit functions of the form $y = \log(\text{ali}(\exp(x)^b))$, and in the last two $y = \text{ali}(x^b)$. The best-fit exponents b we obtain vary depending on our choice of model and range of X values. We present these best-fit values of b in [Table 4.7](#).

Data	X range	Best-fit exponents			
		5	8	12	13
$(\log X, \log \mathcal{C}'_{d,D}(X))$	$1 - 2 \cdot 10^6$	0.73	0.61	0.39	0.42
$(\log X, \log \mathcal{C}'_{d,D}(X))$	$10^4 - 2 \cdot 10^6$	0.64	0.54	0.39	0.38
$(X, \mathcal{C}'_{d,D}(X))$	$1 - 2 \cdot 10^6$	0.67	0.56	0.37	0.37
$(X, \mathcal{C}'_{d,D}(X))$	$10^4 - 2 \cdot 10^6$	0.65	0.53	0.37	0.35

Table 4.7. Best-fit values of b when fitting functions of the form $y = \log(\text{ali}(\exp(x)^b))$ or $y = \text{ali}(x^b)$ as appropriate to data of counts of degree 2 newform orbits with prime level and prescribed discriminant

The best-fit exponents we obtain are all higher than the lower bounds proposed in [Question 3.4](#), in many cases substantially, except for discriminant 12, where the value is slightly lower than the 0.4 appearing in [Question 3.4](#). There is only one form of discriminant 17 and prime level less than $2 \cdot 10^6$, at level 75653.

The data presented in [Table 4.5](#), [Figure 4.6](#), and [Table 4.7](#), as well as the heuristics from [Question 3.4](#), support [Conjecture 1.2](#). Namely, the suggested lower bounds for counts by quadratic rationality field are largest for $\mathbb{Q}(\sqrt{5})$. Since the heuristics do not rely on a restriction to prime level, one is led to ask:

Question 4.8. *Do 100% of quadratic twist classes (ordered by minimal level) of weight 2 degree 2 non-CM newforms have rationality field $\mathbb{Q}(\sqrt{5})$?*

There is an arithmetic reason to expect a relative scarcity of certain quadratic fields in prime level compared to the lower bounds for arbitrary levels suggested in [Question 3.4](#). Namely, if C is a genus 2 curve with RM D , then we typically expect odd primes dividing I_{10}^{\min} to divide the conductor N_C . Based on the factorizations of I_{10} 's in [Table 3.1](#), we expect I_{10}^{\min} to be a 2-power times a prime power very infrequently for $D = 12, 13, 17$. Indeed, the LMFDB [\[25\]](#) lists 5485, 3948 2189, 1230, and 1643 forms in all levels $N \leq 10000$ for $D = 5, 8, 12, 13$, and 17, respectively. Restricting to squarefree level, these numbers are 1820, 1124, 445, 319, and 461.

4.3. Data counts for cubic fields. While we have not attempted to carry out the approach outlined in [Section 3.1](#) to estimate counts of degree 3 forms with a given cubic rationality field K , the data, though more limited, behaves similarly as in the degree 2 case. [Table 4.9](#) and [Figure 4.10](#) present counts of all prime level degree 3 newform orbits by their Hecke field discriminant. These discriminants are sufficient to specify the Hecke field, in the sense that if f and g are degree 3 forms of prime level less than $2 \cdot 10^6$ and $\text{Disc}(K_f) = \text{Disc}(K_g)$, then $K_f = K_g$.

Disc	Level range			Total
	$1 - 10^4$	$10^4 - 10^6$	$10^6 - 2 \cdot 10^6$	
49	34	90	30	154
81	3	13		16
148	12	6		18
169	2	6	3	11
229	8	20	1	29
257	9	6	1	16
321	2	1		3
404	2			2
469	1			1
473	2			2
621	1			1

Table 4.9. Number of prime-level degree 3 newform orbits by discriminant and level. Blank entries are 0.

