# DISTRIBUTION OF THE 2-SELMER RANK UNDER TWISTING

Version of 18 Jul 2021.

MARK WATKINS

Abstract. Consider the elliptic curve $y^2 = x^3 - x$ (which is associated to congruent numbers). Heath-Brown described the 2-Selmer rank distribution of the quadratic twists of this curve over the set of odd squarefree integers, using methods involving computing the moments of the 2-Selmer rank via an analysis of quadratic residue symbols (the appear in a Monsky matrix).

Swinnerton-Dyer then investigated the distribution of the 2-Selmer rank under twisting for a wider family (elliptic curves with full 2-torsion and no 4-torsion), using a Markov chain analysis to show the expected distribution – however, his result used an unnatural ordering of the integers, namely by the number of prime factors. This was remedied by Kane, who showed the result under the natural ordering via methodology similar to Heath-Brown's.

More recently, Smith (as part of a larger work) has shown the same result using Swinnerton-Dyer's method, essentially showing that the input (involving quadratic residue symbols of the prime divisors of the twist factor) to his Markov chain analysis can be shown to have the expected equi-distribution under the natural ordering. We give an exposition of Smith's work, with an explicit (and effective) error bound. We also discuss the related problem of the 4-rank of quadratic class groups, initially done by Fouvry and Klüners.

## 1. Introduction

Let $E/\mathbf{Q}$ be an elliptic curve with full 2-torsion, writing it in an integral model as $y^2 = (x - c_1)(x - c_2)(x - c_3)$. We can assume that $c_1 < c_2 < c_3$ and $c_2 = 0$ by variable transformations. Moreover, we can assume that there is no nontrivial common square factor of the $c_i$, as else this could be divided out to give a simpler model. We write $\delta_{ij} = c_i - c_j$, and will also assume that none of the quantities given by $\delta_{ij}\delta_{ik} = (c_i - c_j)(c_i - c_k)$ are square; as Swinnerton-Dyer notes [50, Theorem 1ff], this is equivalent to there being no rational 4-torsion point.[1]

We let $\Omega$ be the set of bad primes of the above model of $E$ (this is the set of prime divisors of $\delta_{12}\delta_{13}\delta_{23}$ and thus always contains 2) and $\tilde{\Omega}$ when appending the infinite place. We will consider the quadratic twists $E_d : y^2 = (x - dc_1)(x - dc_2)(x - dc_3)$ of $E$ by squarefree $d$ that are coprime to $\Omega$. As we are interested in twist families, we can assume there is no prime that has the same nonzero valuation at all the $\delta_{ij}$; if there is, we can twist $E$ by this prime $p$ and the resulting curve is then good at $p$. On the other hand, we do allow the $\delta_{ij}$ to have a common prime divisor in the alternative case where the twisted curve would not be good at $p$.

Note that, at least with our assumptions as above, the coprimality of $d$ with $\Omega$ is not really restrictive; one can consider the other twists via replacing $E$ by a twist $\tilde{E}$

[1] Kane [23] seems to conflate a rational 4-torsion point with a (cyclic) rational 4-isogeny at various places, and moreover simply elides the condition in the statement of his Theorem 3.

Note that a curve with full 2-torsion can always be brought into the form $y^2 = x(x+1)(x+\lambda)$ and the condition of having a 4-torsion point is that $\lambda = \mu^2$ is square, whence $(\mu, \mu^2 + \mu)$ is indeed a 4-torsion point. On the other hand, in the context of such a curve with full 2-torsion, the primitive part of the 4-division polynomial factors as $(x^2 - \lambda)(x^2 - 2x + \lambda)(x^2 - 2\lambda x + \lambda)$, which can have linear factors without $\lambda$ being square, whereupon there is a 4-isogeny but no 4-torsion.

In any event, the condition could be a red herring for rank distribution, as if $E$ has full 2-torsion there is always (see §13.1) an isogenous curve with full 2-torsion and no 4-torsion – as the rank (though not the 2-Selmer rank) is an isogeny invariant, this then allows one to avert the issue.

of it by products of primes in $\Omega$ and applying the results to $\tilde{E}$ instead.[2] In this case the $\delta_{ij}$ can have common prime factors $p$; if so, upon twisting, any such $p$ will still divide at least one of them.

1.1. Our subject of interest shall be the 2-Selmer group of $E_d$, or more precisely how its size is distributed for $|d| \le X$. What exactly is this 2-Selmer group? We give a fuller discussion below in §7 but for now it suffices to give a brief description (following [50, §3]) in terms of everywhere locally soluble 2-coverings. These are intersections of three quadric equations $m_i y_i^2 = x - dc_i$ where $m_1 m_2 m_3$ is a nonzero square, and we denote such a curve by $\mathcal{C}(\vec{m})$. Any rational point $(X, Y)$ on $E_d$ gives a quadruple $(X, Y_1, Y_2, Y_3)$ on[3] some $\mathcal{C}(\vec{m})$, and finding such a quadruple would conversely yield a point on $E_d$. We need only consider the $m_i$ up to squares, and upon treating the $m_i$ as elements of $\mathbf{Q}^\star/(\mathbf{Q}^\star)^2$, the triples $\vec{m}$ form an abelian group under component-wise multiplication: $\vec{m} \times \vec{m}' = (m_1 m_1', m_2 m_2', m_3 m_3')$. The 2-Selmer group is then the subset of $\vec{m}$ such that $\mathcal{C}(\vec{m})$ is everywhere locally soluble.

In particular, we can restrict attention to $\vec{m}$ where the $m_i$ are units at all primes outside $\tilde{\Omega}$ that do not divide $d$, as if $m_i$ has odd valuation at such a prime $l$, then the equation $m_i y_i^2 = x - dc_i$ is already insoluble in $\mathbf{Q}_l$.

The 2-Selmer group is then computable in terms of Legendre symbols involving the primes dividing $d$ and the various places in $\tilde{\Omega}$ (including a specification of the sign of $d$ corresponding to the infinite place), and it contains the group of $\vec{m}$ such that $\mathcal{C}(\vec{m})$ has a global solution. This latter group is isomorphic to $E_d(\mathbf{Q})/2E_d(\mathbf{Q})$ in terms of the Mordell-Weil group $E_d(\mathbf{Q})$. Thus an upper bound on the rank of the 2-Selmer group gives a bound on the Mordell-Weil rank. Somewhat trivially, due to the 2-torsion points on $E_d$, the 2-Selmer rank is at least 2.

1.1.1. We shall ultimately show the following result, in terms of the numbers

$$\rho_s = \frac{2^s}{\prod_{v=1}^{s}(2^v - 1)} \prod_{n=0}^{\infty} \left(1 - 1/2^{2n+1}\right) = \frac{1/2^{s(s-1)/2}}{\prod_{v=1}^{s}(1 - 1/2^v)} \prod_{n=1}^{\infty} \frac{1}{1 + 1/2^n}, \tag{1}$$

for which $\rho_0 + \rho_2 + \rho_4 + \cdots = \rho_1 + \rho_3 + \rho_5 + \cdots = 1$ (the $n$-product is $\approx 0.419422$). We assume $E$ is twist-minimal, meaning there are no bad primes $p$ that can be removed by quadratic twisting; as above, this says the $v_p(\delta_{ij})$ are not all the same.

**Theorem 1.1.2.** *Let $E/\mathbf{Q}$ be a twist-minimal elliptic curve with full rational 2-torsion and no rational 4-torsion point, and consider quadratic twists $E_d$ of $E$ by[4] odd squarefree integers $|d| \le X$ coprime to the product of the bad primes of $E$. Then for any $\omega < 1/2$, the proportion of such $d$ such that $E_d$ has 2-Selmer rank of $(s+2)$ is $\rho_s/2 + O_{E,\omega}\big(1/(\log\log X)^\omega\big)$, with an effective constant in the error.*

[2] At least in his preprint version, Smith [45] doesn't seem to regard this point too transparently, though it is a minor quibble in any event once one can consider $\tilde{E}$ to begin with.

On the other hand, Heath-Brown [18] only investigates the twists by odd (squarefree) integers for $y^2 = x^3 - x$, and as he doesn't additionally consider the same for $y^3 = x^3 - 4x$, his results (pedantically) do not cover all twists for the case of congruent numbers. Meanwhile, Kane [23] notes (*inter alia*) this idea of replacing $E$ by $\tilde{E}$ when deriving Corollary 4 from his Theorem 3.

