Received: 5 June 2020 | Revised: 7 July 2021 | Accepted: 29 July 2021

DOI: 10.1112/jlms.12518

RESEARCH ARTICLE

Journal of the London Mathematical Society

The least degree of a CM point on a modular curve

Pete L. Clark[1] | **Tyler Genao**[1] | **Paul Pollack**[2] | **Frederick Saia**[1]

[1] Department of Mathematics, University of Georgia, Athens, Georgia, USA

[2] Department of Mathematics, University of Georgia, Athens, Georgia, USA

Correspondence
Pete L. Clark, Department of Mathematics, University of Georgia, Athens, GA 30602.
Email: plclark@gmail.com

Funding information
Research and Training Group, Grant/Award Number: DMS-1344994; National Science Foundation Graduate Research Fellowship, Grant/Award Number: 1842396; National Science Foundation, Grant/Award Number: DMS-2001581

Abstract

For a modular curve $X = X_0(N)$, $X_1(N)$ or $X_1(M,N)$ defined over $\mathbb{Q}$, we denote by $d_{\mathrm{CM}}(X)$ the least degree of a CM point on X. For each discriminant $\Delta < 0$, we determine the least degree of a point on $X_0(N)$ with CM by the order of discriminant Δ. This places us in a position to study $d_{\mathrm{CM}}(X)$ as an 'arithmetic function' and we do so, obtaining various upper bounds, lower bounds and typical bounds. We deduce that all but finitely many curves in each of the families have sporadic CM points. Finally, we supplement these results with a computational study, for example, computing $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ exactly for $N \leqslant 10^6$ and determining whether $X_0(N)$ (respectively, $X_1(N)$, $X_1(M,N)$) has sporadic CM points for all but 106 values of N (respectively, 227 values of N, 146 pairs (M,N) with $M \geqslant 2$).

MSC (2020)
11G15, 11G30 (primary)

Contents

1 | INTRODUCTION

We study elliptic curves with complex multiplication (CM) over number fields, an ongoing project of the present authors and our collaborators [5–8, 10, 13, 15–20].

In particular, we seek to understand the extremal behavior of torsion points on CM elliptic curves over number fields. For each $d \in \mathbb{Z}^+$, as one varies over all CM elliptic curves $E_{/F}$ defined over all number fields of degree d, up to isomorphism there are only finitely many possibilities for the torsion subgroup $E(F)[\mathrm{tors}]$ [56, Corollary 7]. (The same holds for non-CM elliptic curves but lies much deeper [46].) One way to measure the extremal behavior is to study the upper order of the function $T_{\mathrm{CM}}(d)$ which is the maximum size of the torsion subgroup of a CM elliptic curve defined over a degree d number field. The following result of Clark and Pollack, building upon work of Breuer [11], completely determines the extremal behavior in this sense.

Theorem 1.1 (Clark–Pollack [20, Theorem 1.1]). *We have*

$$\limsup_{d\to\infty} \frac{T_{\mathrm{CM}}(d)}{d \log\log d} = \frac{e^\gamma \pi}{\sqrt{3}}.$$

Here is a 'dual' measure: we fix $M, N \in \mathbb{Z}^+$ with M dividing N and ask for the least degree of a number field F over which there is a CM elliptic curve $E_{/F}$ and an injection of groups

$$\mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \hookrightarrow E(F)[\mathrm{tors}].$$

Let us now recast and generalize this problem in terms of modular curves. As we will recall in more detail in Section 7.1, to a subgroup H of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ we attach a modular curve $X(H)_{/\mathbb{Q}}$ that is smooth, projective and integral (but not necessarily geometrically integral), and such a curve comes equipped with a $\mathbb{Q}$-morphism to the j-line $\pi : X(H) \to X(1)$ of degree

$$I(H) := [\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\} : H].$$

If F is a field of characteristic 0 and $E_{/F}$ is an elliptic curve for which the modulo N Galois $\pm$-representation

$$\overline{\rho}_N : \mathrm{Aut}(\overline{F}/F) \to \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \twoheadrightarrow \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$$

lies in H, then $E_{/F}$ induces an F-valued point on $X(H)$. Conversely every F-valued point on $X(H)$ is induced by at least one such elliptic curve, and any two elliptic curves inducing the same point have the same j-invariant. We say a closed point $p \in X(H)$ is a **CM point** if the corresponding $\mathbb{Q}(p)$-valued point is induced by a CM elliptic curve — equivalently, if $\pi(p)$ is a CM j-invariant. For an imaginary quadratic order $\mathcal{O}$, we say that a closed point $p \in X(H)$ is an **$\mathcal{O}$-CM point** if the corresponding elliptic curve has endomorphism ring $\mathcal{O}$.

For a modular curve $X = X(H)$, let $d_{\mathrm{CM}}(X)$ be the least degree $[\mathbb{Q}(p) : \mathbb{Q}]$ of a CM point $p \in X$. For an imaginary quadratic discriminant Δ, let $d_{\Delta,\mathrm{CM}}(X)$ be the least degree of a point with CM by the order[†] of discriminant Δ. (This constrains p to lie in the fiber over a single closed point on $X(1)_{/\mathbb{Q}}$.) Thus, we have

$$d_{\mathrm{CM}}(X) = \min_{\Delta} d_{\Delta,\mathrm{CM}}(X).$$

We now find a crucial distinction: to compute $d_{\Delta,\mathrm{CM}}(X(H))$ for each Δ is a problem in arithmetic geometry. This problem was solved for the curves $X(N)$ by Stevenhagen [57] and again by Bourdon–Clark [5]. It was solved for the curves $X_1(M,N)$ by Bourdon–Clark [6]. We solve it here for the curves $X_0(N)$, building on [6]. In every such case the formula for $d_{\Delta,\mathrm{CM}}(X(H))$ involves the class number h_Δ of the order of discriminant Δ: indeed, if $j : X(H) \to X(1)$ is the natural map, then for any Δ-CM point p on $X(H)$ we have $h_\Delta = [\mathbb{Q}(j(P)) : \mathbb{Q}] \mid [\mathbb{Q}(p) : \mathbb{Q}]$.

Even when $d_{\Delta,\mathrm{CM}}(X(H))$ is known for all Δ, the minimization over Δ remains an interesting problem in its own right, and one with a more analytic flavor.

It is exactly these analytic problems that are our focus in the present work. By combining our arithmetic-geometric results with methods of elementary and analytic number theory, we derive several statistical results on the distribution of $d_{\mathrm{CM}}(X_0(N))$, $d_{\mathrm{CM}}(X_1(N))$, and $d_{\mathrm{CM}}(X_1(M,N))$, as N (or as M,N) vary. Our theorems are listed below, beginning with bounds for the extremal orders. When needed to obtain sharp (or close to sharp) results, we assume the Riemann hypothesis for Dirichlet L-functions, henceforth denoted as GRH (but see Section 6 for unconditional results).

Theorem 1.2 (Lower order of $d_{\mathrm{CM}}(X_0(N))$).

(a) *We have $d_{\mathrm{CM}}(X_0(N)) \geqslant 2$ for all $N > 163$. (See Table 1 for the list of N such that $d_{\mathrm{CM}}(X_0(N)) = 1$.)*

[†] Here and hereafter, 'an order' is always a $\mathbb{Z}$-order in an imaginary quadratic field.

TABLE 1 All integers $N \in \mathbb{Z}^+$ for which $d_{\mathrm{CM}}(X_0(N)) = 1$

N	$d_{\mathrm{CM}}(X_0(N))$	$d_{\mathrm{CM}}(X_1(N))$
1	1	1
2	1	1
3	1	1
4	1	1
6	1	1
7	1	2
9	1	3
11	1	5
14	1	3
19	1	6
27	1	9
43	1	14
67	1	22
163	1	54

(b) *Fix $C \geqslant 2$. Then as $X \to \infty$ we have*

$$\#\{1 \leqslant N \leqslant X \mid d_{\mathrm{CM}}(X_0(N)) \leqslant C\} \asymp \frac{X}{\sqrt{\log X}}.$$

Combining part (a) and part (b) with $C = 2$ we get: $\liminf_{N\to\infty} d_{\mathrm{CM}}(X_0(N)) = 2$.

Theorem 1.3 (Upper order of $d_{\mathrm{CM}}(X_0(N))$). *Assume GRH.*

(a) *There is a sequence of N tending to infinity along which we have*

$$d_{\mathrm{CM}}(X_0(N)) \geqslant \exp\left(\left(\frac{1}{4}\log 2 + o(1)\right)\frac{\log N}{\log\log N}\right).$$

(b) *As N tends to infinity through all positive integers, we have*

$$d_{\mathrm{CM}}(X_0(N)) \leqslant \exp\left((\log 2 + o(1))\frac{\log N}{\log\log N}\right).$$

Further questions about the upper order of $d_{\mathrm{CM}}(X_0(N))$ are raised in Remarks 5.3 and 5.4.

Theorem 1.4 (Lower order of $d_{\mathrm{CM}}(X_1(N))$). *We have $d_{\mathrm{CM}}(X_1(N)) \gg \frac{N}{\sqrt{\log\log N}}$ for all large N, and $d_{\mathrm{CM}}(X_1(N)) \ll \frac{N}{\sqrt{\log\log N}}$ on a sequence of N tending to infinity.*

Theorem 1.5 (Upper order of $d_{\mathrm{CM}}(X_1(N))$). *Assume GRH.*

(a) *There is a sequence of N tending to infinity along which we have*

$$d_{\mathrm{CM}}(X_1(N)) \geqslant N \exp\left(\left(\frac{1}{4}\log 2 + o(1)\right)\frac{\log N}{\log\log N}\right).$$

(b) *As N tends to infinity through all positive integers, we have*

$$d_{\mathrm{CM}}(X_1(N)) \leqslant N \exp\left((\log 2 + o(1))\frac{\log N}{\log\log N}\right).$$

Turning to $d_{\mathrm{CM}}(X_1(M,N))$, we first observe the following consequence of Theorem 1.1.

Theorem 1.6 (Lower order of $d_{\mathrm{CM}}(X_1(M,N))$). *Let M, N be positive integers for which $M \mid N$. If N is sufficiently large, then*

$$d_{\mathrm{CM}}(X_1(M,N)) \gg \frac{MN}{\log\log N}. \tag{1}$$

Moreover, there are infinitely many pairs of positive integers M, N, where $M \mid N$ and

$$d_{\mathrm{CM}}(X_1(M,N)) \ll \frac{MN}{\log\log N}. \tag{2}$$

Proof. Theorem 1.1 implies that if F is a degree d number field and $E_{/F}$ is a CM elliptic curve with $\mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \hookrightarrow E(F)[\mathrm{tors}]$, then $MN \leqslant \#E(F)[\mathrm{tors}] \ll d\log\log d$. So if $d \leqslant MN$, then $MN \ll d\log\log MN$, and $d \gg MN/\log\log MN$. This final estimate holds also when $d > MN$. Therefore,

$$d_{\mathrm{CM}}(X_1(M,N)) \gg \frac{MN}{\log\log MN} \gg \frac{MN}{\log\log N}.$$

The proof of the second half of Theorem 1.6 is similar. By Theorem 1.1, we can choose a sequence of $d \to \infty$ and a corresponding sequence of elliptic curves $E_{/F}$ with $[F : \mathbb{Q}] = d$, having $E(F)[\mathrm{tors}] \cong \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ (where $M \mid N$) and $MN = \#E(F)[\mathrm{tors}] \gg d\log\log d$. Then $d_{\mathrm{CM}}(X_1(M,N)) \leqslant d \ll MN/\log\log MN \ll MN/\log\log N$. □

We do not have a sharp result for the upper order of $d_{\mathrm{CM}}(X_1(M,N))$. However, we will prove (see Lemma 4.4) that whenever $M \mid N$, we have

$$d_{\mathrm{CM}}(X_1(M,N)) \leqslant 2M \cdot d_{\mathrm{CM}}(X_1(N)).$$

Combining this with the upper bound on $d_{\mathrm{CM}}(X_1(N))$ from Theorem 1.5, we see that under GRH, for all $\epsilon > 0$ we have

$$d_{\mathrm{CM}}(X_1(M,N)) \ll_\epsilon MN^{1+\epsilon}.$$

This universal upper bound differs from the universal lower bound (1) by a factor of order smaller than any fixed positive power of N.

Of special interest in the study of $d_{\mathrm{CM}}(X_1(M,N))$ is the full-torsion case $M = N$, corresponding to the modular curve $X(N)$. We prove that $d_{\mathrm{CM}}(X(N))$ has lower order $\frac{N^2}{\log\log N}$ and upper order N^2. This lower order result refines the second half of Theorem 1.6 by showing that (2) holds with $M = N$ for infinitely many N.

The last of our main analytic results concerns the sizes of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ for *typical* inputs N. It should be read as asserting that for most inputs, we have

$$d_{\mathrm{CM}}(X_0(N)) \approx (\log N)^{\frac{1}{2}\log 2} \quad \text{and} \quad d_{\mathrm{CM}}(X_1(N)) \approx N(\log N)^{\frac{1}{2}\log 2}.$$

Theorem 1.7. *Assume GRH. Fix $\epsilon > 0$.*

(a) *As $x \to \infty$, all but $o(x)$ positive integers $N \leqslant x$ satisfy*

$$2^{(\frac{1}{2}-\epsilon)\log\log x} \leqslant d_{\mathrm{CM}}(X_0(N)) \leqslant 2^{(\frac{1}{2}+\epsilon)\log\log x}. \tag{3}$$

(b) *All but $o(x)$ integers $N \leqslant x$ satisfy*

$$N \cdot 2^{(\frac{1}{2}-\epsilon)\log\log x} \leqslant d_{\mathrm{CM}}(X_1(N)) \leqslant N \cdot 2^{(\frac{1}{2}+\epsilon)\log\log x}.$$

The proofs of these analytic theorems draw from the stream of ideas in [19, 20] but also require tools from probabilistic number theory and the 'anatomy of integers'. In some cases our methods seem to be of wider interest. Here is one example. For each $N \in \mathbb{Z}^+$, let $R(N)$ denote the least positive integer that is not a square but reduces to a square mod N. The proof of Theorem 1.7 can be modified to show that on GRH we have $R(N) = (\log N)^{\log 2 + o(1)}$ as $N \to \infty$ along a set of integers of asymptotic density 1. It is somewhat surprising, given the simple definition of $R(N)$ and the long history of investigations into the distribution of power residues, that this normal order theorem for $R(N)$ also seems to be new.

It has been asked, for example, by Hindry and Silverman [32], whether the CM case gives the extremal behavior of torsion points on elliptic curves over number fields. One way to analyze this is to compare the functions $T(d)$, $T_{\mathrm{CM}}(d)$ and $T_{\neg\,\mathrm{CM}}(d)$, where $T(d)$ is the largest size of a torsion subgroup of an elliptic curve defined over a degree d number field and $T_{\neg\,\mathrm{CM}}(d)$ is the same but restricted to elliptic curves *without* CM. Evidently for each $d \in \mathbb{Z}^+$ we have

$$T(d) = \max\{T_{\mathrm{CM}}(d), T_{\neg\,\mathrm{CM}}(d)\},$$

so which is it? Breuer showed [11] that the upper order of $T_{\neg\,\mathrm{CM}}(d)$ is at least $\sqrt{d\log\log d}$:

$$\limsup_d \frac{T_{\neg\,\mathrm{CM}}(d)}{\sqrt{d\log\log d}} > 0.$$

It may well be that the above limit supremum is finite, but showing this seems out of present reach. If that finiteness holds, in view of Theorem 1.1 we would have $\limsup_d \frac{T_{\mathrm{CM}}(d)}{T_{\neg\,\mathrm{CM}}(d)} = \infty$. On the other hand, we have [18, Remark 2.3]

$$\liminf_d \frac{T_{\neg\,\mathrm{CM}}(d)}{\sqrt{d}} > 0, \tag{4}$$

whereas [8, Theorem 1.4]

$$\liminf_d T_{\mathrm{CM}}(d) = 6.$$

Moreover, combining (4) and [7, Theorem 1.1] — that 'torsion is typically bounded on CM elliptic curves over number fields' — it follows that $T_{\neg\,\mathrm{CM}}(d) > T_{\mathrm{CM}}(d)$ on a set of positive integers of density 1.

For a modular curve $X_{/\mathbb{Q}}$, let $d(X)$ (respectively $d_{\neg\,\mathrm{CM}}(X)$) be the least degree of a non-cuspidal (respectively of a non-cuspidal, non-CM) closed point on X, so

$$d(X) = \min\{d_{\mathrm{CM}}(X), d_{\neg\,\mathrm{CM}}(X)\}.$$

Again, which is it? For the modular curves $X_0(N)$ and $X_1(M,N)$, it seems likely that $d(X) = d_{\mathrm{CM}}(X)$ 'most of the time'. For any modular curve $X(H)$ we have the upper bound

$$d_{\neg\,\mathrm{CM}}(X(H)) \leqslant I(H) = \deg(X(H) \to X(1)).$$

Indeed, starting with any non-CM $j \in \mathbb{Q}$ and pulling back along $X(H) \to X(1) \cong \mathbb{P}^1$ gives a closed non-CM point of degree at most $I(H)$. Moreover, as $\mathbb{P}^1(\mathbb{Q})$ is infinite and there are only finitely many closed CM points on a modular curve of any fixed degree, $d_{\neg\,\mathrm{CM}}(X(H))$ is at most the degree of any finite morphism to $\mathbb{P}^1$, but by a result of Abramovich (see Section 7) this only improves the upper bound by an absolute constant, and thus it is as yet unknown whether as $X(H)$ ranges over all modular curves we have $\liminf_H \frac{d_{\neg\,\mathrm{CM}}(X(H))}{I(H)} > 0$. If this holds, then there would be a constant $\mathbf{A}$ such that for all $N \geqslant \mathbf{A}$ and all $M \mid N$ we have

$$d_{\mathrm{CM}}(X_0(N)) < d_{\neg\,\mathrm{CM}}(X_0(N)) \quad \text{and} \quad d_{\mathrm{CM}}(X_1(M,N)) < d_{\neg\,\mathrm{CM}}(X_1(M,N)).$$

In summary, for many modular curves X our upper bounds on $d_{\mathrm{CM}}(X)$ give the best known upper bounds on $d(X)$, which may lie close to the truth. To show this is again beyond present reach, but this time we can establish a result in this direction.

Theorem 1.8. *There is a constant* $\mathbf{A}$ *such that for all* $N \geqslant \mathbf{A}$ *and all* $M \mid N$*, the curves* $X_0(N)_{/\mathbb{Q}}$ *and* $X_1(M,N)_{/\mathbb{Q}(\zeta_M)}$ *have sporadic CM points.*

The general definition of a sporadic point on a modular curve $X(H)$ involves some field of definition considerations, so we defer it to Section 7. The curves $X_0(N)$ and $X_1(N)$ are defined and geometrically integral over $\mathbb{Q}$, and a sporadic point is a closed point p whose degree $[\mathbb{Q}(p) : \mathbb{Q}]$ is smaller than that of the degree of all but finitely many closed points.

In Section 8, we give explicit finite sets of N (respectively pairs (M,N)) such that away from these sets, the curves $X_0(N)$ and $X_1(N)$ (respectively $X_1(M,N)$)) have sporadic CM points. These results show in particular that in the setting of Theorem 1.8 we may take $\mathbf{A} = 8581$. We also report on computations of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ for all $N \leqslant 10^6$ and of $d_{\mathrm{CM}}(X_1(M,N))$ for all (M,N) with $M \mid N \leqslant 53$.

Notation

Most of our notation is standard, or will be introduced as necessary, but one exception is worth highlighting. *For the remainder of the paper*, we adopt the following convention for logarithms and iterated logarithms, borrowed from [3]. We write $\ln x$ for the usual natural logarithm and we set

$$\log x := \max\{2, \ln x\}.$$

Thus, $\log x \geqslant 2$ for all $x > 0$, and the same holds for iterated logarithms. This allows us to state several upper and lower bounds in a uniform way, without cavil over small arguments. Moreover, since $\log x$ is bounded away from 1, it allows us to absorb positive constants into bounded powers of $\log x$ (or of iterates of $\log x$). Owing to this convention, the constant 0.693147 ... , which plays a role in several of our results, will be written as $\ln 2$ from now on rather than as $\log 2$.

2 | THE CLASS NUMBER h_Δ

For each negative integer Δ that is 0 or 1 modulo 4, there is a unique order $\mathcal{O}(\Delta)$ of discriminant Δ. The fraction field of $\mathcal{O}(\Delta)$ is $K = \mathbb{Q}(\sqrt{\Delta})$. In all cases, the unit group $\mathcal{O}(\Delta)^\times$ is generated by $e^{\frac{2\pi i}{u(\Delta)}}$, where

$$u(\Delta) = \begin{cases} 6 & \text{if } \Delta = -3 \\ 4 & \text{if } \Delta = -4. \\ 2 & \text{if } \Delta < -4 \end{cases}$$

For a discriminant $\Delta = \mathfrak{f}^2 \Delta_K < 0$, let $H_\Delta(j) \in \mathbb{Z}[j]$ be the Hilbert class polynomial: it is the monic separable polynomial whose roots in $\mathbb{C}$ are the j-invariants of the $\mathcal{O}(\Delta)$-CM elliptic curves. It has degree h_Δ and is irreducible over K. We put

$$\mathcal{R}^\circ(\Delta) := \mathbb{Q}[j]/H_\Delta.$$

Then

$$\mathcal{R}(\Delta) := K\mathcal{R}^\circ(\Delta)$$

is the ring class field of K of conductor $\mathfrak{f}$ and we have $[\mathcal{R}^\circ(\Delta) : \mathbb{Q}] = [\mathcal{R}(\Delta) : K] = h_\Delta$.

(Warning aside: for all $\Delta < 0$, the number field $\mathcal{R}(\Delta)$ is Galois over $\mathbb{Q}$ and thus well-defined as a subfield of $\mathbb{C}$. On the other hand, the number field $\mathcal{R}^\circ(\Delta)$ is only Galois over $\mathbb{Q}$ for finitely many values of Δ and thus in general it has several conjugate copies inside $\mathbb{C}$. This causes no trouble for us here, but it is worth keeping in mind.)

We put

$$h_\Delta := \#\operatorname{Pic}\mathcal{O}(\Delta),$$

the class number of $\mathcal{O}(\Delta)$.

Let $\mathbb{Z}_K$ be the ring of integers of K. This is the maximal order in K, and for each $\mathfrak{f} \in \mathbb{Z}^+$ there is a unique order $\mathcal{O}$ such that $[\mathbb{Z}_K : \mathcal{O}] = \mathfrak{f}$ and $\Delta(\mathcal{O}) = \mathfrak{f}^2\Delta_K$. Throughout this paper, we work with all imaginary quadratic orders (equivalently, with all CM elliptic curves) and not just maximal orders, so we need information about h_Δ for all $\Delta < 0$, not just for fundamental discriminants Δ_K. But in fact the class number of a nonmaximal order is easily understood in terms of the class number of the corresponding maximal order, as the following result shows.

Theorem 2.1 (Relative Class Number Formula). *Let K be an imaginary quadratic field with ring of integers $\mathbb{Z}_K$, and let $\mathfrak{f} \in \mathbb{Z}^+$. Then we have*

$$\frac{h_{\mathfrak{f}^2\Delta_K}}{h_{\Delta_K}} = \frac{2}{\#\mathbb{Z}_K^\times}\mathfrak{f}\prod_{p\mid\mathfrak{f}}\left(1-\left(\frac{\Delta_K}{p}\right)\frac{1}{p}\right). \tag{5}$$

Proof. See [21, Corollary 7.24]. □

Next we record some upper and lower bounds for h_Δ. They are easy consequences of well-known results, but we will include proofs for completeness.