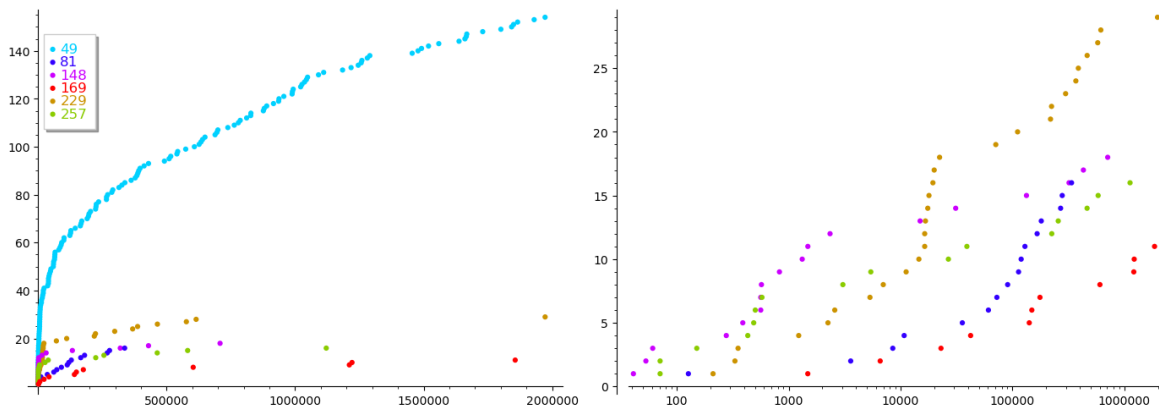


Figure 4.10. Counts of degree 3 forms by discriminant. The plot on the right excludes discriminant 49 and is on a log scale. Not shown are the 9 forms with discriminant 321, 404, 469, 473, or 621.

Like in [Section 4.2](#), we fit functions of the form either $y = \log(\text{ali}(\exp(x)^b))$ or $y = \text{ali}(x^b)$ as appropriate to data points of the form either $(\log X, \log \mathcal{C}'_{d,D}(X))$ or $(X, \mathcal{C}'_{d,D}(X))$, for prime X between either 1 and $2 \cdot 10^6$ or 10^4 and $2 \cdot 10^6$. The best-fit exponents we obtain in each of these four cases, for $D = 49, 81, 148, 169, 229$, and 257 , are shown in [Table 4.11](#).

Data	X range	Best-fit exponents by D					
		49	81	148	169	229	257
$(\log X, \log \mathcal{C}'_{d,D}(X))$	$1 - 2 \cdot 10^6$	0.42	0.34	0.12	0.39	0.24	0.18
$(\log X, \log \mathcal{C}'_{d,D}(X))$	$10^4 - 2 \cdot 10^6$	0.42	0.32	0.11	0.38	0.20	0.19
$(X, \mathcal{C}'_{d,D}(X))$	$1 - 2 \cdot 10^6$	0.43	0.23	0.11	0.37	0.19	0.19
$(X, \mathcal{C}'_{d,D}(X))$	$10^4 - 2 \cdot 10^6$	0.43	0.22	0.11	0.37	0.18	0.19

Table 4.11. Best-fit values of b when fitting functions of the form $y = \log(\text{ali}(\exp(x)^b))$ or $y = \text{ali}(x^b)$ as appropriate to data of counts of degree 3 newform orbits with prime level and prescribed discriminant

Analogous to the degree 2 case, it is natural to ask:

Question 4.12. *Among squarefree levels $N \rightarrow \infty$, do 100% of degree 3 newforms in $S_2(N)$ have rationality field $\mathbb{Q}(\zeta_7)^+$?*