[3] Namely, given $(X, Y)$ we can take $m_i = X - dc_i$ and $Y_i = 1$ for all $i$.

[4] Kane states his Theorem 3 in terms of *positive* squarefree integers, but I do not think it is correct (for parity reasons). For instance, with the elliptic curve of conductor 225 given by twisting 15a by $-15$, thus $(c_1, c_2, c_3) = \big(-15 \cdot 16, 0, 15 \cdot 9\big)$ in our model, all of its twists by positive squarefree integers coprime to 15 have even parity. There are also decent reasons to twist by fundamental discriminants instead of squarefree integers, but we shall opt for the latter.

*In other words, writing $s(E_d)$ for the 2-Selmer rank of $E_d$, we have*

$$\frac{\#\{|d|\le X : \mu(d)\neq 0,\ \gcd(d,\Omega)=1 \mid s(E_d)=s+2\}}{\#\{|d|\le X : \mu(d)\neq 0,\ \gcd(d,\Omega)=1\}} = \frac{\rho_s}{2} + O_{E,\omega}\Big(\frac{1}{(\log\log X)^{\omega}}\Big).$$

Let us immediately say that our methods are largely based on those of Smith [45] and Swinnerton-Dyer [50], with some various technical aspects added to our exposition of their work, mostly to explicitly obtain the stated error bound.[5]

1.1.3. We describe our above observation more fully, namely that the $\gcd(d,\Omega)=1$ condition can be softened by working with twists of $E$ involving bad primes.

As an example, we might consider $E$ as 15a with $(c_1,c_2,c_3)=(-16,0,9)$, where $\Omega=\{2,3,5\}$. Then, for instance, the twists of $E$ by $3l$ (where $3\nmid l$, since $3l$ is squarefree) are the same as the twists of $E_3$ by $l$, where $E_3$ is the twist of $E$ by 3. This $E_3$ is twist-minimal and meets the torsion conditions, and we can thus apply Theorem 1.1.2 to it. This then gives

$$\begin{aligned}&\frac{\#\{|d|\le X : \mu(d)\neq 0,\ 3|d,\ \gcd(d/3,\Omega)=1 \mid s(E_d)=s+2\}}{\#\{|d|\le X : \mu(d)\neq 0,\ 3|d,\ \gcd(d/3,\Omega)=1\}}\\ &= \frac{\#\{|d|\le X/3 : \mu(d)\neq 0,\ \gcd(d,\Omega)=1 \mid s((E_3)_d)=s+2\}}{\#\{|d|\le X/3 : \mu(d)\neq 0,\ \gcd(d,\Omega)=1\}} = \frac{\rho_s}{2} + O_{E_3,\omega}(\cdots),\end{aligned}$$

which in particular has the same ratio of twists with a given 2-Selmer rank. The same analysis also applies when twisting by any other product of primes from $\Omega$.

1.1.4. We can also mention what happens, or is expected to happen, when we loosen the requirements on $E$.

First we consider the effect of modifying the 2-torsion condition. This is perhaps best considered in terms of the Galois nature of the cubic $f$ in $E: y^2=f(x)$. When $f$ has exactly one rational root (so there is one 2-torsion point) the distribution is quite different, as noted by Xiong [55]. The results in this genre are more commonly described in terms of the $\phi$-Selmer group where $\phi: E_d\to E'_d$ is the isogeny corresponding to the 2-torsion point; up to a correction for the 2-torsion point, this $\phi$-Selmer group injects into the 2-Selmer group, as indeed we have the exact sequence

$$0\to E'_d[2]/\phi(E_d[2])\to \mathrm{Sel}_\phi(E_d)\to \mathrm{Sel}_2(E_d)\xrightarrow{\phi}\mathrm{Sel}_{\hat\phi}(E'_d)$$

so that $s_\phi(E_d)-1\le s(E_d)\le s_\phi(E_d)+s_{\hat\phi}(E_d)$ in terms of the $\phi$- and $\hat\phi$-Selmer ranks. In particular, Xiong's main result shows that typically[6] the $\phi$-Selmer distribution has mean $\sqrt{(\log\log d)/2}$ as $d\to\infty$, with the 2-Selmer mean thus also being

[5]Although Smith references Kane's work, his setup does not depend on it. He is somewhat uncareful (see [45, Corollary 1.2]) with the issue we highlighted in Footnote 4 regarding the necessity of including both positive and negative twists in the average, but this is merely an accounting item in any event (note that for the congruent number curve, which is indeed Smith's main application, there is no distinction between positive and negative twists).

[6]One atypical case is for $A(a,b): y^2=x(x^2+ax+b)$ with $b(a^2-4b)$ a square; here we see that [55, Remark (2) to Theorem 1] notes in passing that both the $\phi$- and $\hat\phi$-Selmer means are bounded (as considered by Yu [57]), so the 2-Selmer mean is also bounded. The other atypical case is when $b$ is square (so that the 2-isogenous curve $A(-2a,a^2-4b)$ has full 2-torsion), where Xiong and Zaharescu [56] show that the $\phi$-Selmer mean is even larger, namely $(\log\log d)/2$, at least when $A(-2a,a^2-4b)$ has no rational 4-torsion point. (It seems Xiong's Remark glosses over the fact that Theorem 1 of [56] excludes the case where $A(-2a,a^2-4b)$ has a 4-torsion point (phrased therein as "$ab$ not a square" in the model $y^2=x(x+a)(x+b)$), so we mention it here).

unbounded. So this is dramatically different than the full 2-torsion case, where the 2-Selmer mean is bounded.

When $f$ is irreducible with Galois group $\mathrm{Sym}_3$, the work of Klagsbrun, Mazur, and Rubin [25] (specialized to $\mathbf{Q}$) gives the distribution of the 2-Selmer rank (albeit under Swinnerton-Dyer's ordering, see §1.2.1 below), with again the mean being finite.[7] I am unaware of any work for the case where $f$ has Galois group $\mathrm{Alt}_3$.

Secondly, when $E$ has full 2-torsion, the 4-torsion condition can likely be mollified somewhat (it ultimately comes from Swinnerton-Dyer's analysis, appearing for us in Lemma 8.5.4), perhaps handling some additional cases that have different transition probabilities (see §1.2.1) – though again one presumably wants to avoid cases where the 2-Selmer mean is in fact unbounded.[8]

Finally, when $E$ is not twist-minimal but meets the stated torsion conditions, this corresponds to a re-ordering of the $d$ with respect to the associated twist-minimal curve $C$. For instance, if $E$ is $C$ twisted by 5, the twists of $E$ up to $X$ are given by the twists of $C$ up to $X/5$ that are coprime to 5, together with the twists of $C$ up to $5X$ that are multiples of 5. Although we don't give the details, it is possible to include congruential information (say mod $m$) about $d$ in the above Theorem 1.1.2, with the main term being the same, though the error term then also depends on the modulus $m$. Upon re-combining the progressions modulo $m$, we then get a result for the twists of $E$ with an error term depending on $C$ and $m$, or in other words on $E$. Thus Theorem 1.1.2 is also true as stated in the non-twist-minimal case.

1.2. **History.** As noted in the Abstract, the first to consider the 2-Selmer distribution was Heath-Brown [17, 18], who looked at odd squarefree twists of the congruent number curve $E : y^2 = x^3 - x$. In [17] he computed the average size of the 2-Selmer group, then in [18] all of the integral moments of this, which then gave the distribution.[9]

His method consisted of noting that the 2-Selmer rank can be written in terms of (a matrix of) Legendre symbols $(p_i|p_j)$ for prime divisors of $d = p_1 \cdots p_r$, and then showed suitable equi-distribution of the values of such symbols. Dividing the primes into dyadic-like intervals, when both of the primes are large a bilinear bound (such as [17, Lemma 4], which he notes perhaps originated with Heilbronn [20]) saves a suitable amount, while when one of the primes is small and the other is large one can apply (albeit ineffectively in some ranges) results about primes in arithmetic progressions. One significant difficulty is then in handling the cases where both primes are small – I must say that I do not completely understand Heath-Brown's mechanism here, though it seems to do with a careful consideration (in [18, §§6-7]) of "unlinked indices" and the main terms therein (see also [9, §7.3-7.6] or [8, (48ff)]).