Lemma 2.2. *Fix $\epsilon > 0$. For all $\Delta < 0$, we have*

$$|\Delta|^{\frac{1}{2}-\epsilon} \ll_\epsilon h_\Delta \ll_\epsilon |\Delta|^{\frac{1}{2}+\epsilon}.$$

Proof. A celebrated theorem of Siegel [55] gives $h_{\Delta_K} \gg |\Delta_K|^{\frac{1}{2}-\epsilon}$ whenever Δ_K is a fundamental discriminant. For $\Delta = \mathfrak{f}^2\Delta_K < 0$, using (5) and $1 \leqslant \#\mathcal{O}(\Delta)^\times \leqslant 6$, we get

$$h_\Delta \gg h_{\Delta_K}\mathfrak{f}\prod_{p\mid\mathfrak{f}}\left(1-\frac{1}{p}\right) = h_{\Delta_K}\phi(\mathfrak{f}).$$

Since $\phi(\mathfrak{f}) \gg_\epsilon \mathfrak{f}^{1-2\epsilon}$ (cf. [31, p. 352, Theorem 327]), Siegel's theorem gives

$$h_\Delta \gg_\epsilon |\Delta_K|^{\frac{1}{2}-\epsilon}\mathfrak{f}^{1-2\epsilon} = |\mathfrak{f}^2\Delta_K|^{\frac{1}{2}-\epsilon} = |\Delta|^{\frac{1}{2}-\epsilon}.$$

As for the upper bound: it is elementary to prove (for example, from Dirichlet's class number formula) that $h_{\Delta_K} \ll |\Delta_K|^{1/2}\log|\Delta_K|$ for fundamental discriminants Δ_K. Writing $\Delta = \mathfrak{f}^2\Delta_K$ as above, we find that $h_\Delta \ll h_{\Delta_K}\mathfrak{f} \ll |\Delta|^{1/2}\log|\Delta_K| \ll_\epsilon |\Delta|^{\frac{1}{2}+\epsilon}$. □

Lemma 2.3. *Assume GRH. For any negative discriminant Δ, we have*

$$h_\Delta = |\Delta|^{1/2}(\log\log|\Delta|)^{O(1)}.$$

Proof. Write $\Delta = \mathfrak{f}^2\Delta_K$, where Δ_K is a fundamental discriminant. Then

$$h_\Delta \asymp h_{\Delta_K}\mathfrak{f}\prod_{p\mid\mathfrak{f}}\left(1-\left(\frac{\Delta_K}{p}\right)\frac{1}{p}\right) = h_{\Delta_K}\mathfrak{f}\cdot(\log\log|\Delta|)^{O(1)}.$$

(In the final step we use that the product on p is bounded below by $\phi(\mathfrak{f})/\mathfrak{f}$, which is $\gg 1/\log\log|\Delta|$, and bounded above by $\mathfrak{f}/\phi(\mathfrak{f})$, which is $\ll \log\log|\Delta|$.) By Dirichlet's analytic class number formula, we have

$$h_{\Delta_K} \asymp \sqrt{|\Delta_K|} \cdot L(1, \left(\frac{\Delta_K}{\cdot}\right)).$$

Littlewood [44] has shown that under GRH,

$$\begin{aligned} L\left(1, \left(\frac{\Delta_K}{\cdot}\right)\right) &= (\log\log|\Delta_K|)^{O(1)} \\ &= (\log\log|\Delta|)^{O(1)}. \end{aligned}$$

Collecting our estimates, we find that $h_\Delta = \mathfrak{f}\sqrt{|\Delta_K|} \cdot (\log\log|\Delta|)^{O(1)}$. □

3 | EXACT RESULTS ON $d_{\Delta,\mathrm{CM}}(X)$ AND $d_{\mathrm{CM}}(X)$

3.1 | $X_1(M,N)$ and $X(N)$

Let $M \mid N$, let $\Delta = \mathfrak{f}^2\Delta_K < 0$ be a discriminant, and put $K = \mathbb{Q}(\sqrt{\Delta})$. In [6, Section 8], Bourdon–Clark give exact formulae for $d_{\Delta,\mathrm{CM}}(X_1(M,N))$. The results are somewhat intricate and involve several cases. Most of the complexity is not relevant to our asymptotic study (though it is relevant to our computational work): for instance, the hardest part is to decide whether the answer is the same for the curve $X_1(M,N)_{/\mathbb{Q}}$ as it is for the curve $X_1(M,N)_{/K}$ — but this involves only a factor of $[K:\mathbb{Q}] = 2$, a discrepancy that is absolutely harmless for analytic purposes. We will not record the general result here.

In the case of $M = N$ we get a much nicer formula. Rather than specializing the results of [6, Section 8], it is cleaner to make use of earlier results of Bourdon–Clark–Stankewicz [8] and Bourdon–Clark [5]. The latter result, giving an exact formula for the least degree of an $\mathcal{O}(\Delta)$-CM point on $X(N)_{/K}$, is essentially due to Stevenhagen [57].

As mentioned above, for $N \leqslant 2$ we have $d_{\mathrm{CM}}(X(N)) = 1$.

Theorem 3.1. *Let $\mathcal{O}$ be an order of discriminant $\Delta < 0$. For all $N \geqslant 3$ we have*

$$d_{\Delta,\mathrm{CM}}(X(N)) = \frac{2h_\Delta}{\#\mathcal{O}^\times}\phi(N)N\prod_{p\mid N}\left(1 - \left(\frac{\Delta}{p}\right)\frac{1}{p}\right).$$

Proof. Let K be the fraction field of $\mathcal{O}$. Let $E_{/F}$ be an $\mathcal{O}$-CM elliptic curve defined over a field F of characteristic 0, and let $\mathfrak{h} : E \to E/\mathrm{Aut}(E) \xrightarrow{\sim} \mathbb{P}^1$ be a Weber function on E. By [8, Lemma 3.15], if $(\mathbb{Z}/N\mathbb{Z})^2 \hookrightarrow E(F)$, then $F \supset K$, so $F \supset \mathcal{R}(\Delta)(\mathfrak{h}(E[N]))$. By [5, Theorem 1.4], we have

$$[\mathcal{R}(\Delta)(\mathfrak{h}(E[N])) : \mathcal{R}(\Delta)] = \frac{\#(\mathcal{O}/N\mathcal{O})^\times}{\#\mathcal{O}^\times}.$$

We also know [5, Corollary 1.7] that there is an $\mathcal{O}$-CM elliptic curve E defined over $F = \mathcal{R}(\Delta)(\mathfrak{h}(E[N]))$ with $(\mathbb{Z}/N\mathbb{Z})^2 \hookrightarrow E(F)$, so

$$\begin{aligned} d_{\Delta,\mathrm{CM}}(X(N)) &= [\mathcal{R}(\Delta)(\mathfrak{h}(E[N])) : \mathbb{Q}] \\ &= [\mathcal{R}(\Delta)(\mathfrak{h}(E[N])) : \mathcal{R}(\Delta)][\mathcal{R}(\Delta) : K][K : \mathbb{Q}] \\ &= 2h_\Delta \cdot \frac{\#(\mathcal{O}/N\mathcal{O})^\times}{\#\mathcal{O}^\times}. \end{aligned}$$

Moreover by [5, Lemma 2.2], we have

$$\#(\mathcal{O}/N\mathcal{O})^\times = N^2 \prod_{p|N} \left(1 - \left(\frac{\Delta}{p}\right)\frac{1}{p}\right)\left(1 - \frac{1}{p}\right) = \phi(N)N \prod_{p|N} \left(1 - \left(\frac{\Delta}{p}\right)\frac{1}{p}\right),$$

so the formula follows. □

Theorem 3.2. *Let $\ell^a > 2$ be a prime power. Then we have*

$$d_{\mathrm{CM}}(X(\ell^a)) = d_{-3,\mathrm{CM}}(X(\ell^a)) = \frac{1}{3}\ell^{2a-2}(\ell - 1)\left(\ell - \left(\frac{-3}{\ell}\right)\right).$$

Proof. In view of Theorem 3.1 it suffices to prove the first equality. For each prime power $\ell^a > 3$, as we range over all imaginary quadratic discriminants Δ we must show that the quantity $d_{\Delta,\mathrm{CM}}(X(\ell^a))$ is minimized when $\Delta = -3$.

- Suppose $\ell = 2$, so $a \geqslant 2$. Then

$$d_{-3,\mathrm{CM}}(X(2^a)) = d_{-4,\mathrm{CM}}(X(2^a)) = 2^{2a-2},$$

  whereas if $\Delta < -4$, then

$$d_{\Delta,\mathrm{CM}}(X(2^a)) \geqslant 2^{2a}\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right) = 2^{2a-2}.$$

- Suppose $\ell = 3$. Then

$$d_{-3,\mathrm{CM}}(X(3^a)) = 2 \cdot 3^{2a-2},$$

  while

$$d_{-4,\mathrm{CM}}(X(3^a)) = 4 \cdot 3^{2a-2} > d_{-3,\mathrm{CM}}(X(3^a)),$$

  and for $\Delta < -4$, we have

$$d_{\Delta,\mathrm{CM}}(X(3^a)) > 3^{2a}\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{3}\right) = 4 \cdot 3^{2a-2} > d_{-3,\mathrm{CM}}(X(3^a)).$$

- Suppose $\ell \geqslant 5$. Then

$$d_{-3,\mathrm{CM}}(X(\ell^a)) = \frac{1}{3}\ell^{2a-2}(\ell-1)\left(\ell - \left(\frac{-3}{\ell}\right)\right) \leqslant \frac{\ell^{2a-2}(\ell^2-1)}{3},$$

while

$$d_{-4,\mathrm{CM}}(X(\ell^a)) \geqslant \frac{\ell^{2a-2}(\ell-1)^2}{2} = \frac{\ell^{2a-2}(\ell^2-1)}{2}\frac{\ell-1}{\ell+1} \geqslant \frac{2}{3}\frac{\ell^{2a-2}(\ell^2-1)}{2} = d_{-3,\mathrm{CM}}(X(\ell^a)).$$

For $\Delta < -4$, a similar calculation shows that

$$d_{-\Delta,\mathrm{CM}}(X(\ell^a)) \geqslant 2d_{-3,\mathrm{CM}}(X(\ell^a)).$$ □

Remark 3.3. Let $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$.

(a) The proof of Theorem 3.2 gives $d_{\mathrm{CM}}(X(N)) = d_{-3,\mathrm{CM}}(X(N))$ if $\prod_{i=1}^{r} \frac{\ell_i - 1}{\ell_i + 1} \geqslant \frac{2}{3}$.

(b) However, if $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$ with each $\ell_i \equiv 5 \pmod{12}$, then

$$\frac{d_{-4,\mathrm{CM}}(X(N))}{d_{-3,\mathrm{CM}}(X(N))} = \frac{2}{3}\prod_{i=1}^{r}\frac{\ell_i - 1}{\ell_i + 1}.$$

The Prime Number Theorem for Arithmetic Progressions shows that this ratio can be arbitrarily small. A similar argument shows that for $\Delta_1, \ldots, \Delta_n$ any n imaginary quadratic discriminants, the quantity $\frac{d_{\mathrm{CM}}(X(N))}{\min_i d_{\Delta_i,\mathrm{CM}}(X(N))}$ can be arbitrarily small.

3.2 | $X_1(N)$

We have $X_1(N) = X(1,N)$, and the computation of $d_{\Delta,\mathrm{CM}}(X_1(N))$ was done in [6, Section 7] as a stepping stone to the general case of $X_1(M,N)$. For later use, we record these values for $\Delta = -3$ and $\Delta = -4$. Here and below, ψ denotes Dedekind's function, defined by $\psi(N) := N\prod_{p|N}(1+1/p)$.

Theorem 3.4. *Let $N \in \mathbb{Z}^+$.*

(a) *We have $d_{-4,\mathrm{CM}}(X_1(1)) = d_{-4,\mathrm{CM}}(X_1(2)) = 1$.*

(b) *Let $N \geqslant 3$, and write*

$$N = 2^a p_1^{b_1} \cdots p_r^{b_r} q_1^{c_1} \cdots q_s^{c_s}$$

with $a \geqslant 0$, $r, s \geqslant 0$, $b_i, c_j \geqslant 1$, and distinct primes $p_i \equiv 1 \pmod 4$ and $q_j \equiv 3 \pmod 4$. Then

$$d_{-4,\mathrm{CM}}(X_1(N)) = \begin{cases} \phi(N)\cdot\dfrac{\phi(2^a)\prod_{j=1}^{s}\psi(q_j^{r_j})}{4} & \text{if } r = 0, \\ \phi(N)\cdot\dfrac{\phi(2^a)\prod_{j=1}^{s}\psi(q_j^{r_j})}{2} & \text{if } r \geqslant 1. \end{cases}$$

(c) *We have $d_{-3,\mathrm{CM}}(X_1(1)) = d_{-3,\mathrm{CM}}(X_1(2)) = d_{-3,\mathrm{CM}}(X_1(3)) = 1$.*

(d) *Let $N \geqslant 4$, and write*

$$N = 3^a p_1^{b_1} \cdots p_r^{b_r} q_1^{c_1} \cdots q_s^{c_s}$$

with $a \geqslant 0$, $r, s \geqslant 0$, $b_i, c_j \geqslant 1$ and distinct primes $p_i \equiv 1 \pmod 3$ and $q_j \equiv 2 \pmod 3$. Then

$$d_{-3,\mathrm{CM}}(X_1(N)) = \begin{cases} \phi(N) \cdot \dfrac{\lceil 3^{a-1} \rceil \prod_{j=1}^{s} \psi(q_j^{c_j})}{6} & \text{if } r = 0, \\ \phi(N) \cdot \dfrac{\lceil 3^{a-1} \rceil \prod_{j=1}^{s} \psi(q_j^{c_j})}{3} & \text{if } r \geqslant 1. \end{cases}$$

Proof. This is a special case of [5, Theorem 7.2; 6, Theorem 7.1]. □

3.3 | $X_0(N)$

Theorem 3.5. *Let $p \in X_0(N)_{/\mathbb{Q}}$ be a closed point with CM by an order $\mathcal{O}$ of discriminant $\Delta = \mathfrak{f}^2 \Delta_K < -4$ in an imaginary quadratic field K. Let $N \in \mathbb{Z}^+$, and consider the morphism of $\mathbb{Q}$-schemes $\pi : X_1(N) \to X_0(N)$.*

(a) *The morphism π is inert over p: that is, writing the fiber $\pi^*(p)$ as* Spec $A(p)$ *for a $\mathbb{Q}$-algebra $A(p)$, we have that $A(p)$ is a field.*
(b) *For every closed point $P \in X_1(N)$ with $\pi(P) = p$, we have*

$$[\mathbb{Q}(P) : \mathbb{Q}(p)] = \deg(\pi) = \begin{cases} 1 & \text{if } N \leqslant 2 \\ \frac{\varphi(N)}{2} & \text{if } N \geqslant 3 \end{cases}.$$

Proof. Part (a) is the special case $M = 1$ of [15, Theorem 1.2]. Part (b) follows immediately. □

We say that a positive integer N is of *Type I* if $\mathrm{ord}_3(N) \leqslant 1$ and N is not divisible by any prime $\ell \equiv 2 \pmod 3$. We say that a positive integer N is of *Type II* if $\mathrm{ord}_2(N) \leqslant 1$ and N is not divisible by any prime $\ell \equiv 3 \pmod 4$.

Remark 3.6.

(a) A positive integer N is of Type I if and only if there is a primitive ideal I of $\mathcal{O}(-3)$ of norm N, that is, such that $\mathcal{O}(-3)/I \cong \mathbb{Z}/N\mathbb{Z}$. Thus, for any field $F \supset \mathbb{Q}(\sqrt{-3})$ and any $\mathcal{O}(-3)$-CM elliptic curve $E_{/F}$, we have that $E \to E/E[I]$ is an F-rational cyclic N-isogeny.

For $N \in \mathbb{Z}^+$, by [5, Theorem 6.18c)] there is an $\mathcal{O}(-3)$-CM elliptic curve $E_{/\mathbb{Q}(\sqrt{-3})}$ admitting a $\mathbb{Q}(\sqrt{-3})$-rational cyclic N-isogeny if and only if there is $a \in \{1, 2, 3, 6\}$ such that $\frac{N}{a}$ is an integer of Type I. Whereas if N is of Type I the existence of a cyclic $\mathbb{Q}(\sqrt{-3})$-rational N-isogeny is independent of the $\mathbb{Q}(\sqrt{-3})$-rational model, in the exceptional cases some but not all $\mathcal{O}(-3)$-CM elliptic curves defined over $\mathbb{Q}(\sqrt{-3})$ admit such an isogeny.

(b) A positive integer N is of Type II if and only if there is a primitive ideal I of $\mathcal{O}(-4)$ of norm N. Thus, for any field $F \supset \mathbb{Q}(\sqrt{-1})$ and any $\mathcal{O}(-4)$-CM elliptic curve $E_{/F}$, we have that $E \to E/E[I]$ is an F-rational cyclic N-isogeny.

For $N \in \mathbb{Z}^+$, by [5, Theorem 6.18b)] there is an $\mathcal{O}(-4)$-CM elliptic curve $E_{/\mathbb{Q}(\sqrt{-1})}$ admitting a $\mathbb{Q}(\sqrt{-1})$-rational cyclic N-isogeny if and only if there is $a \in \{1, 2\}$ such that $\frac{N}{a}$ is an integer of Type II. As above, if $\frac{N}{2}$ is of Type II but N is not, then some but not all $\mathcal{O}(-4)$-CM elliptic curves defined over $\mathbb{Q}(\sqrt{-1})$ admit a $\mathbb{Q}(\sqrt{-1})$-rational cyclic N-isogeny.

Theorem 3.7. *Let $\mathcal{O}$ be an order of discriminant Δ, and let $N \geqslant 2$.*

(a) *We have $d_{-3,\mathrm{CM}}(X_0(2)) = d_{-4,\mathrm{CM}}(X_0(2)) = 1$. For $\Delta < -4$, we have*

$$d_{\Delta,\mathrm{CM}}(X_0(2)) = \begin{cases} h_\Delta & \text{if } \left(\frac{\Delta}{2}\right) \neq -1, \\ 3h_\Delta & \text{if } \left(\frac{\Delta}{2}\right) = -1. \end{cases}$$

(b) *If $N \geqslant 3$ and $\Delta < -4$, then*

$$d_{\Delta,\mathrm{CM}}(X_0(N)) = \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

(c) *Let $N \geqslant 3$ and $\Delta = -4$. Then:*
   (i) *if N is of Type II, then*

$$d_{-4,\mathrm{CM}}(X_0(N)) = 2 = \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/4};$$

   (ii) *if N is not of Type II, then*

$$d_{-4,\mathrm{CM}}(X_0(N)) = \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

(d) *Let $N \geqslant 3$ and $\Delta = -3$. Then:*
   (i) *if $N = 3$, then*

$$d_{-3,\mathrm{CM}}(X_0(3)) = 1;$$

   (ii) *if $N > 3$ is of Type I, then*

$$d_{-3,\mathrm{CM}}(X_0(N)) = 2 = \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/6};$$

   (iii) *if N is not of Type I, then*

$$d_{-3,\mathrm{CM}}(X_0(N)) = \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

Proof. By [8, Theorem 5.5], if an elliptic curve $E_{/F}$ has an F-rational cyclic N-isogeny, there is a field extension L/F of degree dividing $\frac{\phi(N)}{2}$ and a quadratic twist E^D of $E_{/L}$ such that $E^D(L)$ has a

point of order N. From this it follows

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \geqslant \frac{d_{\Delta,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

(a) We have $X_1(2) = X_0(2)$, so in this case the result is [5, Remark 7.3].
(b) The case in which $\Delta < -4$ is immediate from Theorem 3.5.
(c) Suppose $N \geqslant 3$ is of Type II. Then $N \notin \{1, 2, 4\}$, so by Section 3.1 we have $d_{-4,\mathrm{CM}}(X_0(N)) \geqslant 2$. By Remark 3.6, there is an $\mathcal{O}(-4)$-CM elliptic curve with a $\mathbb{Q}(\sqrt{-1})$-rational cyclic N-isogeny, so $d_{-4,\mathrm{CM}}(X_0(N)) = 2$. By Theorem 3.4, we have

$$d_{-4,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2} = \frac{\phi(N)}{4} d_{-4,\mathrm{CM}}(X_0(N)).$$

Now suppose $N = 2^a p_1^{b_1} \cdots p_r^{b_r} q_1^{c_1} \cdots q_s^{c_s}$ is not of Type II.

- Suppose $a = r = 0$. By Theorem 3.4 we have

$$2\frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)} = \frac{\prod_{j=1}^{s} \psi(q_j^{c_j})}{2} = [\mathcal{R}^\circ(N^2\Delta_K) : \mathbb{Q}].$$

By, for example, [6, Section 2.6], there is an $\mathcal{O}(-4N^2)$-CM elliptic curve $E'_{/\mathcal{R}^\circ(N^2\Delta_K)}$ and a $\mathcal{R}^\circ(N^2\Delta_K)$-rational cyclic N-isogeny

$$\iota_{N,1} : E' \to E,$$

where E is an $\mathcal{O}(-4)$-CM elliptic curve. The dual isogeny $\iota_{N,1}^\vee : E \to E'$ is also a cyclic N-isogeny defined over $\mathcal{R}^\circ(N^2\Delta_K)$, showing that

$$d_{-4,\mathrm{CM}}(X_0(N)) \leqslant [\mathcal{R}^\circ(N^2\Delta_K) : \mathbb{Q}] = \frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

- Suppose $a = 0$, $r \geqslant 1$. Put

$$M_1 := \frac{N}{p_1^{b_1} \cdots p_r^{b_r}} = q_1^{c_1} \cdots q_s^{c_s}, \quad M_2 := p_1^{b_1} \cdots p_r^{b_r}.$$

By Theorem 3.4 we have

$$2\frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)} = \phi(2^a) \prod_{j=1}^{s} \psi(q_j^{c_j}) = [\mathcal{R}(M_1^2\Delta_K) : \mathbb{Q}].$$

As above, there is an $\mathcal{O}(-4)$-CM elliptic curve $E_{/\mathbb{Q}(M_1)}$ admitting a $\mathbb{Q}(M_1)$-rational cyclic M_1-isogeny; let K_1 be its kernel. For $1 \leqslant i \leqslant r$, since $p_i \equiv 1 \pmod 4$ there is a prime ideal $\mathfrak{p}_i$ of $\mathcal{O}(-4)$ of norm p_i. Then $\eta : E \to E/E[\mathfrak{p}_1^{b_1} \cdots \mathfrak{p}_r^{b_r}]$ is a cyclic $K(M_1)$-rational M_2-isogeny; let K_2 be its kernel. Then $K := K_1 \oplus K_2$ is a $K(M_1)$-rational cyclic subgroup scheme of order

N, so $E \to E/K$ is a $K(M_1)$-rational cyclic N-isogeny, showing

$$d_{-4,\mathrm{CM}}(X_0(N)) \leqslant [K(N) : \mathbb{Q}] = \frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

- Suppose $a = 1$. We observe that since N is not of type II, neither is $\frac{N}{2}$. Since every $\mathcal{O}(-4)$-CM elliptic curve E defined over a number field F has the form $y^2 = x^3 + Ax$ and thus has an F-rational point of order 2, we have $d_{-4,\mathrm{CM}}(X_0(N/2)) = d_{-4,\mathrm{CM}}(X_0(N))$ and $d_{-4,\mathrm{CM}}(X_1(N/2)) = d_{-4,\mathrm{CM}}(X_1(N))$. Finally, we have $\phi(N/2) = \phi(N)$, so the result in this case follows from the $a = 0$ case.
- Suppose $a \geqslant 2$. In this case we have

$$\frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)} = 2\frac{d_{-4,\mathrm{CM}}(X_1(N/2))}{\phi(N/2)}.$$

Moreover, since $\deg(X_0(N) \to X_0(N/2)) = 2$, by induction on a we get

$$d_{-4,\mathrm{CM}}(X_0(N)) \leqslant 2d_{-4,\mathrm{CM}}(X_0(N/2)) = \frac{d_{-4,\mathrm{CM}}(X_1(N/2))}{\phi(N/2)/4} = \frac{d_{-4,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

(d) We know $d_{-3,\mathrm{CM}}(X_0(N)) = 1$ if and only if $N \in \{1, 2, 3, 6\}$, so suppose $N \notin \{1, 2, 3, 6\}$: then $d_{-3,\mathrm{CM}}(X_1(N)) \geqslant 2$. If N is of Type I, then by Remark 3.6 there is an $\mathcal{O}(-3)$-CM elliptic curve with a $\mathbb{Q}(\sqrt{-3})$-rational cyclic N-isogeny, so $d_{-3,\mathrm{CM}}(X_0(N)) = 2$. By Theorem 3.4, we have

$$d_{-3,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{3} = \frac{\phi(N)}{6}d_{-3,\mathrm{CM}}(X_0(N)).$$

Now suppose $N = 3^a p_1^{b_1} \cdots p_r^{b_r} q_1^{c_1} \cdots q_s^{c_s}$ is not of Type I.

- Suppose $a = r = 0$. By Theorem 3.4 we have

$$2\frac{d_{-3,\mathrm{CM}}(X_1(N))}{\phi(N)} = \frac{\prod_{j=1}^{s} \psi(q_j^{c_j})}{3} = [\mathcal{R}^\circ(N^2\Delta_K) : \mathbb{Q}].$$

Arguing as in part (c) we get an $\mathbb{Q}(N)$-rational $\mathcal{O}(-3)$-CM point on $X_0(N)$, showing that

$$d_{-3,\mathrm{CM}}(X_0(N)) \leqslant [\mathbb{Q}(N) : \mathbb{Q}] = \frac{d_{-3,\mathrm{CM}}(X_1(N))}{\phi(N)/2}.$$

- The case $a = 0$, $r \geqslant 1$ is handled as in part (c) above.
- Suppose $a = 1$. We observe that since N is not of Type II, neither is $\frac{N}{3}$. Let $\mathfrak{p}_3$ be the prime ideal of $\mathcal{O}(-3)$ of norm 3. Then for any $\mathcal{O}(-3)$-CM elliptic curve E defined over any number field F, the map $E \to E/E[\mathfrak{p}_3]$ is an F-rational 3-isogeny. From this it follows easily that $d_{-3,\mathrm{CM}}(X_0(N/3)) = d_{-3,\mathrm{CM}}(X_0(N))$, and from Theorem 3.4 we see that also $\frac{d_{-3,\mathrm{CM}}(N)}{\phi(N)} = \frac{d_{-3,\mathrm{CM}}(N/3)}{\phi(N/3)}$, so the result holds in this case.