REFERENCES

- [1] Paul T. Bateman and Roger A. Horn. A heuristic asymptotic formula concerning the distribution of prime numbers. *Math. Comp.*, 16:363–367, 1962. [1](#)
- [2] Michael A. Bennett, Adela Gherga, and Andrew Rechnitzer. Computing elliptic curves over \mathbb{Q} . *Math. Comp.*, 88(317):1341–1390, 2019. [1](#), [2](#), [4.1](#)
- [3] Alex J. Best, Jonathan Bober, Andrew R. Booker, Edgar Costa, John E. Cremona, Maarten Derickx, Min Lee, David Lowry-Duda, David Roe, Andrew V. Sutherland, and John Voight. Computing classical modular forms. In *Arithmetic geometry, number theory, and computation*, Simons Symp., pages 131–213. Springer, Cham, [2021] ©2021. [4](#)
- [4] Lev A. Borisov and Paul E. Gunnells. On Hilbert modular threefolds of discriminant 49. *Selecta Math. (N.S.)*, 19(4):923–947, 2013. [3.4](#)
- [5] Nils Bruin, E. Victor Flynn, Josep González, and Victor Rotger. On finiteness conjectures for endomorphism algebras of abelian surfaces. *Math. Proc. Cambridge Philos. Soc.*, 141(3):383–408, 2006. [3.1](#)
- [6] Armand Brumer. The rank of $J_0(N)$. Number 228, pages 3, 41–68. 1995. Columbia University Number Theory Seminar (New York, 1992). [1.3](#), [3.3](#)
- [7] Armand Brumer and Oisín McGuinness. The behavior of the Mordell-Weil group of elliptic curves. *Bull. Amer. Math. Soc. (N.S.)*, 23(2):375–382, 1990. [1](#), [2](#)
- [8] Alex Cowan. Computing newforms using supersingular isogeny graphs. *Res. Number Theory*, 8(4):Paper No. 96, 2022. [1](#), [4](#)
- [9] Alex Cowan and Kimball Martin. Moduli for rational genus 2 curves with real multiplication for discriminant 5. *J. Théor. Nombres Bordeaux*, to appear. [3.2](#)
- [10] J. Cremona. ecdatal. <https://github.com/JohnCremona/ecdata>, January 2023. [2](#)
- [11] Luis Victor Dieulefait, Ariel Pacetti, and Panagiotis Tsaknias. On the number of Galois orbits of newforms. *J. Eur. Math. Soc. (JEMS)*, 23(8):2833–2860, 2021. [2.1](#)
- [12] W. Duke and E. Kowalski. A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations. *Invent. Math.*, 139(1):1–39, 2000. With an appendix by Dinakar Ramakrishnan. [1](#), [2](#)
- [13] Noam Elkies. Families of genus-2 curves with 5-torsion, Preprint. [2](#)
- [14] Noam Elkies and Abhinav Kumar. K3 surfaces and equations for Hilbert modular surfaces. *Algebra Number Theory*, 8(10):2297–2411, 2014. [3.2](#), [3](#)
- [15] Eyal Z. Goren. *Lectures on Hilbert modular varieties and modular forms*, volume 14 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 2002. With the assistance of Marc-Hubert Nicole. [3.1](#), [3.2](#)