In his calculations for the $k$th moment, Heath-Brown ends up with an error term of $X(\log\log X)^{4^k}/(\log X)^{1/4^k}$. The transition range for when this becomes useful is when $16^k \sim \log\log X/(\log\log\log X)$, that is when $k \sim (\log\log\log X)/\log 16$. The reduction to his Theorem 2 in §8 (passing from moments to a distribution) is

[7] In various talks, Smith has stated that his results extend to this case.

[8] Similar to the last part of Footnote 6, the final paragraph of [55, Remark (2) to Theorem 1] seems to imply a mis-statement in this regard. Note first that if $E$ has full 2-torsion and a rational 4-torsion point, then there is a 2-isogeny $\phi : E \to E'$ for some $E'$ that also has full 2-torsion. In this case, with $E : y^2 = x(x^2 + ax + b)$, both $a^2 - 4b$ and $b$ are squares, so the last sentence of Xiong's Remark should apply. However, this again seems suspect when $E$ has a 4-torsion point.

[9] As Fouvry and Klüners [8, p. 459] note, the latter need not be automatic from the former.

inexplicit as given, but presumably can be done with explicit error terms. At any rate, I do not expect that one would save more than a (small) power of $(\log\log X)$ in the end.[10] Also note that his work utilizes the ineffective theorem of Siegel and Walfisz in [17, Lemma 6] and [18, Lemma 5].

The results for the average size were generalized by Yu [58] to elliptic curves with full 2-torsion[11] (though restricting twists to certain arithmetic progressions).

1.2.1. The next investigation for the distribution of the 2-Selmer rank was undertaken by Swinnerton-Dyer [50], who considered curves with full 2-torsion but no rational 4-torsion point. Rather than deal with the question of the equi-distribution of $(p_i|p_j)$, he took a different tact, essentially assuming they were equi-distributed by considering ordering $d$ by the number of prime factors (rather than the ordinary ordering of $|d| \leq X$). Fixing the number $r$ of prime factors, he used the description of the 2-Selmer rank in terms of everywhere locally soluble 2-covers. Upon restricting consideration to the first $j$ primes dividing $d$, this yields a sequence[12] of estimations $s_j(E_d)$ for which $s_r(E_d)$ is equal to the 2-Selmer rank $s(E_d)$. Moreover, and critical to his result, Swinnerton-Dyer was able to show that these $s_{j+1}(E_d)$ have a probability distribution derived from that for $s_j(E_d)$ in a sufficiently generic case (and averaging over suitable $d$ for this to make sense). In particular, the probability that $s_{j+1} = s_j + 2$ is $1/2^{2s_j+1}$, the probability that $s_{j+1} = s_j$ is $3/2^{s_j} - 5/2^{2s_j+1}$, and the remainder has $s_{j+1} = s_j - 2$. After showing non-generic cases are rare when there is no 4-torsion, he concludes by applying a Markov chain analysis to this probability distribution, with the stationary state given by the $\rho_s$ listed in (1).

1.2.2. Kane [23] then extended Heath-Brown's results about moments to the general case of full 2-torsion, again excluding the case where $E$ has a rational 4-torsion point. Kane's methods use Swinnerton-Dyer's description of the 2-Selmer group, though after that largely follow Heath-Brown, showing suitable equi-distribution of $(p_i|p_j)$ for "active" indices. Also, the final step, of passing from moments to a distribution, is done by complex analysis rather than linear algebra, using Swinnerton-Dyer's known limiting distribution to bootstrap it (the given argument has an inexplicit style, relying on compactness and convergent subsequences, etc.).

1.2.3. More recently, as part of a larger work, a preprint of Smith [45] gives a different method for showing the distribution of 2-Selmer ranks. Again one needs to consider equi-distribution of $(p_i|p_j)$, but he introduces a new[13] idea with permuting indices: the 2-Selmer rank is the kernel dimension of a matrix over $\mathbf{F}_2$, and is thus invariant under row/column permutations. This allows him to ignore $(i, j)$ where neither the bilinear estimate nor primes in arithmetic progressions gives a viable bound. He then completes the proof via Swinnerton-Dyer's Markov chain analysis.

[10] At the very end of [23], Kane suggests the power might be 1/8.

[11] In fact, Heath-Brown had prognosticated in [17, Remark 1, page 173] that this should be possible, though was unsure whether the constant (that is, the average size of the 2-Selmer group) should remain the same, as indeed Yu showed. On the other hand, extending the results to curves without full 2-torsion (cf. [18, Page 335]) has proven to be more difficult.

[12] It is a bit fanciful to say that the sequence $\{s_j(E_d)\}_j$ "converges" to $s(E_d)$ as $j \to r$ for any given $d$, as only the final value for $j = r$ is relevant, though in the sense of 2-Selmer *distributions* (that is, when considering many $d$) the usage of the term is more reasonable.

[13] Kane's work includes a permuting of the indices, but it seems only to be used in the rather limited context in that he prefers to allow prime variables to run freely rather than be subject to $p_1 < \cdots < p_r$ (say), and he thus overcounts by the natural factor of $r!$.

Although his preprint version gives the error bound as relying on Siegel's ineffective theorem, as discussed in Footnote 25 below I think this is mostly an oversight/inefficiency. His error term saves a small power of $(\log\log X)$, though is inexplicit about what power this is (we shall obtain any power less than $1/2$).

Smith goes on to discuss much more than the 2-Selmer rank, indeed considering the $2^\infty$-Selmer rank (and also the associated problem of the $2^\infty$-rank of quadratic class groups), but we will not consider such matters here.

1.2.4. Finally, there are the related works of Fourvy and Klüners [8, 9], which consider an analogous problem for the 4-rank of quadratic narrow class groups. Again the principal difficulty[14] is in demonstrating adequate equi-distribution for $(p_i|p_j)$. In [8] they show the expected moments of 4-ranks largely by imitating Heath-Brown's method of linked indices, and thereby derive the distribution of 4-ranks in [7]. As the ordinary class group is equal to the narrow class group except for a thin subset of discriminants (positive discriminants with no prime divisor that is 3 mod 4), the results then carry over to it also (cf. [8, Corollary to Lemma 10]).

In [9] they then consider the restriction of the problem to this thin subset of "Gaussian" discriminants, showing an expected distribution of the 4-rank of the narrow class group for them. We will discuss this in §§10-12 below. However, we will not consider their more notable result, namely that they are also able to get the distribution when including the 4-rank of the ordinary class group (which is either the same or 1 less).

The error terms are comparable to Heath-Brown's result (with $2^k$ rather than $4^k$), reduced by the expected relative factor $1/\sqrt{\log X}$ in the latter case. Again the results are ineffective.

A recent preprint of Chan, Koymans, Milovic, and Pagano [3] then extends Smith's results about higher Selmer groups to the 8-rank of quadratic class groups for Gaussian discriminants. More relevant for our discussion, it also shows how to derive the previous result about narrow 4-ranks of such discriminants via Smith's methods, again with an ineffective error term that saves a power of $(\log\log X)$. We put in a (little) bit of extra effort in our setup to be able to include this case in §12, and in a sequel [54] to the current paper we give an exposition of [3].

1.3. Our presentation is largely a reworking of Smith's methods, and we claim almost no novelty. We recall the basic background on primes and divisors in §3, and then in §4 give a convenient way of splitting up squarefree integers into boxes (Cartesian products of dyadic-like intervals). Here we implicitly use that the $j$th prime divisor $p_j$ of a typical integer is expected to have $\log\log p_j \sim j$.

Section 5 is probably our most significant contribution, as we replace Smith's inductive scheme [45, Proposition 6.3] by a more direct consideration of the product $\prod_{(i,j)}\big[1+(p_i|p_j)\big]$ over suitable index pairs $(i,j)$. Upon multiplying out the product, we then use bilinear bounds or primes in arithmetic progressions to bound the resulting sums, and it is here where we are able to avoid using Siegel's ineffective theorem. The main upshot is that we can split boxes by Legendre symbol conditions, except those corresponding to $(i,j)$ that are both small, with the splitting reducing the number of integers in the box roughly by the expected power-of-2.

[14]The earlier work of Gerth [12] had mostly avoided this by considering the problem under the ordering by number of prime factors, as Swinnerton-Dyer later adopted in the 2-Selmer case.