- Suppose $a \geqslant 2$. In this case we have

$$\frac{d_{-3,\mathrm{CM}}(X_1(N))}{\phi(N)} = \frac{d_{-3,\mathrm{CM}}(X_1(N/3))}{\phi(N/3)/3},$$

and since $\deg(X_0(N) \to X_0(N/3)) = 3$, we may argue as in part (c), by induction on a. □

Remark 3.8. In forthcoming work [15] of the first author, the process employed here is reversed: for each $N \in \mathbb{Z}^+$ and Δ we determine all fields of moduli of closed $\mathcal{O}(\Delta)$-CM points on $X_0(N)$. Via Theorem 3.5, we deduce the set of degrees of closed $\mathcal{O}(\Delta)$-CM points on $X_1(N)$ (at least when $\Delta < -4$), which is more precise than the determination of $d_{\Delta,\mathrm{CM}}(X_1(N))$.

3.4 | Comparison of $d_{\mathrm{CM}}(X_1(N))$ with $\phi(N)$

Let K be an imaginary quadratic field, let $\mathcal{O}$ be an order in K of discriminant Δ, and let N be a positive integer. Recall $\mathcal{R}^\circ(\Delta)$ is the minimal field of definition of an $\mathcal{O}$-CM elliptic curve. As in [6], let $T^\circ(\mathcal{O}, N)$ be the least degree $[F : \mathcal{R}^\circ(\Delta)]$ of a field extension $F/\mathcal{R}^\circ(\Delta)$ over which there is an $\mathcal{O}$-CM elliptic curve $E_{/F}$ with an F-rational point of order N. Since $[\mathcal{R}^\circ(\Delta) : \mathbb{Q}] = h_\Delta$ we have

$$d_{\Delta,\mathrm{CM}}(X_1(N)) = h_\Delta T^\circ(\mathcal{O}, N).$$

As we saw in the previous section, for all $\Delta < 0$ and all $N \in \mathbb{Z}^+$ we have

$$\phi(N) \mid 6d_{\Delta,CM}(X_1(N)).$$

So, it is natural and computationally useful to understand the minimal values of $\frac{d_{\mathrm{CM}}(X_1(N))}{\phi(N)}$. The results of this section accomplish this. In particular, this explains the behavior of discriminants $\Delta = -3$ and $\Delta = -4$ relative to all other imaginary quadratic discriminants.

Theorem 3.9. *Let $\mathcal{O}$ be an order of discriminant Δ, and let $N \in \mathbb{Z}^+$.*

(a) *We have $T^\circ(\mathcal{O}, N) \geqslant \frac{\phi(N)}{3}$.*
(b) *The following are equivalent.*
  (i) *We have $T^\circ(\mathcal{O}, N) = \frac{\phi(N)}{3}$.*
  (ii) *We have $\Delta = -3$ and N is of Type I.*
(c) *Let $N \geqslant 3$. If $T^\circ(\mathcal{O}, N) \in (\frac{\phi(N)}{3}, \frac{\phi(N)}{2}]$, then $T^\circ(\mathcal{O}, N) = \frac{\phi(N)}{2}$. If $T^\circ(\mathcal{O}, N) \in (\frac{\phi(N)}{2}, \phi(N)]$, then $T^\circ(\mathcal{O}, N) = \phi(N)$.*

Proof. *Step* 0: If $N \in \{1, 2, 3, 4, 6\}$, all the assertions hold vacuously. So, we assume $N \in \mathbb{Z}^+ \setminus \{1, 2, 3, 4, 6\}$.

Step 1: Following [5, Section 7.1] we denote by $\widetilde{T}(\mathcal{O}, N)$ the least size of an orbit of the Cartan subgroup $(\mathcal{O}/N\mathcal{O})^\times$ on any point P order N in $(\mathcal{O}/N\mathcal{O}, +)$. By [5, Section 7.2] the Cartan orbit on such a point is isomorphic to $(\mathcal{O}/I_P)^\times$, where $I_P := \{x \in \mathcal{O} \mid xP = 0\}$. Since P has order N we have an injection of rings $\mathbb{Z}/N\mathbb{Z} \hookrightarrow \mathcal{O}/I_P$, which induces an injection of unit groups $(\mathbb{Z}/N\mathbb{Z})^\times \hookrightarrow (\mathcal{O}/I_P)^\times$, and thus $\phi(N) \mid \widetilde{T}(\mathcal{O}, N)$. Since $T^\circ(\mathcal{O}, N) \in \{T(\mathcal{O}, N),\ 2T(\mathcal{O}, N)\}$, it follows that if $\Delta < -4$

we have

$$\frac{\phi(N)}{2} \mid T(\mathcal{O},N) \mid T^{\circ}(\mathcal{O},N),$$

establishing the result in this case. It remains to consider the cases $\Delta = -4$ and $\Delta = -3$.

Step 2: Suppose $\Delta = -4$. From [5, Theorem 7.2] we have

$$T(\mathcal{O},N) = \frac{\widetilde{T}(\mathcal{O},N)}{4}.$$

- Suppose $\mathrm{ord}_2(N) \leqslant 1$ and that N is not divisible by any prime $\ell \equiv 3 \pmod 4$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) = \phi(N)$. Since $N = 5$ or $N \geqslant 7$, it follows that N is divisible by a prime $\ell \equiv 1 \pmod 4$, and then by [6, Theorems 6.2, 7.1] we have $T^{\circ}(\mathcal{O},N) = 2T(\mathcal{O},N) = \frac{\phi(N)}{2}$.
- Suppose $\mathrm{ord}_2(N) = 2$ and that N is not divisible by any prime $\ell \equiv 3 \pmod 4$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) = 2\phi(N)$. Again N must be divisible by a prime $\ell \equiv 1 \pmod 4$, and as above this gives $T^{\circ}(\mathcal{O},N) = 2T(\mathcal{O},N) = \frac{\widetilde{T}(\mathcal{O},N)}{2} = \phi(N)$.
- Suppose $\mathrm{ord}_2(N) = t \geqslant 3$. Then by [5, Theorem 7.2] we have

$$\widetilde{T}(\mathcal{O},N) \geqslant 2^{t-1}\phi(N),\ T(\mathcal{O},N) = 2^{t-3}\phi(N),\ T^{\circ}(\mathcal{O},N) \geqslant T(\mathcal{O},N) = 2^{t-3}\phi(N) \geqslant \phi(N).$$

- Suppose N is divisible by some prime $\ell \equiv 3 \pmod 4$. Then [5, Theorem 7.2] gives

$$\widetilde{T}(\mathcal{O},N) \geqslant (\ell+1)\phi(N) \geqslant 4\phi(N),$$

so

$$T^{\circ}(\mathcal{O},N) \geqslant T(\mathcal{O},N) = \frac{\widetilde{T}(\mathcal{O},N)}{4} \geqslant \phi(N).$$

Step 3: Suppose $\Delta = -3$. From [5, Theorem 7.2] we have

$$T(\mathcal{O},N) = \frac{\widetilde{T}(\mathcal{O},N)}{6}.$$

- Suppose $\mathrm{ord}_3(N) \leqslant 1$ and that N is not divisible by any prime $\ell \equiv 2 \pmod 3$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) = \phi(N)$. Since $N \geqslant 7$, N must then be divisible by a prime $\ell \equiv 1 \pmod 3$, and then by [6, Theorems 6.2, 7.1] we have $T^{\circ}(\mathcal{O},N) = 2T(\mathcal{O},N) = \frac{\phi(N)}{3}$.
- Suppose $\mathrm{ord}_3(N) = 2$ and that N is not divisible by any prime $\ell \equiv 2 \pmod 3$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) = 3\phi(N)$. If $N = 9$ then [6, Theorem 6.6] gives $T^{\circ}(\mathcal{O},9) = 3 = \frac{\phi(9)}{2}$. If $N \geqslant 10$ we have that N is divisible by a prime $\ell \equiv 1 \pmod 3$, and then by [6, Theorems 6.2, 7.1] we have $T^{\circ}(\mathcal{O},N) = 2T(\mathcal{O},N) = \phi(N)$.
- Suppose $\mathrm{ord}_3(N) = b \geqslant 3$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) \geqslant 3^{b-1}\phi(N) \geqslant 9\phi(N)$, so $T^{\circ}(\mathcal{O},N) \geqslant T(\mathcal{O},N) = \frac{\widetilde{T}(\mathcal{O},N)}{6} \geqslant \frac{3}{2}\phi(N)$.
- Suppose N is divisible by some prime $\ell \equiv 2 \pmod 3$. Then [5, Theorem 7.2] gives $\widetilde{T}(\mathcal{O},N) \geqslant (\ell+1)\phi(N)$, so $T^{\circ}(\mathcal{O},N) \geqslant \frac{\ell+1}{6}\phi(N)$. Thus, if $\ell > 2$ we get $T^{\circ}(\mathcal{O},N) \geqslant \phi(N)$, so suppose that 2 is the only prime divisor of N that is congruent to 2 modulo 3, in which case we have $T^{\circ}(\mathcal{O},N) \geqslant \frac{\phi(N)}{2}$. However, since $N \geqslant 10$, either N is divisible by some prime $\ell \equiv 1 \pmod 3$

or by 9. In the former case we get $T^\circ(\mathcal{O},N) = 2T(\mathcal{O},N)$, so $T^\circ(\mathcal{O},N) \geqslant \phi(N)$. In the latter case we get $\widetilde{T}(\mathcal{O},N) \geqslant 9\phi(N)$, so $T^\circ(\mathcal{O},N) \geqslant T(\mathcal{O},N) \geqslant \frac{3}{2}\phi(N)$. □

Theorem 3.10. *Let Δ be an imaginary quadratic discriminant, and let $N \in \mathbb{Z}^+ \setminus \{1,2,3,4,6\}$. The following are equivalent:*

(i) *we have $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$;*
(ii) *either ($\Delta = -4$ and N is of Type II) or*

$$(\Delta,N) \in \{(-3,9),\ (-7,7),\ (-7,14),\ (-11,11),\ (-19,19),\ (-27,9),$$
$$(-27,27),\ (-28,7),\ (-28,14),\ (-43,43),\ (-67,67),\ (-163,163)\}.$$

Proof. Step 1: Suppose $\Delta \in \{-4,-3\}$. Then the result follows from the analysis given in the proof of Theorem 3.9. This analysis also shows $d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$ if N is of Type II.

Step 2: Suppose $\Delta < -4$, and let $\mathcal{O} = \mathcal{O}(\Delta)$. Then

$$d_{\Delta,\mathrm{CM}}(X_1(N)) = \#\operatorname{Pic}\mathcal{O} \cdot T^\circ(\mathcal{O},N) \geqslant \#\operatorname{Pic}\mathcal{O} \cdot \frac{\phi(N)}{2}.$$

Equality holds iff we have $\mathcal{R}^\circ(\Delta) = \mathbb{Q}$, $\widetilde{T}(\mathcal{O},N) = \phi(N)$ and $T^\circ(\mathcal{O},N) = T(\mathcal{O},N)$.

So suppose $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$. Then the class number one condition gives

$$\Delta \in \{-7,-8,-11,-12,-16,-19,-27,-28,-43,-67,-163\},$$

and by [5, Theorem 6.15] the Cartan orbit condition $\widetilde{T}(\mathcal{O},N) = \phi(N)$ holds if and only if Δ is a square in $\mathbb{Z}/4N\mathbb{Z}$. This implies: (i) if $\ell \mid N$ then $(\frac{\Delta}{\ell}) \neq -1$ and (ii) if $\ell \mid \Delta$ and $\ell \nmid \mathfrak{f}$ then $\operatorname{ord}_\ell(N) \leqslant 1$. Moreover, if $\ell^b > 2$, $\ell^b \mid N$ and $(\frac{\Delta}{\ell}) = 1$, then by [6, Theorems 6.2, 7.1] we have $T^\circ(\mathcal{O},N) = 2T(\mathcal{O},N) \geqslant \phi(N)$.

Step 3: Suppose $\Delta \in \{-7,-8,-11,-19,-43,-67,-163\}$ is a fundamental discriminant. It follows from the above analysis that if $T^\circ(\mathcal{O},N) = 1$ and $(\frac{\Delta}{2}) \neq 1$ then N is a squarefree divisor of Δ, while if $(\frac{\Delta}{2}) = 1$ then either N or $\frac{N}{2}$ is a squarefree divisor of Δ. Conversely, by [39, Corollary 4.2], when these conditions on N hold there is an $\mathcal{O}$-CM elliptic curve $E_{/\mathbb{Q}}$ admitting a $\mathbb{Q}$-rational cyclic N-isogeny and thus $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$. This gives rise to the following pairs (Δ,N) with $\Delta < -4$, $N \geqslant 7$ and $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$:

$$(-7,7),\ (-7,14),\ (-11,11),\ (-19,19),\ (-43,43),\ (-67,67),\ (-163,163).$$

Step 4:

- Suppose $\Delta = -12$. The analysis of Step 2, together with the fact that Δ is not a square modulo 32, shows that if $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$ then N is of the form 2^a or $2^a \cdot 3$ for some $a \leqslant 2$. Since $N \notin \{1,2,3,4,6\}$, we need only consider $N = 12$. But $\frac{\phi(12)}{2} = 2 < 4 = d_{\mathrm{CM}}(X_1(12))$ as follows from [27] (or already from [8, Theorem 1.4]).

- Suppose $\Delta = -16$. The analysis of Step 2 shows that if $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$ then N is of the form 2^b for some $b \geqslant 3$. In the notation of [6, Proposition 6.4] we have $m = 2$ and $M = 3$, so by [6, Theorem 6.5] we have $T^\circ(\mathcal{O}, 2^b) = 2T(\mathcal{O}, 2^b) \geqslant \phi(N)$.
- Suppose $\Delta = -27$. The analysis of Step 2 shows that if $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$ then N is of the form 3^b for some $b \geqslant 2$. If $b \geqslant 4$ then -27 is not a square modulo $4 \cdot 3^b$. By [5, Theorem 7.2; 6, Theorem 6.5] we have $T^\circ(\mathcal{O}, 9) = \frac{\phi(9)}{2}$ and $T^\circ(\mathcal{O}, 27) = \frac{\phi(27)}{2}$, giving rise to the following pairs with $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$:

$$(-27, 9),\ (-27, 27).$$

- Suppose $\Delta = -28$. The analysis of Step 2 shows that if $d_{\Delta,\mathrm{CM}}(X_1(n)) = \frac{\phi(N)}{2}$ then N is of the form 7^b or $2 \cdot 7^b$. Moreover, if -28 is a square modulo 7^b then $b = 1$. By [39, Corollary 4.2] there are $\mathcal{O}$-CM elliptic curves admitting $\mathbb{Q}$-rational cyclic N-isogenies when $N = 7$ or 14, giving rise to the following pairs with $d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$:

$$(-28, 7),\ (-28, 14).$$ □

Corollary 3.11. *Let $N \geqslant 7$.*

(a) *We have $d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{3}$ if and only if N is of Type I.*
(b) *We have $d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}$ if and only if*
  (i) *N is* not *of Type I, and*
  (ii) *N is of Type II or $N \in \{9, 11, 14, 27\}$.*

Remark 3.12. In [16, Theorem 1] it is shown that there is a prime number ℓ_0 such that for all primes $\ell > \ell_0$ and all discriminants $\Delta < 0$, we have:

- $d_{\Delta,\mathrm{CM}}(X_1(\ell)) \geqslant \frac{\ell-1}{3}$, with equality holding if and only if $\Delta = -3$ and $\ell \equiv 1 \pmod 3$;
- If $d_{\Delta,\mathrm{CM}}(X_1(\ell)) \in (\frac{\ell-1}{3}, \frac{\ell-1}{2}]$ then $\Delta = -4$, $\ell \equiv 1 \pmod 4$ and $d_{\Delta,\mathrm{CM}}(X_1(\ell)) = \frac{\ell-1}{2}$.

The proof given therein uses a CM analogue of Serre's open image theorem given by Serre himself in [53]: this *does not* lead to an explicit value for ℓ_0. On the other hand we now have the quantitative version of Serre's result due to Stevenhagen and Bourdon–Clark [57; 5, Theorem 1.4]. The results of this section show that the optimal value of ℓ_0 is 163 and generalize this work from primes to all positive integers.

4 | PRELIMINARY RESULTS

4.1 | Results used

The elliptic curve $E_{/\mathbb{Q}}: y^2 = x^3 - x$ has CM by the order of discriminant -4 and $(\mathbb{Z}/2\mathbb{Z})^2 \hookrightarrow E(\mathbb{Q})$, showing that

$$d_{\mathrm{CM}}(X(1)) = d_{\mathrm{CM}}(X_0(2)) = d_{\mathrm{CM}}(X_1(2)) = d_{\mathrm{CM}}(X(2)) = 1.$$

Thus, in our study of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(M,N))$, we may assume $N \geqslant 3$. We remind the reader that $\psi(N) = N\prod_{p|N}(1+1/p)$.

Lemma 4.1. *Let $M, N \in \mathbb{Z}^+$ with $M \mid N$ and $N \geqslant 3$. Then:*

(a) $\deg(X_0(N) \to X(1)) = \psi(N)$;
(b) $\deg(X_1(N) \to X(1)) = \frac{\phi(N)\psi(N)}{2}$;
(c) $\deg(X(N) \to X(1)) = \frac{N\phi(N)^2\psi(N)}{2}$;
(d) $\deg(X_1(M,N) \to X(1)) = \frac{M\phi(M)\phi(N)\psi(N)}{2}$.

We defer the proof of Lemma 4.1 to Section 7.2. For now let us emphasize that we are giving the degrees of the maps viewed as curves over $\mathbb{Q}$. In parts (a) and (b) this distinction does not matter, as the subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$ are rational in the sense of Section 7.1. But it does matter in parts (c) and (d): as follows from the discussion in Section 7.1, $\deg(X(N) \to X(1))$ (respectively $\deg(X_1(M,N) \to X(1))$) is $\phi(N)$ times (respectively $\phi(M)$ times) the degree of the corresponding covering of compact Riemann surfaces.

Theorem 4.2 (Silverberg, Bourdon–Clark). *Let $\mathcal{O}$ be an order in an imaginary quadratic field K, and let E be an $\mathcal{O}$-CM elliptic curve defined over a number field $F \supset K$. If $E(F)$ has a point of order $N \in \mathbb{Z}^+$ then*

$$\phi(N) \mid \frac{\#\mathcal{O}^\times}{2}\frac{[F:\mathbb{Q}]}{\#\operatorname{Pic}\mathcal{O}}.$$

Proof. This is [5, Theorem 6.2]. □

Theorem 4.3 (Clark–Pollack). *Let K be an imaginary quadratic field, let $F \supset K$ be a field extension, and let $E_{/F}$ be an elliptic curve with CM by an order in K. Suppose that for $a, b \in \mathbb{Z}^+$ we have an injection $\mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/ab\mathbb{Z} \hookrightarrow E(F)$. Then $[F(E[ab]) : F] \leqslant b$.*

Proof. This is [19, Theorem 7]. □

4.2 | Some key inequalities

The following important result gives inequalities among various $d_{\mathrm{CM}}(X)$'s coming from the tower structure of modular curves $X_1(M,N) \to X_1(N) \to X_0(N)$.

Lemma 4.4. *Let $M, N \in \mathbb{Z}^+$ with $M \mid N$ and $N \geqslant 3$. Then:*

(a) *we have $\frac{\phi(N)}{6}d_{\mathrm{CM}}(X_0(N)) \leqslant d_{\mathrm{CM}}(X_1(N)) \leqslant \frac{\phi(N)}{2}d_{\mathrm{CM}}(X_0(N))$;*
(b) *we have $d_{\mathrm{CM}}(X_1(M,N)) \leqslant 2M d_{\mathrm{CM}}(X_1(N))$.*

Proof.

(a) For $N \in \{3,4,6\}$ we have $d_{\mathrm{CM}}(X_0(N)) = d_{\mathrm{CM}}(X_1(N)) = 1$, from which part (a) follows. If $N \in \mathbb{Z}^+ \setminus \{1,2,3,4,6\}$ the result follows from Theorem 3.7.

(b) Every non-cuspidal closed point p on $X_1(N)$ with residue field $\mathbb{Q}(p)$ is induced by a pair $(E,P)_{/\mathbb{Q}(p)}$, where $E_{/\mathbb{Q}(p)}$ is an elliptic curve and P is a $\mathbb{Q}(p)$-rational point of order N [22, p. 274, Proposition VI.3.2]. So if $d_{\mathrm{CM}}(X_1(N)) = d$ there is a number field F of degree d and a CM elliptic curve $E_{/F}$ such that $E(F)$ has a point of order N. Let K be the endomorphism algebra of E. Applying Theorem 4.3 to $E_{/FK}$ shows that there is an extension L/F of degree at most $2M$ such that $\mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \hookrightarrow E(L)[\mathrm{tors}]$. □

Remark 4.5. Lemma 4.4(a) implies, in particular, that

$$d_{\mathrm{CM}}(X_1(N)) \asymp \phi(N) d_{\mathrm{CM}}(X_0(N)).$$

This tight relationship between $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ often allows one to deduce theorems about $d_{\mathrm{CM}}(X_1(N))$ from corresponding statements about $d_{\mathrm{CM}}(X_0(N))$. For example, (recalling $\frac{N}{\log\log N} \ll \phi(N) \ll N$) Theorem 1.5 follows immediately from Theorem 1.3, while Theorem 1.7(b) is a consequence of part (a) of the same result.

4.3 | Estimates for $d_{\Delta,\mathrm{CM}}(X_0(N))$

Let $\Delta < 0$ be a quadratic discriminant. For $N \in \mathbb{Z}^+$, write $N = N_1N_2N_3$, where:

- N_1 is divisible precisely by primes p with $(\frac{\Delta}{p}) = 1$;
- N_2 is divisible precisely by primes p with $(\frac{\Delta}{p}) = 0$;
- N_3 is divisible precisely by primes p with $(\frac{\Delta}{p}) = -1$.

Proposition 4.6. *If N_2 is squarefree, then $d_{\Delta,\mathrm{CM}}(X_0(N)) \asymp h_\Delta \psi(N_3)$.*

Proof. Write $N = \prod_{i=1}^r \ell_i^{a_i}$ as a product of distinct prime powers. By Lemma 4.4(a) and [5, Theorem 7.2], we have

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \asymp h_\Delta \prod_{i=1}^r \frac{\widetilde{T}(\mathcal{O}(\Delta), \ell_i^{a_i})}{\phi(\ell_i^{a_i})}, \tag{6}$$

where for a prime power ℓ^a, the quantity $\widetilde{T}(\mathcal{O}(\Delta), \ell^a)$ is computed in [6, Theorem 7.2d)], and the implied constants are absolute. In particular:

- if $(\frac{\Delta}{\ell}) = 1$, then $\widetilde{T}(\mathcal{O}(\Delta), \ell^a) = \phi(\ell^a)$;
- if $(\frac{\Delta}{\ell}) = -1$, then $\widetilde{T}(\mathcal{O}(\Delta), \ell^a) = \phi(\ell^a)\psi(\ell^a)$;
- if $(\frac{\Delta}{\ell}) = 0$, then $\widetilde{T}(\mathcal{O}(\Delta), \ell) = \phi(\ell)$.

Since N_2 is squarefree, this accounts for $\widetilde{T}(\mathcal{O}(\Delta), \ell_i^{a_i})/\phi(\ell_i^{a_i})$ for all i and the proposition follows. □

Suppose we are in the case where $(\frac{\Delta}{p}) = 1$ for all $p \mid N$. The proof of Proposition 4.6 then shows $d_{\Delta,\mathrm{CM}}(X_0(N)) \asymp h_\Delta$. The following proposition makes the upper bound explicit.

Proposition 4.7. *Let $N \in \mathbb{Z}^+$, and let Δ be a negative discriminant with the property that*

$$\left(\frac{\Delta}{p}\right) = 1 \qquad \text{for all primes } p \mid N.$$

Then $d_{\Delta,\mathrm{CM}}(X_0(N)) \leqslant 2h_\Delta$.

Proof. Let $K = \mathbb{Q}(\sqrt{\Delta})$ and let $K^{(1)}$ be its Hilbert class field, so that $[K^{(1)} : K] = h_K := \#\operatorname{Pic}\mathcal{O}_K$. There is a (necessarily proper) ideal I of the ring of integers $\mathcal{O}_K$ of K such that $\mathcal{O}_K/I \cong \mathbb{Z}/N\mathbb{Z}$, so if $E_{/K^{(1)}}$ is any $\mathcal{O}_K$-CM elliptic curve then $E \to E/E[I]$ is a cyclic N-isogeny defined over $K^{(1)}$. Thus, we have

$$d_{\mathrm{CM}}(X_0(N)) \leqslant d_{\Delta,\mathrm{CM}}(X_0(N)) \leqslant 2[K^{(1)} : \mathbb{Q}] = 2h_\Delta. \qquad \square$$

Remark 4.8. That one has an $\mathcal{O}_K$-CM point on $X_0(N)$ defined over $K^{(1)}$ when every prime divisor of N splits in K is a basic point in the theory of Heegner points on modular elliptic curves (often called the 'Heegner hypothesis' on N and K). The observation that when the Heegner hypothesis is satisfied, the rational isogeny of small degree leads to a CM point of order N of small degree seems to appear for the first time in work of Sutherland [58, Section 4].

5 | ANALYTIC RESULTS

5.1 | Lower order of $d_{\mathrm{CM}}(X_0(N))$: Proof of Theorem 1.2

The complete list of $N \in \mathbb{Z}^+$ such that $d_{\mathrm{CM}}(X_0(N)) = 1$ is given in Table 1: the largest such N is 163.