- [16] H. G. Grundman and L. E. Lippincott. Hilbert modular threefolds of arithmetic genus one. *J. Number Theory*, 95(1):72–76, 2002. [3.4](#)
- [17] H. G. Grundman and L. E. Lippincott. Computing the arithmetic genus of Hilbert modular fourfolds. *Math. Comp.*, 75(255):1553–1560, 2006. [3.4](#)
- [18] Ki-ichiro Hashimoto. On Brumer’s family of RM-curves of genus two. *Tohoku Math. J. (2)*, 52(4):475–488, 2000. [3.3](#)
- [19] F. Hirzebruch and D. Zagier. Classification of Hilbert modular surfaces. In *Complex analysis and algebraic geometry*, pages 43–77. Iwanami Shoten, Tokyo, 1977. [3.2](#)
- [20] J. William Hoffman, Zhibin Liang, Yukiko Sakai, and Haohao Wang. Genus 3 curves whose Jacobians have endomorphisms by $\mathbb{Q}(\zeta_7 + \bar{\zeta}_7)$. *J. Symbolic Comput.*, 74:561–577, 2016. [3.4](#)
- [21] J. William Hoffman and Haohao Wang. 7-gons and genus three hyperelliptic curves. *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM*, 107(1):35–52, 2013. [3.4](#)
- [22] Koopa Tak-Lun Koo, William Stein, and Gabor Wiese. On the generation of the coefficient field of a newform by a single Hecke eigenvalue. *J. Théor. Nombres Bordeaux*, 20(2):373–384, 2008. [2.1](#)
- [23] Michael Lipnowski and George J. Schaeffer. Detecting large simple rational Hecke modules for $\Gamma_0(N)$ via congruences. *Int. Math. Res. Not. IMRN*, (19):6149–6168, 2020. [1](#)
- [24] Qing Liu. Conducteur et discriminant minimal de courbes de genre 2. *Compositio Math.*, 94(1):51–79, 1994. [3.2](#)
- [25] The LMFDB Collaboration. The L-functions and modular forms database. <http://www.lmfdb.org>, 2023. [Online; accessed 3 January 2023]. [4](#), [4.2](#)
- [26] Kimball Martin. Local conductors bounds for modular abelian varieties. *Acta Arith.*, to appear. [1.3](#), [3.1](#)
- [27] Kimball Martin. Refined dimensions of cusp forms, and equidistribution and bias of signs. *J. Number Theory*, 188:1–17, 2018. [2.1](#)
- [28] Kimball Martin. An on-average Maeda-type conjecture in the level aspect. *Proc. Amer. Math. Soc.*, 149(4):1373–1386, 2021. [1](#), [1](#), [2](#), [2.1](#), [2.1](#)
- [29] Kimball Martin and Jordan Wiebe. Zeroes of quaternionic modular forms and central L -values. *J. Number Theory*, 217:460–494, 2020. [2.1](#)
- [30] J.-F. Mestre. Familles de courbes hyperelliptiques à multiplications réelles. In *Arithmetic algebraic geometry (Texel, 1989)*, volume 89 of *Progr. Math.*, pages 193–208. Birkhäuser Boston, Boston, MA, 1991. [3.3](#), [3.4](#)
- [31] M. Ram Murty and Kaneenika Sinha. Effective equidistribution of eigenvalues of Hecke operators. *J. Number Theory*, 129(3):681–714, 2009. [2](#)
- [32] V. Kumar Murty. Frobenius distributions and Galois representations. In *Automorphic forms, automorphic representations, and arithmetic (Fort Worth, TX, 1996)*, volume 66 of *Proc. Sympos. Pure Math.*, pages 193–211. Amer. Math. Soc., Providence, RI, 1999. [2.1](#)
- [33] David P. Roberts. Newforms with rational coefficients. *Ramanujan J.*, 46(3):835–862, 2018. [1](#), [2.1](#), [2.1](#), [2](#)
- [34] Peter Sarnak and Nina Zubrilina. Convergence to the Plancherel measure of Hecke eigenvalues, 2022. arXiv 2201.03523. [2](#)
- [35] Jean-Pierre Serre. Répartition asymptotique des valeurs propres de l’opérateur de Hecke T_p . *J. Amer. Math. Soc.*, 10(1):75–102, 1997. [1](#), [2](#)
- [36] Ananth N. Shankar, Arul Shankar, and Xiaoheng Wang. Large families of elliptic curves ordered by conductor. *Compos. Math.*, 157(7):1538–1583, 2021. [1](#), [2](#)
- [37] William A. Stein and Mark Watkins. A database of elliptic curves—first report. In *Algorithmic number theory (Sydney, 2002)*, volume 2369 of *Lecture Notes in Comput. Sci.*, pages 267–275. Springer, Berlin, 2002. [2](#)
- [38] Gerard van der Geer. *Hilbert modular surfaces*, volume 16 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1988. [3.2](#)
- [39] Mark Watkins. Some heuristics about elliptic curves. *Experiment. Math.*, 17(1):105–125, 2008. [1](#), [2](#)

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY, CAMBRIDGE, MA 02138 USA
 Email address: cowan@math.harvard.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OKLAHOMA, NORMAN, OK 73019 USA
 Email address: kimball.martin@ou.edu