Section 6 then gives Smith's argument on how to average over permutations of indices, essentially showing that our exclusion of small $(i,j)$-pairs does not introduce a very large error, at least for the 2-Selmer rank distribution (which is invariant under said permutations). However, the error here is still ultimately the largest in our analysis; also, as the box-splitting aspects are not permutation-invariant, we obtain no information about whether each small $(i,j)$-pair reduces the number of integers by roughly $1/2$ (see Smith's comment after his statement of Proposition 6.3).

We then shift gears and in §7 give an outline of Swinnerton-Dyer's description of the 2-Selmer group, and proceed to sketch his argument regarding 2-Selmer estimations, genericity, and the ensuing Markov chain. Combined with the previous analysis about box-splitting, this then gives the main Theorem 1.1.2. However, we intend §7 to largely be an overview, and indeed in §8 we give many more details, largely replicating Swinnerton-Dyer's arguments in a somewhat more robust form (for instance, he speaks of "random primes" which thus need a suitable interpretation, and as Kane notes, one can instead use formal symbols to make this more rigourous). Additionally, as the use of Markov chains seems somewhat of a black box, we give a brief sketch of how this can be done in our specific case (which is somewhat easier in that the transition matrix is essentially tridiagonal). It is admitted that this section is rather long and technical.

We then recapitulate the main argument in §9, and in §§10-12 discuss the analogous problem for quadratic class groups.

Finally, in §13 we do some exercises to show a few "well-known" facts that were mentioned along the way.

1.3.1. As a significant aspect of Smith's work on the 2-Selmer distribution involves transferring Swinnerton-Dyer's results from a less natural ordering to the expected ordering, it seems useful to review this. In fact, the idea of ordering integers by the number of prime factors seems to originate with Gerth [12, (1.1)] in the context of 4-ranks of quadratic class groups. Defining $S_r(X)$ to be the set of squarefree integers up to $X$ that have exactly $r$ prime factors, the basic idea behind the "unnatural" ordering is to consider an arithmetic function $F$ and the limit

$$\lim_{r\to\infty}\lim_{X\to\infty}\frac{\sum_{m\in S_r(X)}F(m)}{\sum_{m\in S_r(X)}1}.$$

On the other hand, the integers of size $X$ have a normal distribution in their number of prime factors, with mean and variance $\log\log X$. Thus the integers with $r\sim\log\log X$ should dominate under the standard ordering.

We first recall Kane's method for passing from Swinnerton-Dyer's ordering to the natural one. His main tool in this regard is [23, Proposition 10]. In our notation, given $r$ with $(\log\log X)/2<r<2\log\log X$, Kane shows that

$$\frac{1}{r!}\sum_{\vec{p}\in S_r^D(X)}F(p_1,\dots,p_r)=\Big(\frac{1}{\#G}\sum_{g\in G}F(g)\Big)\Big(\frac{\#S_r^D(X)}{r!}\Big)+O_D\Big(\frac{X\log\log\log X}{\log\log X}\Big)$$

where $D$ is 4 times the product of the primes in $\Omega$, while $S_r^D(X)$ is the set of squarefree integers up to $X$ coprime to $D$ with each written in $r!$ ways as a product $p_1\cdots p_r$, and $F$ is a function (with suitable boundedness) from $G=\big(U_D/U_D^2\big)^r$ to $\mathbf{C}$, where $U_D=(\mathbf{Z}/D\mathbf{Z})^\star$. Note that this implicitly uses information about the distribution of quadratic residue symbols, relying thereupon on results for primes in arithmetic progressions and thus zero-free regions for Dirichlet $L$-functions. Also,

the relative error is not too striking, as one might expect from results involving divisor distributions. Kane then sums this over $|r - \log\log X| \le (\log\log X)^{3/4}$ to pass to the natural ordering of integers (with sufficiently small error from other $r$).

In Smith's version, he first introduces boxes, which are themselves mostly just a technique to allow the replacement of a constraint like $p_1 \cdots p_r \le X$ by bounds on each individual $p_i$. This gives a relative error that is a (large) power of $1/\log\log X$, but otherwise is harmless. One is then led to consider the distribution of $(p_i|p_j)$ as $p_i, p_j$ range over their respective intervals. When $p_i$ is much smaller than $p_j$, one can simply consider $p_i$ fixed and use results on primes in arithmetic progressions modulo $p_i$. More subtle is the case where $p_i$ and $p_j$ are both large (with respect to $X$), where a mean-value result dating back to Heilbronn shows that a bilinear sum $\sum_i \sum_j \alpha_i \beta_j (p_i|p_j)$ has suitable cancellation.

This leaves the $(i, j)$ for which $p_i$ and $p_j$ are both small, and it is here that Smith (crucially) employs an additional insight. Namely, there is (in Kane's result) a natural action of $\mathrm{Sym}_r$ on $G$, and Smith observes that our desired $F$ (involving the 2-Selmer rank) is invariant under it. He then gives a suitable averaging method over such permutations, so that the contributions from the $(i, j)$ with both small are adequately "mixed" in with the other index pairs. In contrast, both Heath-Brown and Kane had to deal more directly with such $(i, j)$-pairs (albeit in the context of moments). Ultimately, Smith's method provides (in our version) a somewhat stronger error term, and moreover gives an effective constant therein.

## 2. Notation and parameters

2.1. We accumulate various notations and parameters in one place, for the convenience of the aspiring reader. We will be considering the elliptic curve given by $E : y^2 = (x - c_1)(x - c_2)(x - c_3)$, where the integers $c_i$ have no common nontrivial square factor, and $c_1 < c_2 < c_3$ with $c_2 = 0$. We write $\delta_{ij} = (c_i - c_j)$ for the root-differences and $\Omega$ for the set of bad primes (these are divisors of the $\delta_{ij}$), and $\tilde{\Omega}$ when appending the infinite place. We consider twisting $E$ by squarefree integers $|d| \le X$ coprime to $\Omega$. Here $X$ is the main parameter in the paper.

While we eventually consider both positive and negative $d$, in the discussion of prime divisors it is more convenient to only consider positive integers, and therein we write $\tilde{d}$. With this in mind, we can re-interpret the above schema in a slightly different way (§3.1), considering a set of primes $\mathcal{P}$ from which all the divisors of $\tilde{d}$ must come. We will require that $\mathcal{P}$ be the set of all primes in specified coprime residue classes $\mathcal{R}_{\mathcal{P}}$ to some fixed modulus $M_{\mathcal{P}}$. We will write $\xi_{\mathcal{P}}$ for the number of such residue classes, and also $\alpha_{\mathcal{P}} = \xi_{\mathcal{P}}/\phi(M_{\mathcal{P}})$ for their relative density. (The 2-Selmer case has $\alpha_{\mathcal{P}} = 1$).

We then write $S^{\mathcal{P}}(U)$ for the set of positive squarefree integers up to $U$ all of whose prime factors come from $\mathcal{P}$. The size of this is denoted as $\Phi^{\mathcal{P}}(U)$. Moreover, we write $S_t^{\mathcal{P}}(U)$ and $\Phi_t^{\mathcal{P}}(U)$ for the restriction of these to the squarefree integers with exactly $t$ prime factors.

2.1.1. We write $\bar{T}$ for a box (§4.2), which is a Cartesian product of singleton sets of primes and basic (real) intervals, with $T$ the Cartesian product of singleton sets and sets of all primes from the basic intervals that are in $\mathcal{P}$, and $\hat{T}$ as the set of squarefree integers thus represented. This box will have various associated quantities. The size limit on the singleton sets depends on $\eta_0$, which we discuss more in §2.2. The

number of singleton primes in the Cartesian product is denoted $k_0$, and $\tilde{r}$ is the number of such singletons plus the number of basic intervals (thus $\tilde{r}$ will also be the number of prime factors of $\tilde{d}$). We also introduce a parameter $\eta_1$ (see §4.3.3), and write $k_1$ for the number of singletons plus basic intervals that are less than $Q_1$, where here $Q_1$ is $\exp\exp\big((\log\log X)^{\eta_1}\big)$.

We then have (§4.4.1) a third parameter $\eta_{\rm s}$ associated to $P_{\rm s} = \exp\big((\log\log X)^{\eta_{\rm s}}\big)$; this occurs in the analysis of exceptional zeros, with also a putative (though likely empty) sequence $\{\mathcal{M}_i\}$ of exceptional conductors (§3.4).