This is not a new result, but we are not sure of the proper attribution. These N appear as part of a list of known non-cuspidal $\mathbb{Q}$-rational points on $X_0(N)$ in Mazur's work [45], albeit without a proof that it consists of all CM points. The classification of all non-cuspidal $\mathbb{Q}$-rational points on $X_0(N)$ was obtained for prime N in Mazur's paper, and the composite case was done by various people over the next several years, ending in a work of Kenku [37]. When taken together, these papers provide a proof of Theorem 1.2. However, this risks a dependence on significantly more difficult results.

Here is a better way: for each $\Delta < 0$, the set of $N \in \mathbb{Z}^+$ such that $d_{\Delta,\mathrm{CM}}(X_0(N)) = h_\Delta$ is finite (and known). For $\Delta_K < -4$ this is a result of Kwon [39, Corollary 4.2]; for $\Delta_K \in \{-4, -3\}$ it is due to Bourdon–Clark [6, Corollary 5.11]. Table 1 is an immediate consequence.

Let K be a quadratic field, and let $\mathcal{P}_K$ be the set of primes that split in K. Let P be a set of prime numbers whose symmetric difference with $\mathcal{P}_K$ is finite. Then P is a Chebotarev set in the sense of Serre, and he has shown [54, Theorem 2.8] that if $\mathcal{N}_P$ is the set of positive integers N all of whose prime divisors lie in P, then as $X \to \infty$,

$$\#\{1 \leqslant N \leqslant X \mid N \in \mathcal{N}_P\} \sim c_P \frac{X}{\sqrt{\log X}},$$

for a certain positive constant c_P. (Actually this case of Serre's results also follows from earlier work of Landau [42].)

Let $\Delta < 0$ be a discriminant and put $K = \mathbb{Q}(\sqrt{\Delta})$. By Proposition 4.7, if $N \in \mathbb{Z}^+$ is such that $(\frac{\Delta}{p}) = 1$ for all $p \mid N$ then $d_{\mathrm{CM}}(X_0(N)) \leqslant d_{\Delta,\mathrm{CM}}(X_0(N)) \leqslant 2h_\Delta$. By the above result of Serre, the number of N up to X satisfying this 'Heegner hypothesis' is $\asymp \frac{X}{\sqrt{\log X}}$. Taking Δ to be any one of the 13 class number 1 discriminants, we see

$$\{1 \leqslant N \leqslant X \mid d_{\mathrm{CM}}(X_0(N)) \leqslant 2\} \gg \frac{X}{\sqrt{\log X}},$$

which establishes the lower bound in Theorem 1.2(b)).

The idea for the upper bound (as well as several arguments to come) is that this 'Heegner hypothesis' is also close to being necessary. If $N \geqslant 5$, then $d_{-3,\mathrm{CM}}(X_0(N)) \leqslant 2$ if and only if N is of Type I and $d_{-4,\mathrm{CM}}(X_0(N)) \leqslant 2$ if and only if N is of Type II, so some amount of divisibility at ramified primes is permitted, but this only changes the implied constant.

Here are the details: fix $\Delta < 0$ and $C \geqslant 2$. First of all we have $d_{\Delta,\mathrm{CM}}(X_0(N)) \geqslant h_\Delta$, so if $d_{\mathrm{CM}}(X_0(N)) \leqslant C$ then $d_{\mathrm{CM}}(X_0(N)) = d_{\Delta,\mathrm{CM}}(X_0(N))$ for some $\Delta < 0$ with $h_\Delta \leqslant C$. There are only finitely many such Δ, so we may work with a fixed $\Delta < 0$. Suppose ℓ is a prime such that $(\frac{\Delta}{\ell}) = -1$. Then (6) and the second bulleted point below it gives

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \gg \psi(\ell)h_\Delta \geqslant \ell h_\Delta,$$

with an absolute implied constant. Thus, if $d_{\Delta,\mathrm{CM}}(X_0(N)) \leqslant C$ then N can only be divisible by finitely many primes that are inert in $\mathbb{Q}(\sqrt{\Delta})$, so by Serre's result we have

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \ll_C \frac{X}{\sqrt{\log X}},$$

completing the proof of Theorem 1.2(b).

5.2 | Upper order of $d_{\mathrm{CM}}(X_0(N))$: Proof of Theorem 1.3

We begin by establishing a general (GRH-conditional) upper bound on $d_{\mathrm{CM}}(X_0(N))$, expressed in terms of $\log N$ and the number r of distinct prime factors of N.

Theorem 5.1. *Assume GRH. Let $N \in \mathbb{Z}^+$ have $\omega(N) = r$ distinct prime divisors. For all $\epsilon > 0$ we have*

$$d_{\mathrm{CM}}(X_0(N)) \ll 2^r \log N \cdot (\log\log(2^r \log N))^{O(1)}.$$

The argument below was inspired by work of Schinzel on pseudosquares [52].

Proof. Since $d_{\mathrm{CM}}(X_0(N)) \leqslant d_{\mathrm{CM}}(X_0(2N))$, we can (and will) assume $2 \mid N$. (Note that the order of magnitude of our upper bound does not change under replacing N by $2N$.) We will find a small negative discriminant Δ satisfying $(\frac{\Delta}{p}) = 1$ for all $p \mid N$, and then use Proposition 4.7 and class number estimates to bound $d_{\mathrm{CM}}(X_0(N))$. Our Δ will have the form $\Delta = -\ell$, where $\ell \equiv 3 \pmod 4$ is prime.

Let $p_1, \dots, p_r$ be the distinct primes dividing N, where $p_1 = 2$. To ensure $(\frac{\Delta}{p}) = 1$ for all $p \mid N$, and $\ell \equiv 3 \pmod 4$, we choose the prime ℓ to satisfy certain Chebotarev conditions. Observe:

- the condition $\ell \equiv 3 \pmod 4$ on ℓ holds if and only if ℓ is unramified in $\mathbb{Q}(\sqrt{-1})/\mathbb{Q}$ and the Frobenius element at ℓ in this abelian extension is the nontrivial element of $\mathrm{Aut}(\mathbb{Q}(\sqrt{-1})/\mathbb{Q})$;
- the condition $(\frac{-\ell}{2}) = 1$ holds if and only if $\ell \equiv 7 \pmod 8$ if and only if ℓ is unramified in $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ and the Frobenius element at ℓ is trivial;
- for each $i = 2, 3, \dots, r$, by quadratic reciprocity we have $(\frac{-\ell}{p_i}) = 1$ if and only if $(\frac{p_i}{\ell}) = 1$ if and only if ℓ is unramified in $\mathbb{Q}(\sqrt{p_i})/\mathbb{Q}$ and the Frobenius element at ℓ is trivial.

The classes of -1 and $p_1, \dots, p_r$ are $\mathbb{Z}/2\mathbb{Z}$-linearly independent in $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$. By Kummer theory, the fields $\mathbb{Q}(\sqrt{-1})$, $\mathbb{Q}(\sqrt{p_1}), \dots, \mathbb{Q}(\sqrt{p_r})$ are linearly disjoint over $\mathbb{Q}$. Let L be their compositum, an abelian number field. Let Δ_L be its discriminant. We may represent the automorphism group $\mathrm{Aut}(L/\mathbb{Q})$ as $G := \prod_{i=0}^{r} \{\pm 1\}$, where for $0 \leqslant i \leqslant r$, if H_i is the set of vectors with ith coordinate equal to 1, then $L^{H_0} = \mathbb{Q}(\sqrt{-1})$ and for $1 \leqslant i \leqslant r$, $L^{H_i} = \mathbb{Q}(\sqrt{p_i})$. The desired condition on ℓ is that it is unramified in L and its Frobenius element is $(-1, 1, \dots, 1) \in G$. As we are assuming GRH, an effective version of the Chebotarev Density Theorem due to Lagarias–Odlyzko [40] tells us that there is such a prime number ℓ with $\ell \ll \log^2 |\Delta_L|$.

We now find an upper bound on $|\Delta_L|$. For this we use the bijective correspondence between finite abelian extensions of $\mathbb{Q}$ and finite groups of Dirichlet characters (see, for example, [61, Chapter 3]). For $d \in \mathbb{Q}^\times \setminus \mathbb{Q}^{\times 2}$, let χ_d be the Dirichlet character corresponding to the quadratic field $\mathbb{Q}(\sqrt{d})$, let χ_1 be the trivial character, and let $\mathbb{X}$ be the group of Dirichlet characters generated by χ_{-1} and χ_{p_i} for $1 \leqslant i \leqslant r$: thus $\mathbb{X} \cong G$, so $\#\mathbb{X} = 2^{r+1}$. More explicitly we have

$$\mathbb{X} = \{\chi_d \mid d \in \mathcal{D}\},$$

where $\mathcal{D}$ is the set of (positive and negative) divisors of $p_1 \cdots p_r$. For $d \in \mathcal{D}$, let $\mathfrak{f}_d$ be the conductor of χ_d. Recall the conductor discriminant formula [61, Theorem 3.11]:

$$|\Delta_L| = \prod_{d \in \mathcal{D}} \mathfrak{f}_d.$$

We have

$$\mathfrak{f}_d = \begin{cases} |d| & \text{if } d \equiv 1 \pmod 4, \\ 4|d| & \text{if } d \equiv 2, 3 \pmod 4. \end{cases}$$

Since $d \in \mathcal{D} \Rightarrow -d \in \mathcal{D}$, precisely half of the odd elements of $\mathcal{D}$ are $1 \pmod 4$ and none of the even elements of $\mathcal{D}$ are $1 \pmod 4$, and we have

$$|\Delta_L| = 4^{3 \cdot 2^{r-1}} \left(\prod_{1 \leqslant d \mid p_1 \cdots p_r} d \right)^2 = 2^{6 \cdot 2^{r-1}} (p_1 \cdots p_r)^{2^r}.$$

Since $p_1 \cdots p_r$ is a divisor of N, we see

$$\log |\Delta_L| \ll 2^{r-1} + 2^r \log p_1 \cdots p_r \ll 2^r \log N$$

and so we can choose

$$\ell \ll 4^r \log^2 N.$$

From Proposition 4.7, $d_{\mathrm{CM}}(X_0(N)) \leqslant d_{\Delta,\mathrm{CM}}(X_0(N)) \leqslant 2h_{-\ell}$. The claimed estimate for $d_{\mathrm{CM}}(X_0(N))$ now follows from Lemma 2.3. □

The number r of distinct prime factors of N satisfies $r \leqslant (1+o(1))\frac{\log N}{\log\log N}$ as $N \to \infty$ (cf. the discussion on [31, p. 471]). Plugging this into Theorem 5.1 gives

$$d_{\mathrm{CM}}(X_0(N)) \leqslant \exp\left((\ln 2 + o(1))\frac{\log N}{\log\log N}\right),$$

which is part (b) of Theorem 1.3.

We now turn to the proof of (a). The following useful proposition is an easy consequence of Proposition 4.6 and Lemma 2.3.

Proposition 5.2 (Conditional on GRH). *For each squarefree positive integer N and each negative discriminant Δ,*

$$d_{\Delta,\mathrm{CM}}(X_0(N)) = (\log\log|\Delta N|)^{O(1)} \cdot |\Delta|^{1/2} \prod_{\substack{p\mid N \\ \left(\frac{\Delta}{p}\right)=-1}} p. \tag{7}$$

Proof. Write $N = N_1N_2N_3$ where N_1, N_2, N_3 have the same meanings as in the discussion immediately preceding Proposition 4.6. So $N_3 = \prod_{p\mid N,\,(\frac{\Delta}{p})=-1} p$ (since N is squarefree). Proposition 4.6 gives

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \asymp h_\Delta \psi(N_3) = h_\Delta N_3 \cdot (\log\log N)^{O(1)}.$$

(In the last step, we used that $1 \leqslant \psi(N_3)/N_3 \leqslant N_3/\phi(N_3) \ll \log\log N_3$.) To finish off, we use the estimate for h_Δ appearing in Lemma 2.3. □

Proof of part (a) of Theorem 1.3. We show (non-constructively) the existence of a sequence N tending to infinity along which

$$d_{\mathrm{CM}}(X_0(N)) \geqslant \exp\left(\left(\frac{1}{4}\ln 2 + o(1)\right)\frac{\log N}{\log\log N}\right).$$

Let $y = (\log x \cdot \log\log x)^2$, let $\mathcal{P}$ be the set of primes not exceeding y, and let Ω denote the collection of all K-element subsets of $\mathcal{P}$, where

$$K = \left\lfloor \frac{\log x}{\log y} \right\rfloor.$$

We will consider Ω as a finite probability space with the uniform measure.

We associate to each $S \in \Omega$ the squarefree integer $N_S := \prod_{p \in S} p$. Note $N_S \leqslant x$. Put

$$L := \exp\left(\frac{1}{2}\ln 2 \cdot \frac{\log x}{\log\log x}\right).$$

For every negative discriminant Δ with $|\Delta| \leqslant L$, we introduce the random variable

$$D_\Delta(S) = |\Delta|^{1/2} \prod_{\substack{p \mid N_S \\ \left(\frac{\Delta}{p}\right)=-1}} p.$$

Fix $\epsilon \in (0, \frac{1}{10})$, and take $c = \frac{1}{2} - \epsilon$. We estimate, for a given Δ, the probability that $D_\Delta(S) \leqslant L^c$. Clearly,

$$\begin{aligned}\Pr(D_\Delta(S) \leqslant L^c) &\leqslant \mathbb{E}[L^c \cdot D_\Delta(S)^{-1}] \\ &= \frac{L^c}{|\Delta|^{1/2}} \cdot \frac{1}{\#\Omega} \sum_S \prod_{\substack{p \in S \\ \left(\frac{\Delta}{p}\right)=-1}} p^{-1}.\end{aligned}$$

Observe

$$\sum_S \prod_{\substack{p \in S \\ \left(\frac{\Delta}{p}\right)=-1}} p^{-1} \leqslant \frac{1}{K!}\left(\sum_{\substack{p \in \mathcal{P} \\ \left(\frac{\Delta}{p}\right)=-1}} p^{-1} + \sum_{\substack{p \in \mathcal{P} \\ \left(\frac{\Delta}{p}\right)\neq -1}} 1\right)^K.$$

By a theorem of Mertens, $\sum_{p \in \mathcal{P},\, (\frac{\Delta}{p})=-1} p^{-1} \leqslant \sum_{p \in \mathcal{P}} p^{-1} \ll \log\log\log x$. Moreover,

$$\sum_{\substack{p \in \mathcal{P} \\ \left(\frac{\Delta}{p}\right)\neq -1}} 1 = \sum_{p \in \mathcal{P}} \frac{1}{2}\left(1 + \left(\frac{\Delta}{p}\right)\right) + \sum_{\substack{p \in \mathcal{P} \\ p \mid \Delta}} \frac{1}{2} = \frac{1}{2}\sum_{p \in \mathcal{P}} 1 + \frac{1}{2}\sum_{p \in \mathcal{P}} \left(\frac{\Delta}{p}\right) + \frac{1}{2}\sum_{\substack{p \in \mathcal{P} \\ p \mid \Delta}} 1.$$

Put $\Pi = \#\mathcal{P}$, so that $\frac{1}{2}\sum_{p \in \mathcal{P}} 1 = \frac{1}{2}\Pi = (\frac{1}{2} + o(1))y/\log y$, as $x \to \infty$. Under GRH we have

$$\frac{1}{2}\sum_{p \in \mathcal{P}} \left(\frac{\Delta}{p}\right) = O(y^{1/2}\log(|\Delta| y)),$$

which is $o(\Pi)$ for our choice of y. (See, for example, [47, p. 425, Theorem 13.7].) Also, $\sum_{p \in \mathcal{P},\, p \mid \Delta} 1 \ll \log|\Delta| \ll \log x$, which is again $o(\Pi)$. Collecting our estimates, we deduce

$$\Pr(D_\Delta(S) \leqslant L^c) \leqslant \frac{L^c}{|\Delta|^{1/2}} \cdot \frac{1}{\#\Omega} \frac{((\frac{1}{2} + o(1))\Pi)^K}{K!}.$$

With $\Pi = \#\mathcal{P}$, we have $\#\Omega = \binom{\Pi}{K} = (1+o(1)) \cdot \frac{\Pi^K}{K!}$, and $K = (\frac{1}{2} + o(1))\frac{\log x}{\log\log x}$, so that

$$\Pr(D_\Delta(S) \leqslant L^c) \leqslant L^{c-1}|\Delta|^{-1/2} = L^{-1/2-\epsilon+o(1)}|\Delta|^{-1/2}, \tag{8}$$

as $x \to \infty$, uniformly for negative discriminants Δ with $|\Delta| \leqslant L$.

The sum of $|\Delta|^{-1/2}$ on Δ with $|\Delta| \leqslant L$ is $O(L^{1/2})$. It thus follows from (8) that for large enough x, we can choose S with $D_\Delta(S) > L^c$ for all negative discriminants Δ with $|\Delta| \leqslant L$. Let $N = N_S$. We claim that (as long as x is large enough)

$$d_{\mathrm{CM}}(X_0(N)) \geqslant \exp\left(\left(\frac{1}{4}\ln 2 - \epsilon\right)\frac{\log x}{\log\log x}\right). \tag{9}$$

Note that the right-hand side exceeds $\exp((\frac{1}{4}\ln 2 - \epsilon)\frac{\log N}{\log\log N})$, since $N \leqslant x$. To prove (9), let Δ be the discriminant minimizing $d_{\Delta,CM}(X_0(N))$. If $|\Delta| > L$, then the appearance of the factor $|\Delta|^{1/2}$ in (7) gives the desired lower bound. Otherwise, our choice of S shows $|\Delta|^{1/2}\prod_{p|N,\,(\frac{\Delta}{p})=-1} p \geqslant L^c$, and again the stated lower bound follows from (7). Since ϵ can be taken arbitrarily small, and x can be taken arbitrarily large, we have our result. □

Remark 5.3. Let $c_\sharp$ denote the infimum of those constants c for which

$$d_{\mathrm{CM}}(X_0(N)) \leqslant \exp\left((c+o(1))\frac{\log N}{\log\log N}\right), \qquad \text{as } N \to \infty.$$

Theorem 1.3 shows $\frac{1}{4}\ln 2 \leqslant c_\sharp \leqslant \ln 2$. We conjecture $c_\sharp = \frac{1}{2}\ln 2$.

Remark 5.4. The upper bound of Theorem 5.1 on $d_{\mathrm{CM}}(X_0(N))$ is highly sensitive to the number r of distinct prime factors of N. So it could be interesting to study the upper order of $d_{\mathrm{CM}}(X_0(N))$ under restrictions on r. The most obvious such restriction is to ask that $N = \ell$ be prime. Theorem 5.1 shows (in somewhat more precise form) $d_{\mathrm{CM}}(X_0(\ell)) \leqslant (\log\ell)^{1+o(1)}$, as $\ell \to \infty$. In the opposite direction, using Linnik's theorem on primes in progressions [43], one can produce a sequence of primes $\ell \equiv 3 \pmod 4$ tending to infinity for which the smallest quadratic nonresidue *mod* ℓ is $\gg \log\ell$. (See [26] or [51] for a similar argument, but without the condition $\ell \equiv 3 \pmod 4$.) From Proposition 4.6 and Lemma 2.2, one can deduce $d_{\mathrm{CM}}(X_0(\ell)) \geqslant (\log\ell)^{1/2+o(1)}$ along this sequence of ℓ. Probably the lower bound, with the exponent $\frac{1}{2}$ on $\log\ell$, reflects the truth in this upper order problem.

5.3 | Lower order of $d_{\mathrm{CM}}(X_1(N))$: Proof of Theorem 1.4

We first produce a sequence of N tending to infinity such that $d_{\mathrm{CM}}(X_1(N)) \ll \frac{N}{\sqrt{\log\log N}}$. Take N of the form $N = \prod_{p\leqslant T,\ p\equiv 1 \pmod 3} p$. By Section 3.3 we have $d_{\mathrm{CM}}(X_0(N)) \leqslant 2$, and so by

Lemma 4.4(a),

$$d_{-3,\mathrm{CM}}(X_1(N)) \asymp \phi(N) = N \prod_{\substack{p \leqslant T \\ p \equiv 1 \pmod 3}} \left(1 - \frac{1}{p}\right) \ll N \exp\left(- \sum_{\substack{p \leqslant T \\ p \equiv 1 \pmod 3}} \frac{1}{p}\right).$$

By the prime number theorem for arithmetic progressions, we have $\log N = (\frac{1}{2} + o(1))T$ as $T \to \infty$, and $\sum_{p \leqslant T,\ p \equiv 1 \pmod 3} \frac{1}{p} = \frac{1}{2} \log\log T + O(1) = \frac{1}{2} \log\log\log N + O(1)$. Thus,

$$d_{-3,\mathrm{CM}}(X_1(N)) \ll N \exp(-\frac{1}{2} \log\log\log N) \ll \frac{N}{\sqrt{\log\log N}}.$$

As $d_{\mathrm{CM}}(X_1(N)) \leqslant d_{-3,\mathrm{CM}}(X_1(N))$, the upper bound half of Theorem 1.4 is proved.

To finish the proof of Theorem 1.4, we show

$$d_{\Delta,\mathrm{CM}}(X_1(N)) \gg \frac{N}{\sqrt{\log\log N}} \tag{10}$$

for all $N \in \mathbb{Z}^+$ and discriminants $\Delta < 0$. Write $N = N_1 N_2 N_3$ as in the setup for Proposition 4.6, and let n_2 be the product of the distinct primes dividing N_2. Then

$$d_{\Delta,\mathrm{CM}}(X_0(N)) \geqslant d_{\Delta,\mathrm{CM}}(X_0(N_1 n_2 N_3)) \gg h_\Delta \psi(N_3),$$

using the estimate of Proposition 4.6 in the second step. By Lemma 4.4(a), we have

$$d_{\Delta,\mathrm{CM}}(X_1(N)) \gg h_\Delta \phi(N) \psi(N_3).$$

In particular, $d_{\Delta,\mathrm{CM}}(X_1(N)) \gg h_\Delta \phi(N) \gg h_\Delta \frac{N}{\log\log N}$. If $\Delta > (\log\log N)^2$, inserting the lower bound for h_Δ from Lemma 2.2 (with $\epsilon = \frac{1}{4}$) gives (10). So we suppose $\Delta \leqslant (\log\log N)^2$. In this case, we use the last display to obtain

$$\begin{aligned} d_{\Delta,\mathrm{CM}}(X_1(N)) &\gg h_\Delta N \prod_{p \mid N} \left(1 - \frac{1}{p}\right) \prod_{\substack{p \mid N \\ \left(\frac{\Delta}{p}\right) = -1}} p \\ &\geqslant h_\Delta N \prod_{\substack{p \mid N \\ \left(\frac{\Delta}{p}\right) \neq -1}} \left(1 - \frac{1}{p}\right) \\ &\gg h_\Delta N \exp\left(- \sum_{\substack{p \mid N \\ \left(\frac{\Delta}{p}\right) \neq -1}} \frac{1}{p}\right). \end{aligned}$$

Continuing,

$$\sum_{\substack{p\mid N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p} \leqslant \sum_{\substack{p\leqslant \log N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p} + \sum_{\substack{p\mid N\\ p>\log N}} \frac{1}{p}$$

$$\leqslant \sum_{\substack{p\leqslant \log N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p} + \frac{\omega(N)}{\log N} \leqslant \sum_{\substack{p\leqslant \log N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p} + O(1).$$

Now

$$\sum_{\substack{p\leqslant \log N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p} \leqslant \sum_{p\mid \Delta} \frac{1}{p} + \frac{1}{2} \sum_{p\leqslant \log N} \left(1+\left(\frac{\Delta}{p}\right)\right)\frac{1}{p}$$

$$\leqslant O(\log\log\log|\Delta|) + \frac{1}{2}\log\log\log N + \frac{1}{2}\sum_{p\leqslant \log N}\left(\frac{\Delta}{p}\right)\frac{1}{p}.$$

By the Siegel–Walfisz theorem and partial summation, we have $\sum_{U<p\leqslant V}(\frac{\Delta}{p})\frac{1}{p} \ll_\epsilon 1$ whenever $V > U > \exp(|\Delta|^\epsilon)$. It follows that the sum on p appearing in the last display is at most $\sum_{p\leqslant \exp(|\Delta|^\epsilon)} \frac{1}{p} + O_\epsilon(1) \leqslant \epsilon \log|\Delta| + O_\epsilon(1)$. Taking $\epsilon = \frac{1}{2}$ and collecting estimates, we find

$$\exp\left(-\sum_{\substack{p\mid N\\ \left(\frac{\Delta}{p}\right)\neq -1}} \frac{1}{p}\right) \gg \frac{1}{|\Delta|^{1/3}\sqrt{\log\log N}},$$

and so

$$d_{\mathrm{CM}}(X_1(N)) \gg h_\Delta \Delta^{-1/3} \cdot \frac{N}{\sqrt{\log\log N}}.$$

As $h_\Delta \gg \Delta^{1/3}$, we once again have (10).

5.4 | Upper order of $d_{\mathrm{CM}}(X_1(N))$: Proof of Theorem 1.5

Theorem 1.5 follows immediately from Theorem 1.3 and Lemma 4.4(a).

5.5 | Upper and lower order of $d_{\mathrm{CM}}(X(N))$

Theorem 1.6 was proved already in the introduction, as a consequence of Theorem 1.1. So, we concentrate on the claims about $X(N)$.