A culmination of all this jargon is then the definition of a pleasant box in §4.5.

2.1.2. The process of cutting up boxes in §5 then introduces various decorations of $T$. In particular, we write its Cartesian product as $\prod_l T_l$. We then have two key sets $\mathcal{K}$ and $\mathcal{L}$. The first consists of residue class restrictions (to the modulus $M_{\mathcal{P}}$) for each of the $\tilde{r}$ primes dividing $\tilde{d}$; the second has Legendre symbol specifications for each pair $(i,j)$, associated to primes $p_i$ and $p_j$ dividing $\tilde{d}$ (so giving $\binom{\tilde{r}}{2}$ conditions).

We then write $T(\mathcal{K})$ for the elements of $T$ that meet the residue class specifications of $\mathcal{K}$, and this is still a Cartesian product as $T(\mathcal{K}) = \prod_l T_l(\mathcal{K})$. We call this the $\mathcal{K}$-trimming of the box. We write $T(\mathcal{K},\mathcal{L})$ for the elements of $T(\mathcal{K})$ which meet the Legendre symbol specifications in $\mathcal{L}$. This need not be a Cartesian product.

Then we have (§5.2.1) the $(\mathcal{K},\mathcal{L},[k_0,k_1])$-trimming of a box, denoted $\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$, restricting $T$ by $\mathcal{K}$ for the first $k_1$ primes and $\mathcal{L}$ for $(i,j)$ with $1 \le i < j \le k_1$ and $i \le k_0$. This will have a Cartesian product $\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1} = \prod_l \tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$.

We write $(m|n)$ for the Kronecker (and thus also Legendre or Jacobi) symbol.

In §6 we then write $\mathcal{D}(\tilde{r},\mathcal{P})$ for the set of all possible choices of $(\mathcal{K},\mathcal{L})$ for given $\tilde{r}$ and $\mathcal{P}$, and $\mathrm{Sym}_u$ for the symmetric group on $u$ symbols. This symmetric group acts (with $u=\tilde{r}$) on $\mathcal{K}$ and $\mathcal{L}$ by permuting the indices, and therein we write $\mathcal{K}^\sigma$ and $\mathcal{L}^\sigma$ for an element $\sigma \in \mathrm{Sym}_{\tilde{r}}$. Given a $\tilde{d}$ represented by a box $T$, the set $W_{\tilde{d}}(\mathcal{K},\mathcal{L})$ contains the permuations $\sigma \in \mathrm{Sym}_{\tilde{r}}$ for which $\tilde{d}$ is in $\tilde{T}(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1}$.

While the above template only considers positive $\tilde{d}$, we also want to allow negative $d$ with the 2-Selmer group. It turns out we can do this with a minimal obfuscation of notation, mainly needing just $\hat{T}^{\pm}$, which is twice as large as $\hat{T}$, including both $\tilde{d}$ and $-\tilde{d}$ for the $\tilde{d} \in \hat{T}$.

2.1.3. The discussion of the 2-Selmer group in §7 then introduces another milieu of notation. Firstly there is $\tilde{\mathcal{K}}$, which specifies merely the Legendre symbol for primes in $\tilde{\Omega}$ (rather than a residue class). Moreover, as above, we want to consider $d$ of both signs, and will attach the sign $\varepsilon$ of $d$ as a subscript on $\tilde{\mathcal{K}}$.

We can also mention the local Hilbert symbol at $l$, which we write as $(x,y)_l$.

Then there are various vector spaces such as $Y_l = \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$, and $V_l$ as the set of triples in $Y_l$ whose product is 1, with $V_{\mathcal{B}} = \oplus V_l$ then the sum being over all $l \in \mathcal{B}$, where $\mathcal{B}$ is the union of $\tilde{\Omega}$ with the primes dividing $d$. Finally there is $U_{\mathcal{B}} \subset V_{\mathcal{B}}$, which is generated by $(1,l,l)$ and $(l,l,1)$ for all $l \in \mathcal{B}$; and $W_{\mathcal{B}} \subset V_{\mathcal{B}}$, which is the sum of the $W_l$, which are local images of the Kummer map.

We then have the local pairing $e_l$ on $V_l \times V_l$ and its sum $e_{\mathcal{B}}$, which yields the pairing matrix $\mathbf{M}(E_d)$ on $U_{\mathcal{B}} \times W_{\mathcal{B}}$. This gives the 2-Selmer rank $s_{\tilde{r}}(E_d)$, and as the notation suggests, we have a sequence of estimations $s_c(E_d)$ corresponding to restricting the pairing matrix to the first $c$ primes dividing $d$ (with the $\tilde{r}$th estimation

then being the 2-Selmer rank). As the 2-Selmer rank is the same for all $d$ meeting some given $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ conditions, we can also consider $\mathbf{M}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ and $s_j(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$. The set of all possible $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $\mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$, and the restriction (which by convention includes the class of $d$ in $\prod_{l\in\tilde{\Omega}} \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$) of this to the first $c$ primes is $\mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$.

The expected 2-Selmer distribution $\rho_s$ is given in various places, *e. g.* (1) above.

Section 8 then introduces formal symbols $\dot{p}_j$ in place of primes, and $\mathbf{P}_u$ for the set of $u$ of them. We then have $U_{\mathcal{B}}^c$ and $W_{\mathcal{B}}^c$ restricted to using only the first $c$ primes (or now, formal symbols), with $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ (sometimes abbreviated $\tilde{U}_c$) and $W_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ the restriction of these to the pairing matrix kernels. We have $\vec{w}_p^1$ and $\vec{w}_p^2$ as basis vectors for $W_p$, and similarly for $\vec{\alpha}_p^1$ and $\vec{\alpha}_p^2$ with $U_p$. Finally, in the Markov chain analysis we have the transition matrix $M$ and the vectors $\vec{h}^c$ of 2-Selmer rank distributions.

2.2. Next we discuss the three $\eta$-parameters.

We fix a parameter $\eta_0$ with $0 < \eta_0 < 1$ and let $P_0 = \exp\big((\log\log X)^{\eta_0}\big)$, with $k_0$ the number of prime factors up to $P_0$ of a given squarefree integer $\tilde{d} \le X$, with the expectation that $k_0$ is roughly $\log\log P_0 = \eta_0 \log\log\log X$. We will allow $k_0$ to be as large as $\kappa_0 \log\log P_0$ for some parameter $\kappa_0 > 3$. The error from excluding $\tilde{d}$ with larger $k_0$ will (essentially) be of relative size $\ll 1/(\log P_0)^{\kappa_0(\log\kappa_0 - 1)}$ which is the same as $\ll (1/\log\log X)^{\eta_0[\kappa_0(\log\kappa_0-1)]}$.

We also fix a parameter $\eta_1$ with $0 < \eta_1 < 1$ and let $Q_1 = \exp\exp\big((\log\log X)^{\eta_1}\big)$, with $k_1$ as (essentially) the number of prime factors up to $Q_1$ of a given positive squarefree $\tilde{d} \le X$, with the expectation that $k_1$ is roughly $\log\log Q_1 = (\log\log X)^{\eta_1}$. We will require that $\eta_1 < \eta_0/2$ for our schema of cancellation to work out. Note that here $\log Q_1$ is $\exp\big((\log\log X)^{\eta_1}\big)$ and so is itself asymptotically smaller than any fixed power of $(\log X)$, but larger than any power of $(\log\log X)$.

We will allow $k_1$ to be as large as $\kappa_1 \log\log Q_1$ for a parameter $\kappa_1$, in fact we will simply take $\kappa_1 = 3$, and the error from ignoring $\tilde{d}$ with larger $k_1$ will be of relative size $\ll 1/(\log Q_1)^{\kappa_1(\log\kappa_1 - 1)}$, which saves more than any power of $(\log\log X)$ asymptotically, so that this error will be smaller than the previous one.

2.2.1. The main impediment against taking $\eta_0$ large is a combinatorial argument (Corollary 6.2.3) that will have a relative error of $2^{k_0/2} k_1/\sqrt{\log\log X}$ which we can see is $\le \kappa_1 (\log\log X)^{\kappa_0\eta_0(\log 2)/2 + \eta_1 - 1/2}$. To optimize the error it does not matter much how we take $\eta_1$ (subject to $0 < \eta_1 < \eta_0/2$), and are left to maximize

$$\min\big(\eta_0[\kappa_0(\log\kappa_0 - 1)], 1/2 - \kappa_0\eta_0(\log 2)/2 - \eta_1\big).$$

This can be seen to be arbitrarily close to $1/2$ by taking $\kappa_0 = 1/\eta_0\sqrt{\log 1/\eta_0}$ and letting $\eta_0 \to 0$, so the first term in the minimum tends to $\infty$, and the second to $1/2$.

2.2.2. We also have a parameter $\eta_{\mathrm{s}} > 0$ with $P_{\mathrm{s}} = \exp\big((\log\log X)^{\eta_{\mathrm{s}}}\big)$. This quantity will be the size of allowable moduli in our usage of the prime number theorem in arithmetic progressions for which we do not care if there is an exceptional zero or not. If we apply said theorem with primes of size $\ge Q_1$, we will still have a savings of $Q_1^{1/P_{\mathrm{s}}^\epsilon} = \exp\big((\log Q_1)/P_{\mathrm{s}}^\epsilon\big)$ where $\epsilon = 1/2$ if we want an effective result (and any $\epsilon > 0$ if we relied on Siegel's ineffective theorem). Thus we need $(\log\log Q_1)$ to be somewhat larger than $\epsilon \log P_{\mathrm{s}}$, or that $(\log\log X)^{\eta_1}$ exceeds $\epsilon(\log\log X)^{\eta_{\mathrm{s}}}$. As such, we will require $\eta_{\mathrm{s}} < \eta_1$, whereupon the choice of $\epsilon$ is almost irrelevant, and we

shall indeed take it as $1/2$ to obtain an effective result.[15] We will thus need to exclude conductors $\geq P_{\mathrm{s}}$ associated to exceptional zeros, and the density of such can be bounded by a result of Landau. The relative error therein is $\ll 1/(\log\log X)^{99+\eta_{\mathrm{s}}}$ and so taking $0 < \eta_{\mathrm{s}} < \eta_1 < \eta_0/2$ are the only meaningful constraints.

We will consider $\eta_0$, $\eta_1$, and $\eta_{\mathrm{s}}$ to be fixed throughout the argument, and thus will not include them as subscripts in the $O()$ or $\ll$ notation. We also allow all implicit constants to depend on $E$ (thus $\mathcal{P}$), and similarly elide this from the notation. The letter $c$ will sometimes be used to denote a constant (different at each appearance) that we do not give explicitly, for instance with $\exp\bigl(-c\sqrt{\log U}\bigr)$ as the error in the prime number theorem.

## 3. Background on the distribution of primes and divisors

First we collect results about the distribution of the number of prime divisors for squarefree integers, and also some results about the distribution of primes.

3.1. We will consider positive squarefree integers whose prime divisors come from a specified set $\mathcal{P}$. The instance for 2-Selmer groups will have $\mathcal{P}$ be the set of all primes except those in a finite set $\Omega$. Another typical example (we shall call it the Gaussian case, after the Gaussian integers) would be to restrict all the prime divisors to be 1 mod 4. Although one can derive some results under a lesser notion of regularity (such as merely requiring the set $\mathcal{P}$ to have a natural density), we shall only consider cases where $\mathcal{P}$ contains all the primes in specified coprime residue classes $\mathcal{R}_{\mathcal{P}}$ to a fixed modulus $M_{\mathcal{P}}$.

For instance, in the case of 2-Selmer groups we can take the collection of all coprime residue classes modulo the product of the primes in $\Omega$, and indeed, one can also reverse the above phrasing, noting that to any set of primes $\mathcal{P}$ determined by congruence conditions, there is a set of bad primes $\Omega_{\mathcal{P}}$ given by $p|M_{\mathcal{P}}$. It will sometimes be convenient to require that 8 divides $M_{\mathcal{P}}$, which can typically be done with no loss of generality.

The most pertinent constant associated to $\mathcal{P}$ is the number $\xi_{\mathcal{P}}$ of coprime residue classes that it contains, and we write $\alpha_{\mathcal{P}} = \xi_{\mathcal{P}}/\phi(M_{\mathcal{P}})$, and assume this is nonzero.

3.1.1. We let $\Phi^{\mathcal{P}}(U)$ be the number of squarefree integers up to $U$ that have all their prime divisors in $\mathcal{P}$. An asymptotic for this is rather trivial for the 2-Selmer case, as one has linear behavior, with a factor of $1/\zeta(2)$ for squarefreeness, and a factor of the product of $p/(p+1)$ over $p \in \Omega_{\mathcal{P}}$. Meanwhile, the case with $\mathcal{P}$ was already touched upon by Landau [30], and his methods (of comparing to $\zeta(s)^{\alpha_{\mathcal{P}}}$) readily show that there is some constant $\beta(\mathcal{P}) > 0$ such that

$$\Phi^{\mathcal{P}}(U) \sim \beta(\mathcal{P}) \frac{U}{(\log U)^{1-\alpha_{\mathcal{P}}}}.$$

3.2. We also require results on the number of integers with a specified number of prime factors, and we write $S_t^{\mathcal{P}}(U)$ for the set of squarefree integers up to $U$ with exactly $t$ prime factors, all of which are in $\mathcal{P}$. We also notate its size by $\Phi_t^{\mathcal{P}}(U)$.

[15] Our $P_{\mathrm{s}}$ is first utilized as $t$ and then termed $D_1$ by Smith, and its size is chosen in the Definition on page 71 before Corollary 6.11, wherein the latter in its second paragraph identifies the quantities. We analyze his situation more fully in Footnote 25 below, attempting to indicate why he appears to be reliant on Siegel's ineffective theorem.

In the case where $\mathcal{P}$ is the set of all primes, a classical result due to Erdős[16] and Kac [6] states that the number of prime divisors for squarefree integers up to $U$ is normally distributed with mean $(\log\log U)$ and standard deviation $\sqrt{\log\log U}$. Adapting this to the case with $\mathcal{P}$ yields a similar result with both appearances of $\log\log U$ multiplied by $\alpha_{\mathcal{P}}$.

This approximation is not too sharp in the tails, where one has a Poisson distribution. Indeed, a version of this with uniformity is due to Sathe [42] (see also Selberg [43]), and in the $\mathcal{P}$-situation we find there are constants $\beta_t(\mathcal{P})$ such that[17]

$$\Phi_t^{\mathcal{P}}(U) \sim \beta_t(\mathcal{P})\frac{U}{\log U}\frac{(\alpha_{\mathcal{P}}\log\log U)^{t-1}}{(t-1)!} \tag{2}$$

uniformly for $t < (2-\epsilon)\alpha_{\mathcal{P}}\log\log U$ for any $\epsilon > 0$. Here the $\beta_t(\mathcal{P})$ are uniformly bounded (as $t$ varies) by positive constants that only depend on $\mathcal{P}$.

3.3. A classical result of Mertens [35] gives an asymptotic for the sum of the reciprocal of the primes up to $U$, and this can be adapted to the $\mathcal{P}$-case, with an error term determined by the zero-free region for the $L$-functions with conductors dividing $M_{\mathcal{P}}$. For instance, we find that there are constants $m_{\mathcal{P}}$ and $c_{\mathcal{P}}$ such that

$$\sum_{\substack{p\le U\\ p\in\mathcal{P}}}\frac{1}{p} = \alpha_P\log\log U + m_{\mathcal{P}} + O_{M_{\mathcal{P}}}\big(\exp(-c_{\mathcal{P}}\sqrt{\log U})\big). \tag{3}$$

For upper bounds on the size of sets of integers with many small prime factors taken from $\mathcal{P}$, we have a particularly convenient formulation of Tudesq [51]. Writing $\omega_E(n)$ for the number of prime divisors of a squarefree integer $n$ that are in a set of primes $E$ (this should cause no confusion with the elliptic curve $E$), and $\tilde{E}^{\mathrm{r}}(U)$ for the sum of the reciprocals of the primes in $E$ up to $U$, we have the following.