Proposition 5.5. *$d_{\mathrm{CM}}(X(N))$ has lower order $\frac{N^2}{\log\log N}$.*

Proof. The lower bound $d_{\mathrm{CM}}(X(N)) \gg \frac{N^2}{\log\log N}$ follows from taking $M = N$ in Theorem 1.6. For the upper bound, we apply Theorem 3.1 with $\Delta = -3$ and $N = \prod_{p\leqslant T,\ p\equiv 1 \pmod 3} p$. For large T, that theorem gives $d_{-3,\mathrm{CM}}(X(N)) \ll \phi(N)^2$. As at the beginning of the proof of Theorem 1.4, we have $\phi(N) \ll \frac{N}{\sqrt{\log\log N}}$ for this family of N, and so we have

$$d_{\mathrm{CM}}(X(N)) \leqslant d_{-3,\mathrm{CM}}(X(N)) \ll \frac{N^2}{\log\log N},$$

as desired. □

Proposition 5.6. *$d_{\mathrm{CM}}(X(N))$ has upper order N^2.*

Proof. For all $N \in \mathbb{Z}^+$ we have

$$\begin{aligned} d_{\mathrm{CM}}(X(N)) \leqslant d_{-3,\mathrm{CM}}(X(N)) &= \frac{N^2}{3}\prod_{p\mid N}\left(1-\left(\frac{-3}{p}\right)\frac{1}{p}\right)\left(1-\frac{1}{p}\right) \\ &\leqslant \frac{N^2}{3}\prod_{p}\left(1-\frac{1}{p^2}\right) = \frac{2}{\pi^2}N^2. \end{aligned}$$

By Theorem 3.2, as N tends to infinity along prime powers ℓ^a we have $d_{\mathrm{CM}}(X(\ell^a)) \sim \frac{1}{3}\ell^{2a}$. □

5.6 | Typical behavior of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$: Proof of Theorem 1.7

We consider only Theorem 1.7(a), since part (b) follows from (a) via Lemma 4.4.

We first prove the lower bound in Theorem 1.7(a). That is, we consider $N \leqslant x$ for which

$$d_{\mathrm{CM}}(X_0(N)) < 2^{(\frac{1}{2}-\epsilon)\log\log x}$$

and show that these numbers comprise a set of size $o(x)$, as $x \to \infty$. In what follows we restrict ourselves to N satisfying

$$|\omega(N) - \log\log x| < \log\log\log x \cdot \sqrt{\log\log x}.$$

This is permissible since a well-known theorem of Hardy–Ramanujan [30] shows that this inequality on $\omega(N)$ holds for all but $o(x)$ values of $N \leqslant x$.

Let $\Delta < 0$ be the discriminant minimizing $d_{\Delta,\mathrm{CM}}(X_0(N))$. Writing N' for the product of the distinct primes dividing N, we have

$$d_{\Delta,\mathrm{CM}}(X_0(N')) \leqslant d_{\Delta,\mathrm{CM}}(X_0(N)) = d_{\mathrm{CM}}(X_0(N)) < 2^{(\frac{1}{2}-\epsilon)\log\log x}.$$

Using the estimate of (7) for $d_{\Delta,\mathrm{CM}}(N')$, we see it is necessary to have (for large x)

$$|\Delta| < 2^{(1-\epsilon)\log\log x} \tag{11}$$

as well as

$$\left(\frac{\Delta}{p}\right) = 0 \text{ or } 1 \qquad \text{for all } p \mid N \text{ with } p > 2^{\frac{1}{2}\log\log x}. \tag{12}$$

Our strategy will be to count, for a fixed discriminant $\Delta < 0$ satisfying (11), those N satisfying (12); then we sum on Δ.

We need two lemmas. The first is an estimate for prime character sums weighted by p^{-1}, under GRH.

Lemma 5.7 (under GRH). *Let χ be a nonprincipal Dirichlet character mod m (say). Then $\sum_p \chi(p)/p$ converges, to (say) A_χ. Moreover, for all $X \geqslant 2$,*

$$\sum_{p\leqslant X} \frac{\chi(p)}{p} = A_\chi + O(X^{-1/2}\log\{mX\}).$$

Moreover, $|A_\chi| \leqslant \log\log\log(3m) + O(1)$.

Proof. The convergence of $\sum_p \chi(p)/p$, along with the estimate for its partial sums, follows from the GRH-conditional bound $\sum_{p\leqslant t}\chi(p) = O(t^{1/2}\log\{mt\})$ and summation by parts. To obtain the bound on $|A_\chi|$, we use the trivial estimate $|\chi(p)/p| \leqslant 1/p$ to estimate the contribution from $p \leqslant (\log m)^3$, then apply partial summation to handle the remaining terms. □

The following Hardy–Ramanujan inequality for integers with restricted prime factors is an immediate consequence of [60, Lemma 1].

Lemma 5.8. *There are (absolute) constants A and B for which the following holds: Let $X \geqslant 2$, and let $\mathcal{P}$ be a set of primes not exceeding X. For each positive integer k,*

$$\sum_{\substack{n\leqslant X\\ p\mid n\Rightarrow p\in\mathcal{P}\\ \omega(n)=k}} 1 \leqslant A\frac{X}{\log X}\cdot\frac{1}{(k-1)!}\left(\left(\sum_{p\in\mathcal{P}}\frac{1}{p}\right)+B\right)^{k-1}.$$

Proof of the lower bound half of Theorem 1.7(a). Let Δ be a negative discriminant satisfying (11), let

$$\mathcal{P} = \{p \leqslant x : p \leqslant 2^{\frac{1}{2}\log\log x} \text{ or } \left(\frac{\Delta}{p}\right) \neq -1\},$$

and let k be an integer with

$$|k - \log\log x| < \log\log\log x\cdot\sqrt{\log\log x}. \tag{13}$$

Using Lemma 5.7, we find that for large x,

$$\sum_{p\in\mathcal{P}}\frac{1}{p} \leqslant \sum_{p\leqslant 2^{\log\log x}}\frac{1}{p}+\sum_{p\leqslant x}\frac{1}{2p}\left(1+\left(\frac{\Delta}{p}\right)\right)+\frac{1}{2}\sum_{p\mid\Delta}\frac{1}{p}$$

$$\leqslant O(\log\log\log x)+(\frac{1}{2}\log\log x+O(1))+\frac{1}{2}\sum_{p\leqslant x}\frac{1}{p}\left(\frac{\Delta}{p}\right)+O(\log\log\log|\Delta|)$$

$$=\frac{1}{2}\log\log x+O(\log\log\log x).$$

We plug this estimate for $\sum_{p\in\mathcal{P}}p^{-1}$ into Lemma 5.8. Using Stirling's formula to estimate $(k-1)!$, a short computation reveals that the number of $N\leqslant x$ composed entirely of primes from $\mathcal{P}$ and with $\omega(N)=k$ is at most $x/(\log x)^{\ln 2+o(1)}$. Summing on k satisfying (13) and Δ satisfying (11), we find that the number of N counted for any choice of k,Δ is at most $x/(\log x)^{\epsilon\ln 2+o(1)}$, which is $o(x)$. Keeping in the mind remarks preceding the proof, we have the lower bound half of Theorem 1.7(a). □

We now shift attention to proving that the upper bound in (3) holds for all but $o(x)$ values of $N\leqslant x$, as $x\to\infty$. We can (and will) assume $0<\epsilon<1$.

Write $P^-(N),P^+(N)$ for the least and greatest prime factors of N, with the convention that $P^+(1)=1$ and $P^-(1)=\infty$. We restrict attention to N satisfying all of the following 'anatomical' conditions.

(i) $|\omega(N)-\log\log x|<\log\log\log x\cdot\sqrt{\log\log x}$.
(ii) The largest squarefull divisor of N is at most $\log\log x$.
(iii) The largest d dividing N composed of primes at most $z:=\exp(\sqrt{\log\log x})$ has $d\leqslant z^{\log\log\log x}$.
(iv) $P^+(N)>x^{1/\log\log x}$.

We claim that these conditions exclude only $o(x)$ values of $N\leqslant x$. We have seen this already for (i). The number of $N\leqslant x$ with a squarefull divisor exceeding $\log\log x$ is at most $x\sum_{m>\log\log x,\ \text{squarefull}}1/m\ll\frac{x}{\sqrt{\log\log x}}$, which shows that (ii) is acceptable. By [29, Theorem 07] we get that the number of $N\leqslant x$ violating (iii) is $O(x\exp(-c\log\log\log x))$ for some absolute $c>0$, while [29, Theorem 05] shows that the number of exceptions $N\leqslant x$ to (iv) is $O(x\exp(-c'\log\log x))$ (for some absolute $c'>0$). Hence, (iii) and (iv) are also acceptable.

Put $\theta:=(1+\epsilon)\ln 2$. Let

$$\mathcal{D}_{-1}=\left\{\text{negative fundamental discriminants }\Delta:\frac{1}{2}(\log x)^{\theta}<|\Delta|<(\log x)^{\theta}\right\}.$$

We claim that for all but $o(x)$ of the $N\leqslant x$ satisfying (i)–(iv), there is a $\Delta\in\mathcal{D}_{-1}$ with $(\frac{\Delta}{p})=1$ for all $p\mid N$. Let us see how this claim helps us. Suppose that there is such a Δ. Let N' be the largest squarefull divisor of N, so that $\frac{N}{N'}$ is squarefree and $\gcd(N',\frac{N}{N'})=1$. Then (7) gives

$$d_{\mathrm{CM}}(X_0(N/N'))\leqslant d_{\Delta,\mathrm{CM}}(X_0(N/N'))<(\log x)^{\theta/2}(\log\log x)^{O(1)}.$$

Moreover, we have crudely

$$\frac{d_{\mathrm{CM}}(X_0(N))}{d_{\mathrm{CM}}(X_0(N/N'))} \leqslant \deg(X_0(N) \to X_0(N/N')) = \psi(N)/\psi(N/N') = \psi(N') \ll N'^2 \ll (\log\log x)^2.$$

Noting $(\log x)^{\theta/2} = 2^{(\frac{1}{2}+\frac{1}{2}\epsilon)\log\log x}$, we see from the last two displays that the upper bound in (3) holds (if x is large). The rest of the proof is devoted to proving the claim.

Our strategy is as follows: We sieve $\mathcal{D}_{-1}$, which is a set of size $\asymp (\log x)^{\theta}$, by the set of primes p dividing N. Specifically, for each prime p dividing N, we remove those $\Delta \in \mathcal{D}_{-1}$ for which $(\frac{\Delta}{p}) \neq 1$. The naive expectation, which guides the argument, is that every p should remove very close to half of the values of Δ. Our goal is to show that for almost all N, at least one $\Delta \in \mathcal{D}_{-1}$ survives this sieving process.

We execute the sieve in stages. Let p_0 be the largest divisor of N composed of primes not exceeding $z = \exp(\sqrt{\log\log x})$. (So, contrary to our usual convention, p_0 is not necessarily prime.) By condition (ii) above, N is not divisible by the square of a prime exceeding z. So, we can factor

$$N = p_0 p_1 \cdots p_k,$$

where $p_1 < p_2 < \ldots < p_k$ are primes exceeding z. Note $k \geqslant 1$, since $P^+(N) > x^{1/\log\log x} > z$. From (i), we have

$$k \leqslant \omega(N) < \log\log x + \log\log\log x \cdot \sqrt{\log\log x}.$$

We first sieve by the primes dividing p_0, and then successively by $p_1, p_2, p_3, \ldots$. We let $\mathcal{D}_i$ be the set of discriminants surviving through the ith step. That is,

$$\mathcal{D}_i = \{\Delta \in D_{-1} : \left(\frac{\Delta}{p}\right) = 1 \text{ for all } p \mid p_0 \cdots p_i\}.$$

Clearly,

$$\mathcal{D}_{-1} \supset \mathcal{D}_0 \supset \mathcal{D}_1 \supset \cdots \supset \mathcal{D}_k.$$

In this notation, we are trying to show that for all but $o(x)$ values of $N \leqslant x$ satisfying (i)–(iv), we have $\mathcal{D}_k \neq \emptyset$.

The initial sieving step does not cut down the size of $\mathcal{D}_{-1}$ by very much: As $x \to \infty$,

$$\#\mathcal{D}_0 \geqslant (\log x)^{\theta+o(1)}. \tag{14}$$

To see this, note $p_0 \leqslant z^{\log\log\log x} = (\log x)^{o(1)}$ from (iii), and that

$$\mathcal{D}_0 \supset \left\{\text{negative squarefree } \Delta : \frac{1}{2}(\log x)^{\theta} < |\Delta| < (\log x)^{\theta},\ \Delta \equiv 1 \pmod{8p_0}\right\}.$$

The lower estimate (14) now follows from classical results on the distribution of squarefree numbers in arithmetic progressions. For instance, the estimate (1) of [50] (referred to there as 'near trivial') is more than sufficient.

Suppose $\mathcal{D}_k = \emptyset$. Let δ be a small positive constant, to be specified more precisely momentarily. There must be an index $i \in \{0, 1, \dots, k-1\}$ with

$$\#\mathcal{D}_{i+1} \leqslant \left(\frac{1}{2} - \delta\right)\#\mathcal{D}_i.$$

Let i be the least such index. Then

$$\begin{aligned}\#\mathcal{D}_i &\geqslant \left(\frac{1}{2} - \delta\right)^i \#\mathcal{D}_0 \\ &= \left(\frac{1}{2} - \delta\right)^i (\log x)^{\theta + o(1)} \end{aligned} \tag{15}$$

We now specify our choice of δ: It should be small enough that

$$\theta + \log\left(\frac{1}{2} - \delta\right) > 0.$$

If we now fix a positive $\theta' < \theta + \log(1/2 - \delta)$, then (for large x)

$$\#\mathcal{D}_i \geqslant (\log x)^{\theta'}.$$

Here we have applied the lower bound (15), using $i < k \leqslant (1 + o(1)) \log\log x$. Note that δ and θ' can be chosen to depend only on ϵ (and not on N).

We conclude that when $\mathcal{D}_k = \emptyset$, the number N has a factorization

$$N = Mqr,$$

where

$$M, r \geqslant 1, \quad q \text{ is a prime} > z,$$

$$x/M > x^{1/\log\log x},$$

$$P^-(r) > q > P^+(M),$$

the set

$$\mathcal{D}(M) := \{\Delta \in \mathcal{D}_{-1} : \left(\frac{\Delta}{p}\right) = 1 \text{ for all } p \mid M\}$$

satisfies

$$\#\mathcal{D}(M) \geqslant (\log x)^{\theta'},$$

and

$$\#\{\Delta \in \mathcal{D}(M) : \left(\frac{\Delta}{q}\right) \neq 1\} \geqslant \left(\frac{1}{2} + \delta\right)\#\mathcal{D}(M). \tag{16}$$

Explicitly, we can take $M = p_0 \cdots p_i$, $q = p_{i+1}$, and $r = p_{i+2} \cdots p_k$. We will show that only $o(X)$ values of N admit such a factorization. This will complete the proof of the claim, and thus also of Theorem 1.7(a).

We take two cases, according to whether or not $r = 1$.

When $r = 1$, we count the N with a factorization of this kind by pivoting on the value of M. For each M, we bound the number of primes $q \leqslant x/M$ for which (16) holds. We then sum on M.

To execute this plan, it is again convenient to adopt the language of probability. Let $\Omega := \{\text{primes } q : q \leqslant x/M\}$, viewed as a finite probability space with the uniform measure. For each $\Delta \in \mathcal{D}(M)$, we introduce the random variable X_Δ defined by

$$X_\Delta(q) = \begin{cases} 1 & \text{if } \left(\frac{\Delta}{q}\right) \neq 1, \\ 0 & \text{if } \left(\frac{\Delta}{q}\right) = 1. \end{cases} \tag{17}$$

Set

$$X(q) = \sum_{\Delta \in \mathcal{D}(M)} X_\Delta(q).$$

Then $\mathbb{E}[X] = \sum_{\Delta \in \mathcal{D}(M)} \mathbb{E}[X_\Delta]$. Since $X_\Delta(q) = \frac{1}{2}(1 - (\frac{\Delta}{q})) + \frac{1}{2} \cdot \mathbf{1}_{q|\Delta}$, for each $\Delta \in \mathcal{D}(M)$ we have on GRH

$$\begin{aligned} \mathbb{E}[X_\Delta] &= \frac{1}{2} - \frac{1}{2\#\Omega} \sum_{\substack{q \text{ prime} \\ q \leqslant x/M}} \left(\frac{\Delta}{q}\right) + O\left(\frac{\log|\Delta|}{\#\Omega}\right) \\ &= \frac{1}{2} + O\left(\frac{1}{\#\Omega}\left(\frac{x}{M}\right)^{1/2} \log(|\Delta| x/M)\right). \end{aligned}$$

Since $\frac{x}{M} > x^{1/\log\log x}$ and $\#\Omega \gg x/(M \log x)$, the O-term here is (crudely) $O((\log x)^{-10})$. It follows

$$\mathbb{E}[X] = \frac{1}{2}\#\mathcal{D}(M) + O(\#\mathcal{D}(M) \cdot (\log x)^{-10}).$$

Next, we compute the variance of X. Clearly, $\mathbb{E}[X^2] = \sum_{\Delta,\Delta' \in \mathcal{D}(M)} \mathbb{E}[X_\Delta X_{\Delta'}]$. If $\Delta = \Delta'$, then $X_\Delta X_{\Delta'} = X_\Delta$, and $\mathbb{E}[X_\Delta X_{\Delta'}] = \frac{1}{2} + O((\log x)^{-10})$ (as shown above). For the remaining terms in $\mathbb{E}[X^2]$, we have

$$X_\Delta X_{\Delta'} = \frac{1}{4} - \frac{1}{4}\left(\frac{\Delta}{q}\right) - \frac{1}{4}\left(\frac{\Delta'}{q}\right) + \frac{1}{4}\left(\frac{\Delta\Delta'}{q}\right) + O(\mathbf{1}_{q|\Delta} + \mathbf{1}_{q|\Delta'}).$$

Since Δ and Δ' are distinct fundamental discriminants, their product $\Delta\Delta'$ is not a square, so $(\frac{\Delta\Delta'}{\cdot})$ is a nontrivial character mod $|\Delta\Delta'|$. So under GRH,

$$\begin{aligned}\mathbb{E}[X_\Delta X_{\Delta'}] &= \frac{1}{4} + O\Big(\frac{1}{\#\Omega}\sqrt{\frac{x}{M}}\log\Big\{|\Delta\Delta'|\frac{x}{M}\Big\}\Big) + O\Big(\frac{1}{\#\Omega}\log|\Delta\Delta'|\Big)\\ &= \frac{1}{4} + O((\log x)^{-10}).\end{aligned}$$

Summing on $\Delta, \Delta' \in \mathcal{D}(M)$, we find

$$\mathbb{E}[X^2] = \frac{1}{4}(\#\mathcal{D}(M))^2 + O(\#\mathcal{D}(M) + (\#\mathcal{D}(M))^2 \cdot (\log x)^{-10}),$$

and

$$\mathbb{E}\Big[(X - \frac{1}{2}\#\mathcal{D}(M))^2\Big] = O(\#\mathcal{D}(M) + (\#\mathcal{D}(M))^2 \cdot (\log x)^{-10}).$$

By Chebyshev's inequality,

$$\Pr\Big(X \geqslant \Big(\frac{1}{2} + \delta\Big)\#\mathcal{D}(M)\Big) \ll (\#\mathcal{D}(M))^{-1} + (\log x)^{-10} \ll (\log x)^{-\theta'}.$$

Hence, given M, the number of $q \leqslant x/M$ for which (16) holds is

$$\ll \#\Omega(\log x)^{-\theta'} \ll \frac{x}{M\log(x/M)}(\log x)^{-\theta'} \ll \frac{x\log\log x}{M\log x}(\log x)^{-\theta'}.$$

Now sum on $M \leqslant x$. We find that the total number of N with factorizations of the desired kind, in the case $r = 1$, is $O(x(\log\log x)(\log x)^{-\theta'})$, and so is $o(x)$ as $x \to \infty$.

Now suppose $r > 1$. We first count, for given values of M and q, the number of possible values of r. We need $r \leqslant \frac{x}{Mq}$, and $P^-(r) > q > P^+(M)$. By Brun's upper bound sieve (see [28, p. 68, Theorem 2.2]),

$$\begin{aligned}\#\{r \leqslant \frac{x}{Mq} : P^-(r) > P^+(M)\} &\ll \frac{x}{Mq}\prod_{p \leqslant P^+(M)}\Big(1 - \frac{1}{p}\Big)\\ &\ll \frac{x}{Mq\cdot\log P^+(M)}.\end{aligned} \tag{18}$$

In order to verify the hypotheses of the upper bound sieve, we used that $\frac{x}{Mq} > P^+(M)$. (Indeed, $\frac{x}{Mq} \geqslant r \geqslant P^-(r) > P^+(M)$.)

Now we fix M and sum the bound of (18) on q satisfying (16). For this we employ a second moment argument similar to that seen above for the case $r = 1$, but now counting the primes q with a weight of q^{-1}.

We let $\Omega := \{q \text{ prime} : z < q \leqslant \frac{x}{M}\}$, and we put

$$W := \sum_{q\in\Omega}\frac{1}{q}.$$

A short calculation gives

$$W = \log\log x + O(\log\log\log x).$$

We turn Ω into a finite probability space by assigning to each $q \in \Omega$ the probability $(qW)^{-1}$. We define X_Δ, for $\Delta \in \mathcal{D}(M)$, as in (17), and we let $X = \sum_{\Delta\in\mathcal{D}(M)} X_\Delta$. Then $\mathbb{E}[X] = \sum_{\Delta\in\mathcal{D}(M)} \mathbb{E}[X_\Delta]$, and for each $\Delta \in \mathcal{D}(M)$,

$$\begin{aligned}\mathbb{E}[X_\Delta] &= \frac{1}{2} - \frac{1}{2W} \sum_{\substack{q \text{ prime} \\ z<q\leqslant x/M}} \left(\frac{\Delta}{q}\right)\frac{1}{q} + O\left(\frac{1}{W}\sum_{\substack{q \text{ prime} \\ q|\Delta,\ q>z}} \frac{1}{q}\right) \\ &= \frac{1}{2} + O\Big(\frac{1}{W}\cdot z^{-1/2}\log|\Delta z|\Big) + O\Big(\frac{1}{W}\cdot z^{-1}\log|\Delta|\Big).\end{aligned}$$

The second O-term is subsumed by the first. Looking back, we see that $\log|\Delta| \asymp \log\log x \asymp W$, while $\log z = (\log\log x)^{1/2}$. So $\mathbb{E}[X_\Delta] = \frac{1}{2} + O(z^{-1/2})$, and

$$\mathbb{E}[X] = \frac{1}{2}\#\mathcal{D}(M) + O(\#\mathcal{D}(M)\cdot z^{-1/2}).$$

A similar calculation shows that for all distinct $\Delta, \Delta' \in \mathcal{D}(M)$,

$$\mathbb{E}[X_\Delta X_{\Delta'}] = \frac{1}{4} + O(z^{-1/2}),$$

so that $\mathbb{E}[X^2] = \frac{1}{4}(\#\mathcal{D}(M))^2 + O(\#\mathcal{D}(M) + (\#\mathcal{D}(M))^2\cdot z^{-1/2})$ and

$$\mathbb{E}\Big[(X - \frac{1}{2}\#\mathcal{D}(M))^2\Big] = O(\#\mathcal{D}(M) + (\#\mathcal{D}(M))^2\cdot z^{-1/2}).$$

Hence,

$$\Pr\Big(X \geqslant \Big(\frac{1}{2}+\delta\Big)\#\mathcal{D}(M)\Big) \ll (\#\mathcal{D}(M))^{-1} + z^{-1/2} \ll z^{-1/2}.$$

It follows that given M,

$$\sum_q \frac{1}{q} \ll Wz^{-1/2} \ll z^{-1/3},$$

where q runs over the primes in $(z, x/M]$ for which (16) holds. Putting this back into (18), we see that the number of N that have a factorization of the desired kind, with $r > 1$, and with M given, is $O(xz^{-1/3}\frac{1}{M\log P^+(M)})$. So, the total number of N arising in the case $r > 1$ is

$$\ll xz^{-1/3}\sum_{M\leqslant x}\frac{1}{M\log P^+(M)}.$$

Of course, the term $M = 1$ contributes $O(1)$ to the sum on M. To bound the contribution of the remaining terms, we pivot on the value of $p = P^+(M)$ to find

$$\sum_{1<M\leqslant x} \frac{1}{M \log P^+(M)} \leqslant \sum_{p\leqslant x} \frac{1}{\log p} \sum_{M:\ P^+(M)=p} \frac{1}{M}$$

$$\leqslant \sum_{p\leqslant x} \frac{1}{p \log p} \sum_{P^+(M')\leqslant p} \frac{1}{M'} = \sum_{p\leqslant x} \frac{1}{p \log p} \prod_{\substack{\ell \text{ prime} \\ \ell\leqslant p}} \left(1 - \frac{1}{\ell}\right)^{-1} \ll \sum_{p\leqslant x} \frac{1}{p} \ll \log\log x.$$

Substituting this back above bounds the number of N by $xz^{-1/3} \cdot \log\log x$, which is $\ll xz^{-1/4}$, and so is $o(x)$. This completes the proof.