**Lemma 3.3.1.** *(Tudesq* [51, Theorem 2]*). There exist absolute constants $T_1$ and $T_2$ such that*

$$\#\big\{n\le U : \omega_{E_j}(n) = e_j,\ 0\le j\le l\big\} \le T_1\cdot U\exp\Big(-\sum_{j=0}^{l}\tilde{E}_j^{\mathrm{r}}(U)\Big)\prod_{j=0}^{l}\frac{\big(\tilde{E}_j^{\mathrm{r}}(U)+T_2\big)^{e_j}}{e_j!}$$

*for all choices of $U\ge 1$, disjoint sets of primes $E_j$ indexed 0 to $l$, and $e_j\ge 0$ (which are thus allowed to depend on $U$).*

3.4. We also require basic results about the distribution of primes in arithmetic progressions. For this we recall the definition of an exceptional zero (sometimes called a Siegel zero) of a Dirichlet $L$-function. Given a parameter $\lambda > 0$, a (necessarily quadratic) Dirichlet character of conductor $M$ is said to be $\lambda$-*exceptional* if its $L$-function has a real zero $\beta$ with $\beta \ge 1-\lambda/(\log M)$.

A result[18] of Landau [31, (14), (8)] implies that the set of exceptional characters is sparse, and indeed by taking $\lambda$ arbitrarily small their conductors can be made to grow faster than any power-sequence. In particular, writing $\{\mathcal{M}_i(\lambda)\}$ for the

[16]The paper in question places a Germanic umlaut on the 'o' in his name.

[17]For fixed $t$ such a result essentially follows from the prime number theorem (see Landau's *Handbuch*, §56), and including uniformity (at least *après* Selberg) is mostly a technical matter. There is a transition at $t\sim 2\alpha_{\mathcal{P}}\log\log U$; when $\mathcal{P}$ is the set of all primes, see Hwang [21].

[18]The historical pedant will note that Landau only considered imaginary quadratic fields, and indeed the most direct methods for generalizing to the real quadratic case did not arise for a couple of decades.

putatively infinite sequence of exceptional conductors ordered increasingly, there is some $\lambda$ such that $\mathcal{M}_{i+1}(\lambda) \geq \mathcal{M}_i(\lambda)^2$ for all $i \geq 0$. We shall fix such a $\lambda$ for this paper, thusly also fixing the attendant sequence (possibly finite, or indeed likely empty) of exceptional conductors for this $\lambda$.

3.4.1. We then recall the prime number theorem for arithmetic progressions (written in terms of cancellation for character sums). Let $\psi$ be a nontrivial Dirichlet character of conductor $M$, and assume that $\psi$ is non-exceptional. There is an absolute constant $c > 0$ such that (see [53] for instance)

$$\sum_{p \leq U} \psi(p) \ll U \exp\Big(-c\frac{\log U}{\sqrt{\log U} + 9\log M}\Big)(\log UM)^4.$$

When $\psi$ is exceptional with zero $\beta$, the right side has an extra term $U^\beta$.

We also have results on how bad an exceptional zero $\beta$ can be. The ineffective theorem of Siegel [44] says that for any $\epsilon > 0$ there is some $c_\epsilon$ so that $\beta < 1 - c_\epsilon/M^\epsilon$ for all $M$. A flaw in this result is that given $\epsilon$, one has no means to compute $c_\epsilon$, even in theory. From Dirichlet's class number formula and rudimentary bounds on $L'_\chi(s)$ near $s = 1$ one can show that $\beta < 1 - c/\sqrt{M}(\log M)^2$ for some explicit $c > 0$. (The analysis for the derivative can also be handled more carefully when $L_\chi(1)$ is small, obtaining the result of Goldfeld and Schinzel [15, Corollary] that $\beta < 1 - c/\sqrt{M}$ for $c = 6/\pi + o(1)$; this could then be improved by nearly $(\log M)$ in the numerator by the bounds on $L_\chi(1)$ from the work of Goldfeld [14] and Gross and Zagier [16]).

3.4.2. We will also require a bound for a bilinear sum over primes joined by a Legendre symbol. Various results of this type exist in the literature, with Heath-Brown commenting that Heilbronn [20] seems to be the first to broach the subject. His method is rather simple – apply Cauchy's inequality twice and use the estimate of Pólya and Vinogradov for partial character sums – though as later authors noted, even the later is unneeded (character periodicity suffices).

**Lemma 3.4.3.** *Let $\{\alpha_p\}$ be complex numbers bounded by 1 and supported on primes $p$ with $P \leq p \leq 2P$ in a fixed residue class modulo 8, and similarly for $\{\beta_q\}$. Then*

$$\sum_{p \sim P} \sum_{q \sim Q} \alpha_p \beta_q (p|q) \ll \frac{PQ}{\min(P,Q)^{1/9}}.$$

We recall the proof of this in §13.3, with sundry comments about the literature.

## 4. Parametrising squarefree integers via boxes

We next describe our basic partitioning of the (positive) squarefree integers, dividing them up into boxes, and showing that almost all such integers are represented by boxes of certain types.

4.1. For $C$ a positive integer, the $C$-compressed basic intervals[19] starting at $Z$ (briefly, basic intervals of parameters $(C, Z)$) are defined by

$$(2^v Z + 2^v Z(u-1)/C, 2^v Z + 2^v Z u/C] \quad \text{for } v \geq 0 \text{ and } 1 \leq u \leq C.$$

[19] I think Smith's version of this (see $t'_i/t_i$ near the top of page 68 and $D_1$ in the Definition on page 71) has the compression factor be a small power of $(\log\log X)$ for small primes, and then grow to $(\log X)$ for larger ones. Our choice shall suffice for our purposes.

The ratio of the endpoints is thus $(1+u/C)/\big(1+(u-1)/C\big) \le 1+1/C$. We shall use this for $Z = P_0 = \exp\big((\log\log X)^{\eta_0}\big)$ and $C = \lfloor(\log\log X)^{99}\rfloor$. Note that this ensures that each basic interval is of some size, since $Z$ is on a different exponential scale than $C$ (so in particular $\exp(\sqrt{\log Z})$ asymptotically exceeds any power of $C$).

For any $C$, every prime (indeed, every real number) greater than $Z$ is in exactly one basic interval of parameters $(C, Z)$, so every squarefree integer is represented in a unique nondecreasing basic product of said parameters, namely $\prod_i\{p_i\} \times \prod_j \mathcal{I}_j$ for some increasing sequence $\{p_i\}$ of distinct primes with all $p_i \le Z$ and some sequence $\mathcal{I}_j$ of basic intervals $(A_j, B_j]$ of parameters $(C, Z)$ with $B_j \le B_{j+1}$.

4.2. We let $\mathcal{P}$ be a set of primes determined by congruence conditions as in §3.1.

For a given $\eta_0$ with $0 < \eta_0 < 1$, we define an $(X, \eta_0, \mathcal{P})$-box $\bar{T}$ to be a Cartesian product $\prod_u\{\mathbf{p}_u\} \times \prod_t \bar{T}_t$ where the primes $\mathbf{p}_u$ are distinct and in $\mathcal{P}$, with the $\mathbf{p}_u$ ordered increasingly with each $\le \exp\big((\log\log X)^{\eta_0}\big)$; while the $\bar{T}_t$ are a strictly increasing sequence of basic intervals $(A_t, B_t]$ of parameters $(\lfloor(\log\log X)^{99}\rfloor, P_0)$ with $\prod_u \mathbf{p}_u \prod_t B_t \le X$ and $P_0 = \exp\big((\log\log X)^{\eta_0}\big)$. As a technical convenience, we require that $P_0$ is larger than every prime in $\Omega_{\mathcal{P}}$ (equivalently, $X$ is large enough).

We let $k_0$ be the number of primes $\mathbf{p}_u \le P_0$ (corresponding to singleton sets), and $\tilde{r}$ be $k_0$ plus the number of basic intervals, and term the box to be of type $(\tilde{r}, k_0)$. We will typically index $1 \le u \le k_0$ and $k_0 < t \le \tilde{r}$, and allow $\bar{T}_u$ to refer to $\{\mathbf{p}_u\}$.

The squarefree integers with prime divisors from $\mathcal{P}$ that are represented by a box come from the set $T = \prod_u\{\mathbf{p}_u\} \times \prod_t T_t$ where $T_t$ consists of the primes in $\bar{T}_t$ that are in $\mathcal{P}$. Indeed, there is a natural injective map from this set to $S^{\mathcal{P}}_{\tilde{r}}(X)$, recalling the latter is the set of (positive) integers up to $X$ with exactly $\tilde{r}$ prime factors, all of which are in $\mathcal{P}$. We say that $\tilde{d}$ is represented by a box if it is in the image $\hat{T}$ of this map, and say that the box-data $\{\mathbf{p}_u\}$ and $(A_t, B_t]$ form the basic product of $\tilde{d}$.