6 | EXPLICIT AND UNCONDITIONAL UPPER BOUNDS

For our later discussion of sporadic points, it will be important to have upper bounds on $d_{\mathrm{CM}}(X_0(N))$, $d_{\mathrm{CM}}(X_1(N))$, and $d_{\mathrm{CM}}(X_1(M,N))$ that are completely explicit and not dependent on unproved hypotheses. For $N \in \mathbb{Z}^+$, let N_0 denote the product of the distinct odd primes dividing N. Put $\Delta = 1 - 8N_0$, and write

$$\Delta = f^2 \Delta_K,$$

where Δ_K is a fundamental discriminant. Then $(\frac{\Delta_K}{p}) = 1$ for all primes p dividing N. Note $1 \equiv \Delta \equiv \Delta_K \pmod 8$, so that $\Delta_K \leqslant -7$.

For an imaginary quadratic field K of discriminant $\Delta_K < -4$ we have [16, Appendix]

$$h_K \leqslant \frac{e}{2\pi} \sqrt{|\Delta_K|} \ln |\Delta_K|.$$

Using Proposition 4.7 and Lemma 4.4, we get

$$d_{\mathrm{CM}}(X_0(N)) \leqslant 2h_K \leqslant \frac{e}{\pi} \sqrt{8N_0} \ln |8N_0|, \tag{19}$$

$$d_{\mathrm{CM}}(X_1(N)) \leqslant \frac{e}{2\pi} \phi(N) \sqrt{8N_0} \ln |8N_0|,$$

and for $M \mid N$,

$$d_{\mathrm{CM}}(X_1(M,N)) \leqslant \frac{e}{\pi} M\phi(N) \sqrt{8N_0} \ln |8N_0|. \tag{20}$$

If we drop the requirement of explicitness, sharper unconditional bounds can be obtained. We focus on $d_{\mathrm{CM}}(X_0(N))$, leaving the reader to supply the corresponding estimates for $d_{\mathrm{CM}}(X_1(N))$ and $d_{\mathrm{CM}}(X_1(M,N))$ that follow from Lemma 4.7.

Theorem 6.1. *Let $\epsilon > 0$. With N_0 the product of the distinct odd primes dividing N, we have $d_{\mathrm{CM}}(X_0(N)) \ll_\epsilon N_0^{\frac{1}{8}+\epsilon}$.*

The proof depends on the following special case of a result of Norton (see [48, Corollary 3.38]), proved using Burgess's fundamental work on character sums.

Lemma 6.2. *Let $M \in \mathbb{Z}^+$, and let H be a subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$ containing $(\mathbb{Z}/M\mathbb{Z})^{\times 2}$. Every coset of H has a positive integer representative that is $O_\epsilon(M^{1/4+\epsilon})$.*

Proof of Theorem 6.1. Let $N' = 8N_0$, and write $N' = 2^3 p_1 \cdots p_r$, where $p_1, \dots, p_r$ are odd primes. We define a group homomorphism

$$\iota = (\iota_1, \dots, \iota_{r+2}) := (\mathbb{Z}/N'\mathbb{Z})^\times \to \{\pm 1\}^{r+2},$$

as follows:

- for $1 \leqslant i \leqslant r$, we put $\iota_i(a) = (\frac{a}{p_i})$;
- we put $\iota_{r+1}(a) = 1 \iff a \equiv 1 \pmod 4$;
- we put $\iota_{r+2}(a) = 1 \iff a \equiv \pm 1 \pmod 8$.

Let H be the kernel of ι, so $(\mathbb{Z}/N'\mathbb{Z})^{\times 2} \subset H$. By Lemma 6.2, for all $\epsilon > 0$ there is an integer $1 \leqslant a \ll_\epsilon N'^{1/4+\epsilon} \ll N_0^{1/4+\epsilon}$ such that

$$\iota(a) = \left(\left(\frac{-1}{p_1}\right), \dots, \left(\frac{-1}{p_r}\right), -1, 1\right).$$

Then $\Delta = -a$ is a negative discriminant satisfying $(\frac{\Delta}{p}) = 1$ for all $p \mid N$. The result now follows from Proposition 4.7 and Lemma 2.2. □

Remark 6.3. For prime powers $N = \ell^a$, one can prove the somewhat sharper estimate

$$d_{\mathrm{CM}}(X_0(\ell^a)) \ll_\epsilon \ell^{\frac{1}{8\sqrt{e}}+\epsilon}. \tag{21}$$

In fact, if $\ell = 2$ or $\ell \equiv 1 \pmod 4$, then $d_{\mathrm{CM}}(X_0(\ell^a))$ is bounded; this follows from taking $\Delta = -7$ or $\Delta = -4$ (respectively) in Proposition 4.7. So to prove (21), we may assume $\ell \equiv 3 \pmod 4$. Burgess [12] has shown that for all $\epsilon > 0$ there is a prime

$$q \ll_\epsilon \ell^{\frac{1}{4\sqrt{e}}+\epsilon}$$

which is a quadratic nonresidue modulo ℓ. Now put $\Delta = -q$ or $\Delta = -4q$ according to whether $q \equiv 3 \pmod 4$ or $q \equiv 1, 2 \pmod 4$. Since $\ell \equiv 3 \pmod 4$, we have $(\frac{\Delta}{\ell}) = (\frac{-1}{\ell})(\frac{q}{\ell}) = 1$, and it follows

$$d_{\mathrm{CM}}(X_0(\ell^a)) \leqslant d_{\Delta,\mathrm{CM}}(X_0(\ell^a)) \leqslant 2h_\Delta \ll_\epsilon \ell^{\frac{1}{8\sqrt{e}}+2\epsilon}.$$

Replacing ϵ with $\frac{\epsilon}{2}$ finishes the proof.

7 | SPORADIC CM POINTS ON MODULAR CURVES

7.1 | GL_2 modular curves

In this section, we give a brief review of GL_2-modular curves.

Let $E_{/\mathbb{Q}(t)}$ be an elliptic curve with j-invariant t. For $N \in \mathbb{Z}^+$, let $\mathbb{Q}(X(N))$ be the field obtained by adjoining the x-coordinates of the N-torsion points of E. This field is independent of the choice of E and

$$\mathrm{Aut}(\mathbb{Q}(X(N))/\mathbb{Q}(t)) \cong \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}.$$

When $N \geqslant 3$ the field $\mathbb{Q}(X(N))$ is however not a 'regular' function field: the algebraic closure of $\mathbb{Q}$ in $\mathbb{Q}(X(N))$ is $\mathbb{Q}(\zeta_N)$. Thus, $\mathbb{Q}(X(N))$ is the function field of a curve defined over $\mathbb{Q}$ that is smooth and integral but with $\phi(N)$ geometric connected components, each of which is defined over $\mathbb{Q}(\zeta_N)$. We write $\mathbb{Q}(\zeta_N)(X((N)))$ for the field $\mathbb{Q}(X(N))$ regarded as a function field over $\mathbb{Q}(\zeta_N)$; then there is a corresponding nice (smooth, projective geometrically integral) curve $X(N)_{/\mathbb{Q}(\zeta_N)}$ which is isomorphic to any one of the connected components of the base change of the scheme $X(N)_{/\mathbb{Q}}$ to $\mathbb{Q}(\zeta_N)$.

For $M \mid N$ we have $\mathbb{Q}(X(M)) \subset \mathbb{Q}(X(N))$, so we get a tower of function fields. Let

$$\mathbb{Q}(X(\infty)) := \bigcup_{N \geqslant 1} \mathbb{Q}(X(N)).$$

Then

$$\mathrm{Aut}(\mathbb{Q}(X(\infty))/\mathbb{Q}(X(1))) = \mathrm{GL}_2(\widehat{\mathbb{Z}})/\{\pm 1\}.$$

Let

$$q_N : \ \mathrm{Aut}(\mathbb{Q}(X(\infty))/\mathbb{Q}(X(1))) \to \mathrm{Aut}(\mathbb{Q}(X(N))/\mathbb{Q}(X(1)))$$

be the natural map. For a subgroup H of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$,

$$\mathbb{Q}(X(H)) := \mathbb{Q}(X(\infty))^{q_N^{-1}(H)}$$

is a subextension of $\mathbb{Q}(X(\infty))/\mathbb{Q}(X(1))$ such that

$$I(H) := [\mathbb{Q}(X(H)) : \mathbb{Q}(X(1))] = \frac{\#\,\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}}{\#H}.$$

(If H is any subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, then we will associate the subgroup $H^+ := H\{\pm 1\}$ of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. In the literature, it is common to ignore the distinction between H and H^+. We will try not to do this here.) The algebraic closure of $\mathbb{Q}$ in $\mathbb{Q}(X(H))$ is

$$\mathbb{Q}_H := \mathbb{Q}(\zeta_N)^{\det H}.$$

Thus, we can view $X(H)$ as a curve over $\mathbb{Q}$ that is smooth and integral and also as a curve over $\mathbb{Q}_H$ that is smooth and geometrically integral. Here we will say that H is *rational* if $\mathbb{Q}_H = \mathbb{Q}$. Moreover, for subgroups

$$H_1 \subset \mathrm{GL}_2(\mathbb{Z}/N_1\mathbb{Z})/\{\pm 1\}, \quad H_2 \subset \mathrm{GL}_2(\mathbb{Z}/N_2\mathbb{Z})/\{\pm 1\},$$

we write $H_1 \prec H_2$ if $q_{N_1}^{-1}(H_1) \subset q_{N_2}^{-1}(H_2)$. Thus, we have $H_1 \prec H_2$ if and only if $\mathbb{Q}(X(H_1)) \supset \mathbb{Q}(X(H_2))$ and when these conditions hold there is a finite $\mathbb{Q}$-morphism $X(H_1) \to X(H_2)$.

To $H \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ we can also attach a congruence subgroup $\Gamma(H) \subset \mathrm{PSL}_2(\mathbb{Z}) := \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$: namely, we take the complete preimage $\Gamma(H)$ of $H \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ under the quotient map $\mathrm{PSL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. Then $X(\Gamma(H)) := \Gamma(H)\backslash\overline{\mathcal{H}}$ is a compact Riemann surface — that is, a nice curve defined over $\mathbb{C}$ — and the extension of the nice curve $X(H)$ from $\mathbb{Q}_H$ to $\mathbb{C}$ is isomorphic to $X(\Gamma(H))$. If we put

$$\mathrm{SI}(H) := [\mathrm{PSL}_2(\mathbb{Z}) : \Gamma(H)],$$

then we have $\mathrm{SI}(H) \mid I(H)$, with equality if and only if $\mathbb{Q}_H = \mathbb{Q}$.

7.2 | Proof of Lemma 4.1

Let M and N be positive integers with $M \mid N$. We will give definitions of the modular curves $X_0(N)$, $X_1(N)$, $X(N)$ and $X_1(M,N)$ over $\mathbb{Q}$ and use these definitions to prove Lemma 4.1. This material is well-known to the experts, but in view of the distinction between GL_2 and SL_2 it cannot hurt to be explicit. Moreover, we have (somewhat vexingly) not been able to find a reference for Lemma 4.1(d) — which includes parts (b) and (c) as special cases — so a proof seems necessary for completeness.

As in the previous section, to define each modular curve we need to specify a subgroup H of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ and then $I(H)$ is simply $[\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\} : H]$.

For any commutative rings $R_1, \dots, R_r$ and $n \in \mathbb{Z}^+$ we have a canonical isomorphism

$$\mathrm{GL}_n\left(\prod_{i=1}^{r} R_i\right) = \prod_{i=1}^{r} \mathrm{GL}_n(R_i).$$

If $N = p_1^{a_1} \cdots p_r^{a_r}$, for each subgroup $\mathcal{H} \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ considered below,[†] it will be clear that for all $1 \leqslant i \leqslant r$ we have subgroups $\mathcal{H}_i$ of $\mathrm{GL}_2(\mathbb{Z}/p_i^{a_i}\mathbb{Z})$ such that $\mathcal{H} = \prod_{i=1}^{r} \mathcal{H}_i$ and thus

$$[\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) : \mathcal{H}] = \prod_{i=1}^{r} [\mathrm{GL}_2(\mathbb{Z}/p_i^{a_i}\mathbb{Z}) : \mathcal{H}_i].$$

We call this phenomenon 'primary decomposition'.

We may view $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ as the automorphism group of the $\mathbb{Z}/N\mathbb{Z}$-module $V = V(N) := \mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$; let $e_1 := (1,0)$ and $e_2 := (0,1)$.

[†] This is certainly *not* the case for an arbitrary subgroup H.

- Let $\widetilde{H_0} = \widetilde{H_0(N)}$ be the subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ consisting of matrices $g \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ such that $ge_1 \in \langle e_1 \rangle$. Otherwise put, we have

$$\widetilde{H_0} = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a, d \in (\mathbb{Z}/N\mathbb{Z})^\times,\ b \in \mathbb{Z}/N\mathbb{Z} \right\}.$$

  We put $H_0 = H_0(N) := \widetilde{H_0} \cdot \{\pm 1\} = \widetilde{H_0}$. The subgroup H_0 defines the modular curve $X_0(N)$. A $\mathbb{Z}/N\mathbb{Z}$-submodule of V is free of rank 1 if and only if it is generated by a primitive vector, that is, an element $v \in V$ of order N. It is an easy consequence of the structure theory of finitely generated $\mathbb{Z}$-modules that $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ acts transitively on primitive vectors and thus also on free, rank one $\mathbb{Z}/N\mathbb{Z}$-submodules of V. The Orbit-Stabilizer theorem gives that $I(H_0)$ is equal to the number of free, rank one $\mathbb{Z}/N\mathbb{Z}$-submodules of V, and this in turn is $\frac{1}{\phi(N)}$ times the number of primitive vectors in V. Primary decomposition reduces us to the case of $N = p^a$. Clearly $V(p)$ has $p^2 - 1$ primitive vectors, and a vector in $V(p^a)$ is primitive if and only if its mod p reduction is primitive, so $V(p^a)$ has $p^{2a-2}(p^2 - 1) = \phi(N)\psi(N)$ primitive vectors, hence $\psi(N)$ free, rank one $\mathbb{Z}/N\mathbb{Z}$-submodules of V, so $I(H_0) = \psi(N)$.
- Let $\widetilde{H_1} = \widetilde{H_1(N)}$ be the subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ consisting of matrices $g \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ such that $ge_1 = e_1$. Otherwise put, we have

$$\widetilde{H_1} = \left\{ \begin{pmatrix} 1 & b \\ 0 & d \end{pmatrix} \mid d \in (\mathbb{Z}/N\mathbb{Z})^\times,\ b \in \mathbb{Z}/N\mathbb{Z} \right\}.$$

  The modular curve $X_1(N)$ is defined by the subgroup $H_1 = H_1(N) := \widetilde{H_0} \cdot \{\pm 1\}$ of the group $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. Very similarly to the above we get that the index of $\widetilde{H_1}$ in $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ is equal to the number of primitive vectors in V and thus once again is $\phi(N)\psi(N)$. If $N = 2$ then $-1 \in \widetilde{H_0} = \widetilde{H_1}$ so we have $I(H_1) = 3$. If $N \geqslant 3$, then $I(H_1) = \frac{[\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) : \widetilde{H_1}]}{2} = \frac{\phi(N)\psi(N)}{2}$.
- Let $\widetilde{H_2} = \widetilde{H_2(N)}$ be the trivial subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. The modular curve $X(N)$ is defined by the subgroup $H_2 = H_2(N) := \widetilde{H_2} \cdot \{\pm 1\} = \{\pm 1\}$. Thus, $I(H_2) = \frac{\#\,\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})}{2}$. We are left to compute $\#\,\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, which is the number of ordered $\mathbb{Z}/N\mathbb{Z}$-bases of V. Primary decomposition reduces us to the case of $N = p^a$. When $N = p$ we can take any nonzero vector as the first basis vector and any vector not lying in the span of the first vector as the second basis vector, giving

$$\#\,\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z}) = (p^2 - 1)(p^2 - p) = p\phi(p)^2\psi(p).$$

  For $a \geqslant 1$, reduction modulo p^a gives a short exact sequence

$$1 \to K \to \mathrm{GL}_2(\mathbb{Z}/p^{a+1}\mathbb{Z}) \to \mathrm{GL}_2(\mathbb{Z}/p^a\mathbb{Z}) \to 1,$$

  where $K = 1 + p^a M_2(\mathbb{Z}/p^{a+1}\mathbb{Z})$ has order p^4. This leads to the formula

$$\#\,\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) = N\phi(N)^2\psi(N)$$

  and thus to

$$I(H_2) = \frac{N\phi(N)^2\psi(N)}{2}.$$

- For $M \mid N$, let $\widetilde{H} = \widetilde{H(M,N)} := \widetilde{H_1(N)} \cap \widetilde{H_2(M)} \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. Otherwise put, we have

$$\begin{aligned}\widetilde{H} &= \widetilde{H(M,N)} \\ &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a \equiv 1 \pmod{N},\ b \equiv 0 \pmod{N}, c \equiv 0 \pmod{M},\ d \equiv 1 \pmod{M} \right\}.\end{aligned}$$

We have $\widetilde{H(2,1)} = H_1(2)$ and $\widetilde{H(2,2)} = H_2(2)$, so we may assume $N \geqslant 3$ and thus

$$I(H(M,N)) = \frac{[\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) : \widetilde{H(M,N)}]}{2}.$$

Primary decomposition reduces us to the case $M = p^s$, $N = p^t$ with $0 \leqslant s \leqslant t$. When $s = 0$ we have $\widetilde{H(p^s, p^t)} = \widetilde{H_1(p^t)}$, so we may assume $s \geqslant 1$. The description using matrices yields

$$[\widetilde{H(p^s, p^s)} : \widetilde{H(p^s, p^t)}] = p^{2t-2s},$$

while

$$[\mathrm{GL}_2(\mathbb{Z}/p^t\mathbb{Z}) : \widetilde{H(p^s, p^s)}] = \#\,\mathrm{GL}_2(\mathbb{Z}/p^s\mathbb{Z}) = p^s\phi(p^s)^2\psi(p^s),$$

so

$$[\mathrm{GL}_2(\mathbb{Z}/p^t\mathbb{Z}) : \widetilde{H(p^s, p^t)}] = (p-1)^2(p+1)p^{2s+2t-3} = M\phi(M)\phi(N)\psi(N).$$

7.3 | Sporadic points

Let X be a nice curve defined over a field F. The *gonality* $\gamma_F(X)$ is the least degree of a finite F-rational morphism $X \to \mathbb{P}^1$. If L/F is a field extension, we put $\gamma_L(X) := \gamma_L(X_{/L})$, that is, the gonality of the base extension to L. We then have $\gamma_L(X) \leqslant \gamma_F(X)$.

A closed point p on $X_{/F}$ has *low degree* if its degree $\deg p = [F(p) : F]$ is less than the gonality $\gamma_F(X)$. A closed point p on $X_{/F}$ is *sporadic* if the set of closed points q of F with $\deg q \leqslant \deg p$ is finite. Since every nice curve has infinitely many closed points of degree at most $\gamma_F(X)$, a sporadic point necessarily has low degree.

Sporadic points on curves are interesting and often elusive. To produce such points, one must first find a point of low degree and second establish that there are only finitely many points of equal or smaller degree. There are few general techniques for showing the latter part. By far the most widely used is the following result, essentially due to Frey [25] but stated there with an unnecessary hypothesis on rational points.

To state it (as well as for other, later purposes) it is convenient to make one more definition: for a nice curve $X_{/F}$ we define $\delta(X)$ to be the least degree d such that X has infinitely many closed points of degree d. Thus, a sporadic point is a point of degree less than $\delta(X)$.

Theorem 7.1. *For a nice curve X defined over a number field F, we have*

$$\frac{\gamma_F(X)}{2} \leqslant \delta(X) \leqslant \gamma_F(X).$$

Proof. Since $\mathbb{P}^1(F)$ is infinite, there are infinitely many closed points on X of degree at most $\gamma_F(X)$.[†] Thus, $\delta(X) \leqslant \gamma_F(X)$.

In [14, Theorem 5], it is shown that if for $d \in \mathbb{Z}^+$ the set of closed points of degree dividing d is infinite, then $\gamma_F(X) \leqslant 2d$. Applying this with $d = \delta(X)$ gives $\frac{\gamma_F(X)}{2} \leqslant \delta(X)$. □

Now we consider the modular curve $X(H)$ attached to a subgroup H of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. As above, $X(H)$ can in all cases be defined over $\mathbb{Q}$ but is geometrically integral (and hence nice) if and only if the subgroup H is rational. Although the definitions of gonality, low degree and sporadic points make sense for such curves, to the best of our knowledge these concepts have only been studied in the literature for nice curves, and moreover Theorem 7.1 applies to nice curves. Thus, we will study sporadic points on the geometrically integral curve $X(H)_{/\mathbb{Q}_H}$. We want to point out the interesting recent paper [9], which obtains results on non-CM sporadic points on modular curves, with a particular emphasis on the case of rational j-invariant.

Lemma 7.2. *Let H be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$, and let $X(H)_{/\mathbb{Q}(H)}$ be the corresponding modular curve. Viewing $X(H)$ as a curve over $\mathbb{C}$ by base extension, we have*

$$\gamma_{\mathbb{C}}(X(H)) > \frac{119}{12000}\,\mathrm{SI}(H).$$

Proof. This is [36, Theorem 1.3]. As explained therein, this is obtained by combining work of Abramovich [1, Theorem 0.1] with the best known partial result on Selberg's eigenvalue conjecture due to Kim and Sarnak [38, p. 176]. □

The following result is a direct generalization of [9, Lemma 6.2].

Theorem 7.3. *Let H_0 be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. Suppose there is a closed point $p_0 \in X(H_0)$ such that*

$$\deg_{\mathbb{Q}} p_0 \leqslant \frac{119}{24000} I(H_0) = \frac{119}{24000} \deg(X(H_0) \to X(1)). \tag{22}$$

For every subgroup $H \prec H_0$, every closed point p of $X(H)_{/\mathbb{Q}(H)}$ lying over p_0 is sporadic.

Proof. For any subgroup $H \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ we have

$$I(H) = \mathrm{SI}(H)[\mathbb{Q}_H : \mathbb{Q}],$$

Thus,

$$\begin{aligned}\deg_{\mathbb{Q}_H} p = \frac{\deg_{\mathbb{Q}} p}{[\mathbb{Q}_H : \mathbb{Q}]} &\leqslant \frac{\deg_{\mathbb{Q}} p_0 \cdot \deg(X(H) \to X(H_0))}{[\mathbb{Q}_H : \mathbb{Q}]} \\ &= \frac{\deg_{\mathbb{Q}} p_0}{[\mathbb{Q}_H : \mathbb{Q}]} \frac{I(H)}{I(H_0)} \leqslant \frac{119}{24000} \frac{I(H)}{[\mathbb{Q}_H : \mathbb{Q}]} = \frac{119}{24000}\,\mathrm{SI}(H).\end{aligned}$$

[†] Though it is not necessary for this argument, the Hilbert irreducibility theorem implies there are infinitely many closed points on X of degree $\gamma_F(X)$.

By Lemma 7.2 we have

$$\frac{\gamma_{\mathbb{Q}_H}(X(H))}{2} \geqslant \frac{\gamma_{\mathbb{C}}(X(H))}{2} > \frac{119}{24000}\,\mathrm{SI}(H) \geqslant \deg_{\mathbb{Q}_H} p,$$

so p is a sporadic point on $X(H)_{/\mathbb{Q}_H}$ by Theorem 7.1. □

Remark 7.4. For a reduced curve $X_{/\mathbb{Q}}$, we may define the gonality $\gamma_{\mathbb{Q}}(X)$ as the least degree of a dominant $\mathbb{Q}$-morphism $X \to \mathbb{P}^1$ and the invariant $\delta(X)$ as the least $d \in \mathbb{Z}^+$ such that X admits infinitely many closed points of degree d, and then call a closed point $p \in X$ sporadic if $\deg_{\mathbb{Q}} p < \delta(X)$. Evidently, we still have $\delta(X) \leqslant \gamma_{\mathbb{Q}}(X)$. This applies in particular to the curves $X_1(M,N)_{/\mathbb{Q}}$ for all $M \mid N$. Since the residue field of every closed point on $X_1(M,N)$ contains $\mathbb{Q}(\zeta_M)$, we observe that the set of sporadic closed points on $X_1(M,N)_{/\mathbb{Q}}$ is the same as the set of sporadic closed points on $X_1(M,N)_{/\mathbb{Q}(\zeta_M)}$, and the bound (22) of Theorem 7.3 can be viewed in terms of $X_1(M,N)_{/\mathbb{Q}}$.

We call a closed point p_0 on a modular curve $X(H_0)$ *super-sporadic* if for every $H \prec H_0$, every closed point of $X(H)$ lying over p_0 is sporadic. Thus Theorem 7.3 says that $p_0 \in X(H_0)$ is super-sporadic if $\frac{\deg_{\mathbb{Q}} p_0}{I(H_0)} < \frac{119}{24000}$.

Example 7.5. Let ℓ be a prime number. The modular curve $X_0(\ell)$ has two cusps, 0 and ∞, and both are $\mathbb{Q}$-rational. Let p_0 be either of these points. Then (22) holds for p_0 and $H_0(\ell)$ if and only if $\ell \geqslant 211$. It follows that for every modular curve $X(H)$ associated to a subgroup $H \prec H_0(\ell)$, every cusp on $X(H)$ is a sporadic point.

Super-sporadic CM points abound on modular curves. Indeed, see the following.

Theorem 7.6. *There is a constant* $\mathbf{A}$ *such that for all* $N \geqslant \mathbf{A}$ *and all* $M \mid N$, *the curves* $X_0(N)_{/\mathbb{Q}}$ *and* $X_1(M,N)_{/\mathbb{Q}(\zeta_M)}$ *have super-sporadic CM points.*

Proof. It follows immediately from the definition of a super-sporadic point that if $H_2 \prec H_1$, $\pi : X(H_2) \to X(H_1)$ is the natural map and $p \in X(H_2)$, then if $\pi(p)$ is super-sporadic, so is p. Now all of the above modular curves cover $X_0(N)$, and all sufficiently large N are divisible by a large prime power ℓ^a, so it is enough to prove that for all but finitely many prime powers ℓ^a, the modular curve $X_0(\ell^a)_{/\mathbb{Q}}$ has a super-sporadic CM point. Let H_0 be the subgroup that defines $X_0(\ell^a)$. Then

$$I(H_0) = \deg(X_0(\ell^a) \to X(1)) = \psi(\ell^a) = \ell^a + \ell^{a-1} \geqslant \ell^a,$$

while (21) gives

$$d_{\mathrm{CM}}(X_0(\ell^a)) \ll \ell^{0.079},$$

so the set of prime powers ℓ^a such that $\frac{d_{\mathrm{CM}}(X_0(\ell^a))}{I(H_0)} \geqslant \frac{119}{24000}$ is finite. □

We will pursue explicit forms of Theorem 7.6 in the following section.

The statement of Theorem 7.6 is the same as that of Theorem 1.8 except that 'sporadic' has been changed to 'super-sporadic'. The latter is stronger: it gives sporadic CM points on every curve lying

over one of these curves in the modular tower. This includes most but not all 'named' modular curves, an exception being modular curves attached to non-split Cartan subgroups, for which one could prove similar results.

Are there sporadic CM points on all but finitely many modular curves $X(H)$? Indeed not: by [59, Remark 1.3] there are infinitely many subgroups H with $\mathbb{Q}_H = \mathbb{Q}$ and such that the modular curve $X(H)$ is $\mathbb{Q}$-rationally isomorphic to $\mathbb{P}^1$, and thus has no sporadic points whatsoever. The paper [59] gives a complete finite classification of rational subgroups H of prime power level such that $X(H)$ has infinitely many $\mathbb{Q}$-points — equivalently, has no sporadic points. It should be possible to show that as H ranges over rational subgroups of prime power level, all but finitely many of the curves $X(H)$ have sporadic CM points, but perhaps there is a larger natural class of modular curves to consider?

Any two nice curves of genus 0 over a field k are 'twists' of each other, that is, over an algebraic closure of k they each become isomorphic to $\mathbb{P}^1$ and thus to each other. In correspondence with one of us, Sutherland has suggested to try to show that all but finitely many $\overline{\mathbb{Q}}$-isomorphism classes of GL_2-modular curves have sporadic CM points. As he points out, in this family of curves the genus tends to infinity, which helps to make the statement plausible.

We leave these as open questions.

8 | COMPUTATIONS

8.1 | Computing $d_{\mathrm{CM}}(X_0(N))$, $d_{\mathrm{CM}}(X_1(N))$ and $d_{\mathrm{CM}}(X_1(M,N))$

The main result of this section is the explicit computation of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ for all $N \leqslant 10^6$ and of upper bounds on $d_{\mathrm{CM}}(X_1(M,N))$ for all $M \mid N$ with $N \leqslant 100$. The results are recorded in [27]. In the remainder of this section, we describe how the computations were performed.

For an imaginary quadratic order $\mathcal{O}$ of conductor $\mathfrak{f}$ and discriminant $\Delta = \mathfrak{f}^2\Delta_K$, let $h_\Delta = \#\operatorname{Pic}\mathcal{O}$ be the class number of $\mathcal{O}$. We denote by $\mathcal{R}^\circ(\Delta)$ the field obtained by adjoining to $\mathbb{Q}$ the j-invariant of any $\mathcal{O}$-CM elliptic curve $E_{/\mathbb{C}}$. This is a number field that is well-determined up to isomorphism. For positive integers $M \mid N$, following [6, Section 8] we denote by $T^\circ(\mathcal{O}, M, N)$ the least degree $[F : \mathcal{R}^\circ(\Delta)]$ of a number field F over which there is an $\mathcal{O}$-CM elliptic curve E and an injective group homomorphism $\mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \hookrightarrow E(F)$. We write $T^\circ(\mathcal{O}, N)$ for $T^\circ(\mathcal{O}, 1, N)$. Thus, we have

$$d_{\Delta,\mathrm{CM}}(X_1(M,N)) = h_\Delta T^\circ(\mathcal{O}, M, N),$$

and taking $M = 1$, we get

$$d_{\Delta,\mathrm{CM}}(X_1(N)) = h_\Delta T^\circ(\mathcal{O}, N).$$

Thus, we have

$$d_{\mathrm{CM}}(X_1(M,N)) = \min_\Delta h_\Delta T^\circ(\mathcal{O}, M, N)$$

and in particular

$$d_{\mathrm{CM}}(X_1(N)) = \min_\Delta h_\Delta T^\circ(\mathcal{O}, N).$$

A formula for $T^\circ(\mathcal{O}, M, N)$ is given in [6, Section 8], and Theorem 3.7 computes $d_{\Delta,\mathrm{CM}}(X_0(N))$ in terms of $d_{\Delta,\mathrm{CM}}(X_1(N))$. In order to get from this to the computation of $d_{\mathrm{CM}}(X_1(M,N))$ and $d_{\mathrm{CM}}(X_0(N))$ one must solve the minimization problem, and this clearly requires some information about class numbers of imaginary quadratic orders. In describing the computation of $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ it is useful to single out several classes of values of N. First, Table 1 gives all values of N for which $d_{\mathrm{CM}}(X_0(N)) = 1$, along with the corresponding values of $d_{\mathrm{CM}}(X_1(N))$. Henceforth, we suppose that $d_{\mathrm{CM}}(X_0(N)) \geqslant 2$, or equivalently $N \in \mathbb{Z}^+ \setminus \{1, 2, 3, 4, 6, 7, 9, 11, 14, 19, 27, 43, 67, 163\}$.

Suppose N is of Type I, so $N \geqslant 7$. Theorem 3.7 and Corollary 3.11 give

$$d_{\mathrm{CM}}(X_0(N)) = 2, \quad d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{3}.$$

Suppose N is of Type II and not of Type I. Theorem 3.7 and Corollary 3.11 give

$$d_{\mathrm{CM}}(X_0(N)) = 2, \quad d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2}.$$

Finally, suppose that N is neither Type I nor Type II. Then Theorem 3.7 implies that for all discriminants $\Delta < 0$ we have

$$d_{\Delta,\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2} d_{\Delta,\mathrm{CM}}(X_0(N)).$$

From this it follows

$$d_{\mathrm{CM}}(X_1(N)) = \frac{\phi(N)}{2} d_{\mathrm{CM}}(X_0(N))$$

and also that for all $\Delta < 0$ we have

$$\frac{\phi(N)}{2} \mid d_{\Delta,\mathrm{CM}}(X_1(N)).$$

For the same class of N, we put $I^\circ(\Delta, N) := \frac{T^\circ(\mathcal{O}(\Delta),N)}{\phi(N)/2}$, so that

$$d_{\Delta,\mathrm{CM}}(X_0(N)) = h_\Delta I^\circ(\Delta, N).$$

For imaginary quadratic discriminants Δ_1, Δ_2 we have

$$d_{\Delta_1,\mathrm{CM}}(X_1(N)) \leqslant d_{\Delta_2,\mathrm{CM}}(X_1(N)) \iff h_{\Delta_1} I^\circ(\Delta_1, N) \leqslant h_{\Delta_2} I^\circ(\Delta_2, N).$$

Work of Watkins [62] gives all fundamental discriminants $\Delta_K < 0$ of class number at most 100. Using this and (5), we built the list of all discriminants $\Delta < 0$ such that $h_\Delta \leqslant 100$ [27].

Here then is our method to compute $d_{\mathrm{CM}}(X_1(N))$ — or equivalently, to compute $d_{\mathrm{CM}}(X_0(N))$. For N as above, let $\Delta_1 < 0$ be a discriminant which minimizes the quantity $h_\Delta I^\circ(\Delta, N)$ among all discriminants $\Delta < 0$ with $h_\Delta \leqslant 100$. If $h_{\Delta_1} I^\circ(\Delta_1, N) \leqslant 100$ then clearly for any Δ_2 with $h_{\Delta_2} > 100$ we have

$$h_{\Delta_1} I^\circ(\Delta_1, N) < h_{\Delta_2} I^\circ(\Delta_2, N),$$

and it follows that

$$d_{\mathrm{CM}}(X_1(N)) = d_{\Delta_1,\mathrm{CM}}(X_1(N)) \quad \text{and} \quad d_{\mathrm{CM}}(X_0(N)) = d_{\Delta_1,\mathrm{CM}}(X_0(N)).$$

Our method fails for N precisely when $d_{\mathrm{CM}}(X_0(N)) > 100$. By Theorem 1.2 asymptotically 100% of values of N satisfy this condition. However, it turns out that the smallest such N is 50450400, so our method has a sizable range of effectiveness — more than sufficient for the applications to sporadic CM points. We have computed and recorded $d_{\mathrm{CM}}(X_0(N))$ and $d_{\mathrm{CM}}(X_1(N))$ for all $N \leqslant 10^6$, with a total run time of approximately 8.75 days. Among $N \leqslant 10^6$ the largest value of $d_{\mathrm{CM}}(X_0(N))$ is 48, which occurs for $N \in \{277200,\ 554400,\ 831600,\ 932400,\ 956340,\ 985320\}$.

We next consider $X_1(M,N)$: for fixed Δ and $M \mid N$, the quantity $d_{\Delta,\mathrm{CM}}(X_1(M,N))$ is computed in [6, Section 8]. Using Lemma 8.3(c), the upper bound (20) on $d_{\mathrm{CM}}(X_1(M,N))$ and the lower bound $\psi(N) \geqslant N+1$, we find that $X_1(M,N)$ has a sporadic CM point for all $N \geqslant 474059054$. For each $M \mid N$ with $N \leqslant 474059054$, our method to compute $d_{\mathrm{CM}}(X_1(M,N))$ is as follows. Since $d_{\mathrm{CM}}(X_1(M,N)) = 1$ if $N \leqslant 2$, let us assume $N \geqslant 3$. Then Theorem 4.2 implies that for all $\Delta < -4$ we have $\frac{\phi(N)}{2} \mid T^\circ(\mathcal{O},M,N)$. Thus if there is a Δ_1 such that $\frac{d_{\Delta_1,\mathrm{CM}}(X_1(M,N))}{\phi(N)/2} \leqslant 100$, then $d_{\mathrm{CM}}(X_1(M,N))$ can be computed by minimizing $d_{\Delta,\mathrm{CM}}(X_1(M,N))$ over all discriminants $\Delta < 0$ of class number at most 100.

Unfortunately, this method is valid for a much smaller range of pairs (M,N) with $M > 1$. It is valid for all $M \mid N \leqslant 52$, and we record $d_{\mathrm{CM}}(X_1(M,N))$ for these values [27]. For other values, our method yields only an upper bound on $d_{\mathrm{CM}}(X_1(M,N))$. However it turns out that for each pair (M,N) for which we do not know whether $X_1(M,N)$ has a sporadic CM point we are able to determine $d_{\mathrm{CM}}(X_1(M,N))$ and not just bound it above (cf. Table 7).

Remark 8.1. Conditionally on GRH, the list of fundamental imaginary quadratic discriminants of class number at most 9052 is known [33; 41, Corollary 1.3]. So on GRH one could compute $d_{\mathrm{CM}}(X_0(N))$, $d_{\mathrm{CM}}(X_1(N))$ and $d_{\mathrm{CM}}(X_1(M,N))$ for a much larger range of M,N.

8.2 | Sporadic CM points

Here is the main result of this section.

Theorem 8.2.

(a) *For all $N \geqslant 721$, both of the modular curves $X_0(N)$ and $X_1(N)$ have super-sporadic CM points.*
(b) *For all $M \mid N$ with $N \geqslant 8581$, the modular curve $X_1(M,N)$ has super-sporadic CM points.*
(c) *For the 50 values of N listed in Table 2, the modular curve $X_0(N)$* does not *have sporadic CM points.*

TABLE 2 Some N for which $X_0(N)$ has no sporadic CM points

1	2	3	4	5	6	7	8	9	10
12	13	15	16	17	18	20	21	22	23
24	25	26	28	29	30	31	32	33	35
36	37	39	40	41	46	47	48	49	50
53	59	61	65	71	79	83	89	101	131

TABLE 3 All N for which we do not know whether $X_0(N)$ has a sporadic CM point

N	60	70	72	80	87	90	94	96	105	108	110	120	126	132	138
$d_{CM}(X_0(N))$	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
N	140	144	150	152	160	168	174	180	188	190	192	195	204	208	210
$d_{CM}(X_0(N))$	6	6	4	4	4	4	4	6	4	4	4	4	4	4	4
N	216	220	230	231	234	236	238	240	248	252	255	261	264	270	272
$d_{CM}(X_0(N))$	4	6	4	4	4	4	4	4	4	6	4	4	4	4	4
N	276	280	282	285	286	287	288	300	303	304	310	312	315	318	320
$d_{CM}(X_0(N))$	4	6	4	4	4	4	6	6	4	4	4	4	4	4	4
N	324	332	334	336	348	350	357	360	376	380	384	392	395	400	413
$d_{CM}(X_0(N))$	4	4	4	4	6	4	4	8	4	4	4	4	4	4	4
N	416	420	426	429	430	432	435	440	447	455	468	472	476	483	496
$d_{CM}(X_0(N))$	4	8	6	4	4	6	4	8	4	4	6	4	6	4	4
N	501	504	519	524	528	535	558	560	572	576	591	600	623	635	672
$d_{CM}(X_0(N))$	4	8	4	4	8	4	6	6	6	6	4	8	4	4	8
N	720														
$d_{CM}(X_0(N))$	12														

TABLE 4 Some N for which $X_1(N)$ has no sporadic CM points

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$d_{CM}(X_1(N))$	1	1	1	1	2	1	2	4	3	2	5	4	4	3
$\gamma_Q(X_1(N)) \leq$	1	1	1	1	1	1	1	1	1	1	2	1	2	2
N	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$d_{CM}(X_1(N))$	8	8	8	6	6	8	4	10	22	16	10	6	9	12
$\gamma_Q(X_1(N)) \leq$	2	2	4	2	5	3	4	4	7	4	5	6	6	6
N	29	30	32	33	35	36	38	40	42	44	45	46	47	48
$d_{CM}(X_1(N))$	14	8	16	20	24	24	18	16	12	20	24	22	46	32
$\gamma_Q(X_1(N)) \leq$	11	6	8	10	12	8	12	12	12	15	18	19	29	16
N	51	52	54	55	56	59	60	63	64	69	70	71	72	75
$d_{CM}(X_1(N))$	32	24	18	40	24	58	32	36	32	44	48	70	48	40
$\gamma_Q(X_1(N)) \leq$	24	21	18	30	24	46	24	36	32	44	36	66	32	40
N	77	80	81	87	90	94	96	105	108	140	144			
$d_{CM}(X_1(N))$	60	64	54	112	48	92	64	96	72	144	144			
$\gamma_Q(X_1(N)) \leq$	60	48	54	70	48	83	56	96	72	144	128			

(d) *The 106 values of N listed in Table 3 include all values for which we do not know whether* $X_0(N)$ *has sporadic CM points.*

(e) *For the 67 values of N listed in Table 4, the modular curve* $X_1(N)$ does not *have sporadic CM points.*

(f) *The 227 values of N listed in Table 5 include all values for which we do not know whether* $X_1(N)$ *has sporadic CM points.*

(g) *For the 37 pairs* (M, N) *with* $M \geqslant 2$ *listed in Table 6, the modular curve* $X_1(M, N)$ does not *have sporadic CM points.*

TABLE 5 All N for which we do not know whether $X_1(N)$ has a sporadic CM point

N	37	41	43	49	50	53	57	58	61	62	65	66	67	68	73
$d_{CM}(X_1(N))$	12	20	14	14	10	26	12	14	20	30	24	20	22	32	24
N	74	76	78	79	82	83	84	85	86	88	89	91	92	93	95
$d_{CM}(X_1(N))$	18	36	24	26	20	82	24	32	42	40	44	24	44	20	72
N	97	98	99	100	101	102	103	104	106	107	109	110	112	113	114
$d_{CM}(X_1(N))$	32	42	60	40	50	32	34	48	26	106	36	80	48	56	36
N	115	116	117	118	119	120	121	122	123	124	125	126	127	128	131
$d_{CM}(X_1(N))$	88	56	72	58	96	64	110	30	80	60	50	72	42	64	130
N	132	134	135	136	137	138	141	142	143	145	148	149	150	152	153
$d_{CM}(X_1(N))$	80	66	72	64	68	88	92	70	120	56	72	74	80	144	96
N	154	155	156	158	159	160	161	162	164	165	166	167	168	171	172
$d_{CM}(X_1(N))$	60	120	48	78	104	128	132	54	80	80	82	166	96	108	84
N	173	174	175	176	177	179	180	182	184	186	187	188	189	190	191
$d_{CM}(X_1(N))$	86	112	120	80	116	178	144	72	88	60	160	184	108	144	190
N	192	195	196	197	200	203	204	206	207	208	209	210	212	213	214
$d_{CM}(X_1(N))$	128	192	84	98	80	168	128	102	132	192	180	96	104	140	106
N	215	216	220	224	225	227	230	231	232	234	235	236	238	239	240
$d_{CM}(X_1(N))$	168	144	240	96	120	226	176	240	112	144	184	232	192	238	128
N	242	243	244	245	248	249	251	252	253	254	255	256	261	262	263
$d_{CM}(X_1(N))$	110	162	120	168	240	164	250	216	220	126	256	128	336	130	262
N	264	267	270	272	275	276	279	280	282	285	286	287	288	295	299
$d_{CM}(X_1(N))$	160	176	144	256	200	176	180	288	184	288	240	480	288	232	264
N	300	303	304	310	311	312	315	318	319	320	323	324	329	332	334
$d_{CM}(X_1(N))$	240	400	288	240	310	192	288	208	280	256	288	216	276	328	332
N	336	341	347	348	350	357	359	360	376	380	383	384	392	395	400
$d_{CM}(X_1(N))$	192	300	346	336	240	384	358	384	368	288	382	256	336	624	320
N	413	416	420	426	429	430	432	435	440	447	455	468	472	476	483
$d_{CM}(X_1(N))$	696	384	384	420	480	336	432	448	640	592	576	432	464	576	528
N	496	501	504	519	524	528	535	558	560	572	576	591	600	623	635
$d_{CM}(X_1(N))$	480	664	576	688	520	640	848	540	576	720	576	784	640	1056	1008
tN	672	720													
$d_{CM}(X_1(N))$	768	1152													

(h) *The 146 pairs (M,N) listed in Table 7 include all pairs with $M \geqslant 2$ for which we do not know whether $X_1(M,N)$ has sporadic CM points.*

The proof is of course quite computational. We break it up into a sequence of smaller results.

Lemma 8.3. *Let $N \geqslant 3$ and $M \mid N$.*

(a) *A closed point on $X_0(N)_{/\mathbb{Q}}$ of degree at most $\frac{119}{24000}\psi(N)$ is super-sporadic.*
(b) *A closed point on $X_1(N)_{/\mathbb{Q}}$ of degree at most $\frac{119}{48000}\phi(N)\psi(N)$ is super-sporadic.*

TABLE 6 Some (M,N) with $M \geqslant 2$ for which $X_1(M,N)$ has no sporadic CM points

(M,N)	$d_{\mathrm{CM}}(X_1(M,N))$	(M,N)	$d_{\mathrm{CM}}(X_1(M,N))$
(2, 2)	1	(4, 16)	16
(3, 3)	2	(2, 18)	12
(2, 4)	2	(3, 18)	12
(4, 4)	4	(2, 20)	8
(5, 5)	8	(2, 22)	10
(2, 6)	2	(2, 24)	16
(3, 6)	4	(3, 24)	32
(6, 6)	6	(2, 26)	12
(2, 8)	4	(3, 27)	36
(4, 8)	8	(2, 28)	12
(3, 9)	6	(2, 30)	16
(2, 10)	4	(2, 32)	16
(2, 12)	8	(2, 36)	24
(3, 12)	12	(2, 40)	32
(4, 12)	8	(2, 48)	32
(6, 12)	16	(2, 54)	36
(2, 14)	6	(2, 70)	72
(3, 15)	16	(2, 72)	72
(2, 16)	8		

(c) *A closed point on $X_1(M,N)_{/\mathbb{Q}}$ of degree at most $\frac{119}{48000}M\phi(M)\phi(N)\psi(N)$ is super-sporadic.*

Proof. This is an immediate consequence of Theorem 7.3 and Lemma 4.1. □

Lemma 8.4. *For all $N \geqslant 102641930$, the curves $X_0(N)_{/\mathbb{Q}}$ and $X_1(N)_{/\mathbb{Q}}$ have super-sporadic CM points.*

Proof. We compute this for $X_0(N)$ using Lemma 8.3(a), the upper bound (19) on $d_{\mathrm{CM}}(X_0(N))$ and $\psi(N) \geqslant N+1$. Since every preimage in $X_1(N)$ of a super-sporadic CM point on $X_0(N)$ is super-sporadic, the result follows for $X_1(N)$. □

Lemma 8.5. *For a positive integer N, let $D(N)$ be the minimum of $|\Delta|$ as Δ ranges over imaginary quadratic discriminants satisfying $(\frac{\Delta}{p}) = 1$ for all primes $p \mid N$. (Thus N and $K = \mathbb{Q}(\sqrt{-D(N)})$ satisfy the 'Heegner hypothesis'.) Let $\mathcal{E}$ be the set of positive integers $N \leqslant 102641930$ such that*

$$d_{-D(N),\mathrm{CM}}(X_0(N)) > \frac{119}{24000}\psi(N).$$

Then $\#\mathcal{E} = 689$ and the largest element of $\mathcal{E}$ is 4290.

Proof. By direct computation. □

TABLE 7 All pairs (M, N) with $M \geqslant 2$ for which we do not know whether $X_1(M, N)$ has a sporadic CM point

(M,N)	$(7,7)$	$(8,8)$	$(9,9)$	$(5,10)$	$(10,10)$	$(11,11)$	$(12,12)$
$d_{\mathrm{CM}}(X_1(M,N))$	12	8	18	8	16	40	24
(M,N)	$(7,14)$	$(14,14)$	$(5,15)$	$(8,16)$	$(6,18)$	$(9,18)$	$(4,20)$
$d_{\mathrm{CM}}(X_1(M,N))$	24	36	16	32	18	36	16
(M,N)	$(5,20)$	$(10,20)$	$(3,21)$	$(4,24)$	$(6,24)$	$(12,24)$	$(5,25)$
$d_{\mathrm{CM}}(X_1(M,N))$	32	32	12	16	32	48	40
(M,N)	$(4,28)$	$(7,28)$	$(3,30)$	$(5,30)$	$(6,30)$	$(4,32)$	$(8,32)$
$d_{\mathrm{CM}}(X_1(M,N))$	24	72	16	48	32	32	64
(M,N)	$(3,33)$	$(2,34)$	$(3,36)$	$(4,36)$	$(6,36)$	$(2,38)$	$(3,39)$
$d_{\mathrm{CM}}(X_1(M,N))$	40	16	36	48	48	18	24
(M,N)	$(4,40)$	$(5,40)$	$(2,42)$	$(3,42)$	$(6,42)$	$(2,44)$	$(4,44)$
$d_{\mathrm{CM}}(X_1(M,N))$	32	64	12	36	36	20	40
(M,N)	$(3,45)$	$(5,45)$	$(2,46)$	$(3,48)$	$(4,48)$	$(6,48)$	$(2,50)$
$d_{\mathrm{CM}}(X_1(M,N))$	48	96	22	64	64	64	20
(M,N)	$(3,51)$	$(2,52)$	$(4,52)$	$(3,54)$	$(6,54)$	$(5,55)$	$(2,56)$
$d_{\mathrm{CM}}(X_1(M,N))$	64	24	48	36	72	160	24
(M,N)	$(4,56)$	$(2,58)$	$(2,60)$	$(3,60)$	$(4,60)$	$(2,62)$	$(2,64)$
$d_{\mathrm{CM}}(X_1(M,N))$	48	28	32	64	64	30	32
(M,N)	$(4,64)$	$(2,66)$	$(2,68)$	$(3,69)$	$(3,72)$	$(4,72)$	$(6,72)$
$d_{\mathrm{CM}}(X_1(M,N))$	64	40	32	88	96	96	144
(M,N)	$(2,74)$	$(3,75)$	$(2,76)$	$(2,78)$	$(2,80)$	$(4,80)$	$(3,81)$
$d_{\mathrm{CM}}(X_1(M,N))$	36	80	72	24	64	128	108
(M,N)	$(2,82)$	$(2,84)$	$(3,84)$	$(2,86)$	$(3,87)$	$(2,88)$	$(2,90)$
$d_{\mathrm{CM}}(X_1(M,N))$	40	48	72	42	112	40	72
(M,N)	$(3,90)$	$(2,92)$	$(2,94)$	$(2,96)$	$(3,96)$	$(4,96)$	$(2,98)$
$d_{\mathrm{CM}}(X_1(M,N))$	96	44	92	64	128	128	42
(M,N)	$(2,100)$	$(2,102)$	$(2,104)$	$(3,105)$	$(2,106)$	$(2,108)$	$(3,108)$
$d_{\mathrm{CM}}(X_1(M,N))$	40	64	96	192	52	72	144
(M,N)	$(2,110)$	$(2,112)$	$(2,116)$	$(2,118)$	$(2,120)$	$(2,122)$	$(2,124)$
$d_{\mathrm{CM}}(X_1(M,N))$	120	48	56	116	64	60	120
(M,N)	$(2,126)$	$(2,128)$	$(2,132)$	$(2,136)$	$(2,138)$	$(2,140)$	$(4,140)$
$d_{\mathrm{CM}}(X_1(M,N))$	108	64	80	128	88	144	288
(M,N)	$(2,144)$	$(3,144)$	$(4,144)$	$(2,150)$	$(2,152)$	$(2,156)$	$(2,160)$
$d_{\mathrm{CM}}(X_1(M,N))$	144	288	288	120	144	96	128
(M,N)	$(2,162)$	$(2,166)$	$(2,168)$	$(2,174)$	$(2,180)$	$(2,188)$	$(2,190)$
$d_{\mathrm{CM}}(X_1(M,N))$	108	164	96	168	192	184	144
(M,N)	$(2,192)$	$(2,196)$	$(2,200)$	$(2,208)$	$(2,210)$	$(2,216)$	$(2,220)$
$d_{\mathrm{CM}}(X_1(M,N))$	128	168	160	192	192	216	320
(M,N)	$(2,234)$	$(2,236)$	$(2,238)$	$(2,248)$	$(2,252)$	$(2,262)$	$(2,264)$
$d_{\mathrm{CM}}(X_1(M,N))$	216	232	288	240	288	260	320
(M,N)	$(2,280)$	$(2,286)$	$(2,288)$	$(2,300)$	$(2,336)$	$(2,360)$	
$d_{\mathrm{CM}}(X_1(M,N))$	288	360	288	320	384	576	

Lemmas 8.3 and 8.5 imply that $X_0(N)$ and $X_1(N)$ have super-sporadic CM points for all $N \in \mathbb{Z}^+ \setminus \mathcal{E}$.

Lemma 8.6. *Let $\mathcal{F}_0$ be the subset of $\mathcal{E}$ consisting of N for which $d_{\mathrm{CM}}(X_0(N)) \geqslant \frac{119}{24000}\psi(N)$, and let $\mathcal{F}_1$ be the subset of $\mathcal{E}$ consisting of N for which $d_{\mathrm{CM}}(X_1(N)) \geqslant \frac{119}{48000}\phi(N)\psi(N)$.*

(a) *We have*

$$\begin{aligned}\mathcal{F}_0 = {} & [1,197] \cup [199,221] \cup [223,227] \cup [229,245] \cup [247,249] \cup [251,257] \cup \{259\} \\ & \cup [261,265] \cup \{267\} \cup [269,272] \cup [275,277] \cup [279,283] \cup [285,289] \\ & \cup \{291, 293, 295, 299, 300, 301, 303, 304, 305, 307, 310, 311, 312, 313, 315, 317, 318, 319, 320, 323, \\ & 324, 329, 331, 332, 334, 336, 337, 341, 343, 347, 348, 349, 350, 353, 357, 359, 360, 361, 367, 373, 376, \\ & 379, 380, 383, 384, 389, 392, 395, 397, 400, 401, 413, 416, 420, 426, 429, 430, 432, 435, 440, 447, 455, \\ & 468, 472, 476, 483, 496, 501, 504, 519, 524, 528, 535, 558, 560, 572, 576, 591, 600, 623, 635, 672, 720\}.\end{aligned}$$

(b) *We have*

$$\begin{aligned}\mathcal{F}_1 = {} & [1,110] \cup [112,128] \cup [131,132] \cup [134,138] \cup [140,145] \cup [148,150] \cup [152,156] \\ & \cup [158,162] \cup [164,168] \cup [171,177] \cup [179,180] \cup \{182,184\} \cup [186,192] \\ & \cup [195,197] \cup \{200\} \cup [203,204] \cup [206,210] \cup [212,216] \cup \{220, 224, 225, 227, 230 \\ & 231, 232, 234, 235, 236, 238, 239, 240, 242, 243, 244, 245, 248, 249, 251, 252, 253, 254, 255, 256, \\ & 261, 262, 263, 264, 267, 270, 272, 275, 276, 279, 280, 282, 285, 286, 287, 288, 295, 299, 300, 303, \\ & 304, 310, 311, 312, 315, 318, 319, 320, 323, 324, 329, 332, 334, 336, 341, 347, 348, 350, 357, 359, \\ & 360, 376, 380, 383, 384, 392, 395, 400, 413, 416, 420, 426, 429, 430, 432, 435, 440, 447, 455, 468, \\ & 472, 476, 483, 496, 501, 504, 519, 524, 528, 535, 558, 560, 572, 576, 591, 600, 623, 635, 672, 720\}.\end{aligned}$$

Proof. By direct computation. □

Remark 8.7. If $N \geqslant 7$ and is neither Type I nor Type II, then we have

$$\frac{d_{\mathrm{CM}}(X_1(N))}{d_{\mathrm{CM}}(X_0(N))} = \frac{\phi(N)}{2} = \deg(X_1(N) \to X_0(N))$$

and thus $N \in \mathcal{F}_0 \iff N \in \mathcal{F}_1$. However, there are 62 values N of Type I or Type II for which super-sporadic CM points exist on $X_1(N)$ but not on $X_0(N)$.

Lemma 8.8.

(a) *We have $\delta(X_0(N)) = 1 \iff N \in \{1,2,3,4,5,6,7,8,9,10,12,13,16,18,25\}$. For these values of N, $X_0(N)$ has no sporadic points.*

(b) *We have* $\delta(X_1(N)) = 1 \iff N \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12\}$. *For these values of* N, $X_1(N)$ *has no sporadic points.*

(c) *We have* $\delta(X_0(N)) = 2 \iff$

$$N \in \{11, 14, 15, 17, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37,$$
$$39, 40, 41, 43, 46, 47, 48, 49, 50, 53, 59, 61, 65, 71, 79, 83, 89, 101, 131\}.$$

(d) *We have* $\delta(X_1(N)) = 2 \iff N \in \{11, 13, 14, 15, 16, 18\}$.

Proof. (a) and (b) For a nice curve defined over a number field $X_{/F}$, we have $\delta(X) = 1$ if and only if X is F-rationally isomorphic to $\mathbb{P}^1$ or to an elliptic curve with positive rank. Since the cusp at ∞ is a $\mathbb{Q}$-rational point on $X_1(N)$ and $X_0(N)$, these curves are isomorphic to the projective line if and only if they have genus 0 and are elliptic curves if and only if they have genus 1. The curve $X_0(N)$ (respectively $X_1(N)$) has genus 1 if and only if $N \in \{11, 14, 15, 17, 19, 20, 21, 24, 27, 32, 36, 49\}$ (respectively if and only if $N \in \{11, 14, 15\}$), but in all these cases the Mordell-Weil rank is 0, so the curves have $\delta = 2$. The curve $X_0(N)$ has genus 0 if and only if $N \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 16, 18, 25\}$, giving part (a), and the curve $X_1(N)$ has genus 0 if and only if $N \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12\}$, giving part (b).

(c) and (d) By work of Abramovich–Harris [2], a nice curve X defined over a number field has $\delta(X) = 2$ if and only if it is genus 0 with no F-rational points, is an elliptic curve with finitely many rational points, is hyperelliptic with genus at least 2, or has genus at least two and admits a degree 2 F-rational morphism to an elliptic curve with infinitely many F-rational points. Using this, Bars [4, Theorem 4.3] computed all N such that $X_0(N)$ has genus at least 2 and $\delta(X_0(N)) = 2$, and Jeon–Kim [34, Theorem 4.2] computed all N such that $\delta(X_1(N)) \leqslant 2$. □

Corollary 8.9.

(a) *For* $N \in \{11, 14, 19, 27, 43, 67, 163\}$ *we have* $d_{\mathrm{CM}}(X_0(N)) = 1 < \delta(X_0(N))$, *and thus* $X_0(N)$ *has a sporadic CM point.*

(b) *For*

$$N \in \{15, 17, 20, 21, 22, 23, 24, 26, 28, 29, 30, 31, 32, 33,$$
$$35, 36, 37, 39, 40, 41, 46, 47, 48, 49, 50, 53, 59, 61, 65, 71, 79, 83, 89, 101, 131\}$$

we have $d_{\mathrm{CM}}(X_0(N)) \geqslant 2 = \delta(X_0(N))$, *so* $X_0(N)$ does not *have sporadic CM points.*

(c) *For* $N \in \mathcal{F}_0$, *we have* $d_{\mathrm{CM}}(X_0(N)) = 2 < \delta(X_0(N))$ *iff*

$$N \in \{34, 38, 42, 44, 45, 51, 52, 54, 55, 56, 57, 58, 62, 63, 64, 66, 68, 69, 73, 74, 75, 76, 77, 78,$$
$$81, 82, 84, 85, 86, 88, 91, 92, 93, 95, 97, 98, 99, 100, 102, 103, 104, 106, 107, 109, 111, 112,$$
$$113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 125, 127, 128, 129, 130, 133, 134,$$
$$135, 136, 137, 139, 141, 142, 143, 145, 146, 147, 148, 149, 151, 153, 154, 155, 156, 157,$$
$$158, 159, 161, 162, 164, 165, 166, 167, 169, 170, 171, 172, 173, 175, 176, 177, 178, 179,$$

181, 182, 183, 184, 185, 186, 187, 189, 191, 193, 194, 196, 197, 199, 200, 201, 202, 203, 205, 206, 207, 209, 211, 212, 213, 214, 215, 217, 218, 219, 221, 223, 224, 225, 226, 227, 229, 232, 233, 235, 237, 239, 241, 242, 243, 244, 245, 247, 249, 251, 253, 254, 256, 257, 259, 262, 263, 265, 267, 269, 271, 275, 277, 279, 281, 283, 289, 291, 293, 295, 299, 301, 305, 307, 311, 313, 317, 319, 323, 329, 331, 337, 341, 343, 347, 349, 353, 359, 361, 367, 373, 379, 383, 389, 397, 401}.

Thus, for these values of N, the curve $X_0(N)$ has a sporadic CM point.

Proof. By Lemma 8.8 and direct computation. □

Derickx and van Hoeij [24] give upper bounds on $\gamma_{\mathbb{Q}}(X_1(N))$ for $N \leqslant 250$. Since sporadic points have low degree, if $d_{\mathrm{CM}}(X_1(N))$ is at least their upper bound on $\gamma_{\mathbb{Q}}(X_1(N))$, then $X_1(N)$ cannot have sporadic CM points. Of the 296 numbers $N \in \mathcal{F}_1$, there are 67 values of $N \leqslant 250$ for which this is the case; these are listed in Table 4, along with their corresponding $d_{\mathrm{CM}}(X_1(N))$ values and upper bounds on $\gamma_{\mathbb{Q}}(X_1(N))$ from [24].

In fact, for $N \leqslant 40$ the referenced upper bounds are known to be the exact gonality values, which allows us to make a few additional certifications of sporadic CM points on $X_1(N)$.

Lemma 8.10. *For $N \in \{31, 34, 39\}$, the modular curve $X_1(N)$ has sporadic CM points.*

Proof. Let $N \in \{31, 34, 39\}$, and let $J_1(N)_{/\mathbb{Q}}$ be the Jacobian abelian variety of the modular curve $X_1(N)_{/\mathbb{Q}}$. By [23, Theorem 3.1] we have $J_1(N)(\mathbb{Q})$ is finite. From this it follows — see, for example, [9, Theorem 4.2] — that $\delta(X_1(N)) = \gamma_{\mathbb{Q}}(X_1(N))$. Comparing the work of [24] to our own calculations, we find that $d_{\mathrm{CM}}(X_1(N)) < \gamma_{\mathbb{Q}}(X_1(N))$. □

For the remaining 227 values of N, which are recorded in Table 5, we do not yet know whether $X_1(N)$ has a sporadic CM point.

In our computations for $d_{\mathrm{CM}}(X_1(M,N))$, we find based on Lemma 8.3 that there are 183 pairs (M,N) with $M > 1$ such that $X_1(M,N)$ may fail to have a sporadic CM point. Our computations of $d_{\mathrm{CM}}(X_1(M,N))$ for all such pairs are exact — that is, these modular curves are minimized by an order of class number at most 100. Thus, for some of these pairs we can show that there are no sporadic CM points on $X_1(M,N)$, as follows: using the $\mathbb{Q}$-morphism $X_1(M,N) \to X_1(N)$, we get

$$\gamma_{\mathbb{Q}}(X_1(M,N)) \leqslant \deg(X_1(M,N) \to X_1(N)) \cdot \gamma_{\mathbb{Q}}(X_1(N)) = M\phi(M) \cdot \gamma_{\mathbb{Q}}(X_1(N)),$$

and it follows that if

$$d_{\mathrm{CM}}(X_1(M,N)) \geqslant M\phi(M) \cdot \gamma_{\mathbb{Q}}(X_1(N)),$$

then $X_1(M,N)$ has no sporadic CM points. Using this and the upper bounds on $\gamma_{\mathbb{Q}}(X_1(N))$ in [24], we find that $X_1(M,N)$ has no sporadic CM points for 31 pairs (M,N).

Additionally, for the pairs (2, 2), (3, 3), (4, 4), (5, 5), and (3,6) the curves $X_1(M,N)_{/\mathbb{Q}(\zeta_M)}$ are of genus 0 with infinitely many rational points. By [35, Theorem 3.2] we have $\delta(X_1(6,6)_{/\mathbb{Q}}) = 4$, while our computations give $\mathrm{d}_{\mathrm{CM}}(X_1(6,6)) = 6$. We then have 37 total pairs (M,N) with $M > 1$ for which we know we have no sporadic CM points, and we list these in Table 6. For the remaining 146 pairs (M,N) with $M > 1$ (listed in Table 7), we do not yet know whether $X_1(M,N)$ has a sporadic CM point.

Example 8.11. Consider the modular curve $X_1(450)$. In order to use Lemma 8.3 to produce a super-sporadic CM point on $X_1(450)$, we would need

$$d_{\mathrm{CM}}(X_1(450)) \leqslant \left\lfloor \frac{119}{48000}\phi(450)\psi(450) \right\rfloor = 321.$$

If Δ is an imaginary quadratic discriminant such that every prime divisor of $450 = 2 \cdot 3^2 \cdot 5^2$ splits in $\mathcal{O}(\Delta)$, then ($\Delta < -4$ and thus) $d_{\Delta,\mathrm{CM}}(450) = \phi(450)h_\Delta = 120h_\Delta$. Thus such a Δ gives us a super-sporadic CM point if and only if $h_\Delta \leqslant 2$. However, in none of the 27 imaginary quadratic fields of class number 1 or 2 do the primes 2, 3 and 5 all split. Moreover, minimizing $d_{\Delta,\mathrm{CM}}(X_1(450))$ among all *maximal* orders of class number at most 100 only gives

$$d_{\mathrm{CM}}(X_1(450)) \leqslant 360.$$

In fact

$$d_{\mathrm{CM}}(X_1(450)) = 240,$$

with minimizing discriminant $\Delta = -36 = 3^2 \cdot \Delta_{\mathbb{Q}(\sqrt{-1})}$. Equivalently we have

$$d_{\mathrm{CM}}(X_0(450)) = d_{-36,\mathrm{CM}}(X_0(450)) = 4,$$

coming from the fact that $h_{-36} = 2$ so the ring class field F of $\mathbb{Q}(\sqrt{-1})$ of conductor 2 is a number field of degree 4 over which there is an $\mathcal{O}(-36)$-CM elliptic curve with an F-rational cyclic 450-isogeny (cf. [5, Theorem 6.18]).

ACKNOWLEDGEMENTS

We are grateful to Drew Sutherland for his interesting suggestions recorded at the end of Section 7. We thank the referee for the careful work and for contributing Lemma 8.10.

Partial support for the second and fourth author was provided by the Research and Training Group grant DMS-1344994 funded by the National Science Foundation. The second author is also supported in part by the National Science Foundation Graduate Research Fellowship under Grant Number: 1842396. The third author is partially supported by the National Science Foundation grant DMS-2001581.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and

mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. D. Abramovich, *A linear lower bound on the gonality of modular curves*, Internat. Math. Res. Notices (1996), no. 20, 1005–1011.
2. D. Abramovich and J. Harris, *Abelian varieties and curves in $W_d(C)$*, Compos. Math. **78** (1991), 227–238.
3. W. D. Banks, F. Luca, F. Saidak, and I. E. Shparlinski, *Values of arithmetical functions equal to a sum of two squares*, Q. J. Math. **56** (2005), 123–139.
4. F. Bars, *Bielliptic modular curves*, J. Number Theory **76** (1999), 154–165.
5. A. Bourdon and P. L. Clark, *Torsion points and Galois representations on CM elliptic curves*, Pacific J. Math. **305** (2020), 43–88.
6. A. Bourdon and P. L. Clark, *Torsion points and isogenies on CM elliptic curves*, J. Lond. Math. Soc. (2) **102** (2020), 580–622.
7. A. Bourdon, P. L. Clark, and P. Pollack, *Anatomy of torsion in the CM case*, Math. Z. **285** (2017), 795–820.
8. A. Bourdon, P. L. Clark, and J. Stankewicz, *Torsion points on CM elliptic curves over real number fields*, Trans. Amer. Math. Soc. **369** (2017), 8457–8496.
9. A. Bourdon, O. Ejder, Y. Liu, F. Odumodu, and B. Viray, *On the level of modular curves that give rise to sporadic j-invariants*, Adv. Math. **357** (2019), 106824, 33 pp.
10. A. Bourdon and P. Pollack, *Torsion subgroups of CM elliptic curves over odd degree number fields*, Int. Math. Res. Not. **2017**, 4923–4961.
11. F. Breuer, *Torsion bounds for elliptic curves and Drinfeld modules*, J. Number Theory **130** (2010), 1241–1250.
12. D. A. Burgess, *The distribution of quadratic residues and non-residues*, Mathematika **4** (1957), 106–112.
13. M. Chou, P. L. Clark, and M. Milosevic, *Acyclotomy of torsion in the CM case*, Ramanujan J. **55** (2021), no. 3, 1015–1037.
14. P. L. Clark, *On the Hasse principle for Shimura curves*, Israel J. Math. **171** (2009), 349–365.
15. P. L. Clark, *CM elliptic curves: volcanoes, reality and applications*. http://alpha.math.uga.edu/~pete/Isogenies.pdf
16. P. L. Clark, B. Cook, and J. Stankewicz, *Torsion points on elliptic curves with complex multiplication (with an appendix by Alex Rice)*, Int. J. Number Theory **9** (2013), 447–479.
17. P. L. Clark, P. Corn, A. Rice, and J. Stankewicz, *Computation on elliptic curves with complex multiplication*, LMS J. Comput. Math. **17** (2014), no. 1, 509–535.
18. P. L. Clark, M. Milosevic, and P. Pollack, *Typically bounding torsion*, J. Number Theory **192** (2018), 150–167.
19. P. L. Clark and P. Pollack, *The truth about torsion in the CM case*, C. R. Math. Acad. Sci. Paris **353** (2015), 683–688.
20. P. L. Clark and P. Pollack, *The truth about torsion in the CM case, II*, Q. J. Math. **68** (2017), 1313–1333.
21. D. Cox, *Primes of the form $x^2 + ny^2$. Fermat, class field theory and complex multiplication*, John Wiley & Sons, New York, 1989.
22. P. Deligne and M. Rapoport, Les schémas de modules de courbes elliptiques, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972). Lecture Notes in Mathematics, vol. 349 (Springer, Berlin, 1973), pp. 143–316.
23. M. Derickx, A. Etropolski, M. van Hoeij, J. S. Morrow, and D. Zureick-Brown, *Sporadic cubic torsion*, Algebra Number Theory, to appear.
24. M. Derickx and M. van Hoeij, *Gonality of the modular curve $X_1(N)$*, J. Algebra **417** (2014), 52–71.
25. G. Frey, *Curves with infinitely many points of fixed degree*, Israel J. Math. **85** (1994), 79–83.
26. V. R. Fridlender, *On the least nth-power non-residue*, Dokl. Akad. Nauk SSSR **66** (1949), 351–352.
27. T. Genao and F. Saia, *Least CM degree repository*. https://github.com/fsaia/least-cm-degree.
28. H. Halberstam and H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs 4, Academic Press, London-New York, 1974.
29. R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge University Press, Cambridge, 1988.
30. G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Q. J. Math. **48** (1917), 76–92.

31. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 6th ed., Oxford University Press, Oxford, 2008.
32. M. Hindry and J. Silverman, *Sur le nombre de points de torsion rationnels sur une courbe elliptique*, C. R. Acad. Sci. Paris Sér. I Math. **329** (1999), 97–100.
33. M. J. Jacobson, Jr., S. Ramachandran, and H. C. Williams, Numerical results on class groups of imaginary quadratic fields, Algorithmic number theory, Lecture Notes in Computer Science 4076 (Springer, Berlin, 2006), pp. 87–101.
34. D. Jeon and C. H. Kim, *Bielliptic modular curves $X_1(N)$*, Acta Arith. **112** (2004), 75–86.
35. D. Jeon, C. H. Kim, and E. Park, *On the torsion of elliptic curves over quartic number fields*, J. Lond. Math. Soc. **74** (2006), 1–12.
36. D. Jeon, C. H. Kim, and A. Schweizer, *On the torsion of elliptic curves over cubic number fields*, Acta Arith. **113** (2004), 291–301.
37. M. A. Kenku, *On the modular curves $X_0(125)$, $X_1(25)$ and $X_1(49)$*, J. Lond. Math. Soc. (2) **23** (1981), 415–427.
38. H. H. Kim, *Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2*, J. Amer. Math. Soc. **16** (2003), 139–183 (with appendix 1 by D. Ramakrishnan and appendix 2 by H. H. Kim and P. Sarnak).
39. S. Kwon, *Degree of isogenies of elliptic curves with complex multiplication*, J. Korean Math. Soc. **36** (1999), 945–958.
40. J. C. Lagarias and A. M. Odlyzko, Effective versions of the Chebotarev density theorem, Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975), Academic Press, London, 1977, pp. 409–464.
41. Y. Lamzouri, X. Li, and K. Soundararajan, *Conditional bounds for the least quadratic non-residue and related problems*, Math. Comp. **84** (2015), 2391–2412.
42. E. Landau, *Lösung des Lehmer'schen Problems*, Amer. J. Math. **31** (1909), 86–102.
43. U. V. Linnik, *On the least prime in an arithmetic progression. I. The basic theorem*, Mat. Sbornik N.S. **15** (1944), no. 57, 139–178.
44. J. E. Littlewood, *On the class number of the corpus $P(\sqrt{-k})$*, Proc. Lond. Math. Soc. **27** (1928), 358–372.
45. B. Mazur, *Rational isogenies of prime degree*, Invent. Math. **44** (1978), 129–162 (with an appendix by D. Goldfeld).
46. L. Merel, *Bornes pour la torsion des courbes elliptiques sur les corps de nombres*, Invent. Math. **124** (1996), 437–449.
47. H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97 (Cambridge University Press, Cambridge, 2007).
48. K. K. Norton, *A character-sum estimate and applications*, Acta Arith. **85** (1998), 51–78.
49. J. L. Parish, *Rational torsion in complex-multiplication elliptic curves*, J. Number Theory **33** (1989), 257–265.
50. K. Prachar, *Über die kleinste quadratfreie Zahl einer arithmetischen Reihe*, Monatsh. Math. **62** (1958), 173–176.
51. H. Salié, *Über den kleinsten positiven quadratischen Nichtrest nach einer Primzahl*, Math. Nachr. **3** (1949), 7–8.
52. A. Schinzel, *On pseudosquares*, New Trends in Probability and Statistics, vol. 4 (Palanga, 1996), pp. 213–220. VSP, Utrecht, 1997.
53. J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), 259–331.
54. J.-P. Serre, *Divisibilité de certaines fonctions arithmétiques*, Enseignement Math. (2) **22** (1976), 227–260.
55. C. Siegel, *Über die Klassenzahl quadratischer Zahlkörper*, Acta Arith. **1** (1935), 83–86.
56. A. Silverberg, *Torsion points on abelian varieties of CM-type*, Compos. Math. **68** (1988), 241–249.
57. P. Stevenhagen, *Hilbert's 12th problem, complex multiplication and Shimura reciprocity*, Class field theory – its centenary and prospect (Tokyo, 1998), pp. 161–176, Advanced Studies in Pure Mathematics, vol. 30, Math. Soc. Japan, Tokyo, 2001.
58. A. V. Sutherland, *Torsion subgroups of elliptic curves over number fields*. Online lecture notes: https://math.mit.edu/~drew/MazursTheoremSubsequentResults.pdf
59. A. V. Sutherland and D. Zywina, *Modular curves of prime-power level with infinitely many rational points*, Algebra Number Theory **11** (2017), 1199–1229.
60. G. Tenenbaum, *A rate estimate in Billingsley's theorem for the size distribution of large prime factors*, Q. J. Math. **51** (2000), 385–403.
61. L. C. Washington, *Introduction to cyclotomic fields*, 2nd ed. Graduate Texts in Mathematics, vol. 83, Springer, New York, 1997.
62. M. Watkins, *Class numbers of imaginary quadratic fields*, Math. Comp. **73** (2004), 907–938.