Fixing $(X, \eta_0, \mathcal{P})$, every $\tilde{d} \in S^{\mathcal{P}}_{\tilde{r}}(X)$ is represented by at most one $(X, \eta_0, \mathcal{P})$-box, and moreover we can show that almost all such $\tilde{d}$ are represented by such a box. Writing $\Phi^{\mathcal{P}}_{\tilde{r}}(X) = \#S^{\mathcal{P}}_{\tilde{r}}(X)$ and $C = \lfloor(\log\log X)^{99}\rfloor$ we have the following.

**Lemma 4.2.1.** *For $\tilde{r}$ with $|\tilde{r} - \alpha_{\mathcal{P}}\log\log X| \le (\alpha_{\mathcal{P}}/99)(\log\log X)$, the union of all $(X, \eta_0, \mathcal{P})$-boxes with type $(\tilde{r}, k_0)$ (as $k_0$ varies) essentially covers $S^{\mathcal{P}}_{\tilde{r}}(X)$, with the exceptional set having size*

$$\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\frac{\log\log X}{C} \ll \frac{\Phi^{\mathcal{P}}_{\tilde{r}}(X)}{(\log\log X)^{98}}.$$

*Proof.* The exceptional set of $\tilde{d} \in S^{\mathcal{P}}_{\tilde{r}}(X)$ for a given $\tilde{r}$ has:

(a) $\tilde{d}$ whose basic product of parameters $(C, P_0)$ has $\prod_u \mathbf{p}_u \prod_t B_t > X$;
(b) $\tilde{d}$ with at least two prime factors from the same basic interval.

4.2.2. For (a), writing $(A_t, B_t]$ for the basic intervals in the unique basic product containing $\tilde{d}$, the exceptional set is majorized by the $\tilde{d}$ with $\tilde{d}\prod_t(B_t/A_t) \ge X$, and such $\tilde{d}$ are $\ge X/(1+1/C)^{\tilde{r}}$. We write $U$ for this, and the number of exceptions is bounded as $\le \Phi^{\mathcal{P}}_{\tilde{r}}(X) - \Phi^{\mathcal{P}}_{\tilde{r}}(U) \ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\big[1-1/(1+1/C)^{\tilde{r}}\big] \ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)(\log\log X)/C$, where we used the Sathe asymptotic (2) and $\tilde{r} \le (100\alpha_{\mathcal{P}}/99)\log\log X$.

4.2.3. Situation (b) is similar to the "comfortable spacing" of Smith (Definition 5.3 on page 44), which he considers at the end of the proof of Proposition 5.6 (page 56); see also [3, §4.2.1].

Writing $\lambda = 1 + 1/C$, the number $N_{\mathrm{b}}$ of $\tilde{d}$ in (b) is bounded by

$$\sum_{P_0<q<\sqrt{X}} \sum_{q/\lambda<p<q\lambda} \Phi^{\mathcal{P}}_{\tilde{r}-2}(X/pq).$$

When $p$ is small, say $p < X^{1/3}$, we have $\Phi^{\mathcal{P}}_{\tilde{r}-2}(X/pq) \ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)/pq$ following from the Sathe asymptotic (2), thus giving a bound of

$$\ll \sum_{P_0<q\le\sqrt{X}} \frac{\Phi^{\mathcal{P}}_{\tilde{r}}(X)}{q} \sum_{q/\lambda<p<q\lambda} \frac{1}{p}.$$

By the Mertens asymptotic (3) the $p$-sum here is (for some constant $c > 0$)

$$\ll \log\frac{\log q\lambda}{\log q/\lambda} + \exp\bigl(-c\sqrt{\log q}\bigr) \ll \frac{1}{C\log q} + \exp\bigl(-c\sqrt{\log q}\bigr),$$

and the resulting convergent $q$-sums give a contribution to $N_{\mathrm{b}}$ bounded as

$$\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\Bigl[\frac{1/C}{\log P_0} + \exp\bigl(-c\sqrt{\log P_0}\bigr)\Bigr] \ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\Bigl[\frac{1}{(\log\log X)^{99+\eta_0}} + \frac{1}{(\log\log X)^{999}}\Bigr]$$

where $\exp\bigl(-c\sqrt{\log P_0}\bigr) = \exp\bigl(-c(\log\log X)^{\eta_0/2}\bigr) \ll 1/(\log\log X)^{999}$ since $\eta_0 > 0$.

For the remaining $p > X^{1/3}$ we use the crude $\Phi^{\mathcal{P}}_{\tilde{r}-2}(X/pq) \ll X/pq$ and get a contribution bounded as

$$\sum_{P_0<q<\sqrt{X}} \sum_{\substack{q/\lambda<p<q\lambda \\ p>X^{1/3}}} \frac{X}{pq} \ll \sum_{X^{1/3}/2<q<\sqrt{X}} \frac{X}{q}\Bigl[\frac{1}{C(\log q)} + \exp(-c\sqrt{\log q})\Bigr] \ll \frac{X}{C(\log X)},$$

and since $X/\log X \ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)$ for our $\tilde{r}$, this then gives the desired bound. □

4.3. Next we show that the set of $\tilde{d}$ whose prime factorization is not suitably regular are sparse; namely, we expect that $\alpha_{\mathcal{P}} \log\log p_l$ for $p_l$ the $l$th prime factor (written increasingly) is suitably close to $l$. Smith requires this for many $l$ in his regularity condition [45, Definition 5.3], while we shall opt for a reduced version.[20]

Since this will not be the dominant error term, we make no attempt to optimize.[21]

**Lemma 4.3.1.** *The union of $S^{\mathcal{P}}_{\tilde{r}}(X)$ over $|\tilde{r} - \alpha_{\mathcal{P}}\log\log X| \ge \alpha_{\mathcal{P}}(\log\log X)/99$ has size bounded as $\ll \Phi^{\mathcal{P}}(X)/(\log X)^{\alpha_{\mathcal{P}}/20000}$.*

*Proof.* Using Tudesq's formulation (Lemma 3.3.1), we take $E_0$ as the set of all primes not in $\mathcal{P}$ and $e_0 = 0$, while $E_1$ is the set of primes in $\mathcal{P}$ up to $X$ and $e_1 = \tilde{r}$. Writing $a = (98\alpha_{\mathcal{P}}/99)(\log\log X)$, we can bound the number of integers in $S^{\mathcal{P}}(X)$

[20]Note that we do not, and indeed Smith did not, use his condition of "extravagant spacing" for our result; it only becomes relevant in his Section 7 when considering higher Selmer groups.

[21]A referee notes that Turán's method [52] might be able to show these more directly, without relying on the full force of the asymptotic in (2).

with fewer than $a$ prime factors as[22]

$$\ll \frac{X}{\log X}\sum_{e_1<a}\frac{(\alpha_{\mathcal P}\log\log X+\tilde T_2)^{e_1}}{e_1!}\ll\frac{(X/\log X)\cdot\exp(a\log\alpha_{\mathcal P}+a\log\log\log X)}{\exp(a\log(98\alpha_{\mathcal P}/99)+a\log\log\log X-a)}$$
$$=\frac{X}{\log X}\exp\bigl(a+a\log(99/98)\bigr)=\frac{X}{\log X}\bigl((\log X)^{\alpha_{\mathcal P}}\bigr)^{(98/99)[1+\log(99/98)]}$$
$$=\frac{X}{\log X}(\log X)^{\alpha_{\mathcal P}}\cdot\bigl((\log X)^{\alpha_{\mathcal P}}\bigr)^{(98/99)[1+\log(99/98)]-1}\ll\frac{\Phi^{\mathcal P}(X)}{(\log X)^{\alpha_{\mathcal P}(1/2)(1/99^2)}},$$

where we used Stirling's formula, and $(98/99)\bigl[1+\log(99/98)\bigr]-1\le-(1/2)(1/99^2)$ so as to save a (small) power of $(\log X)$.

For the large $e_1$ we write $a=(100\alpha_{\mathcal P}/99)(\log\log X)$ and find the contribution is

$$\ll \frac{X}{\log X}\sum_{e_1>a}\frac{(\alpha_{\mathcal P}\log\log X+\tilde T_2)^{e_1}}{e_1!}\ll\frac{(X/\log X)\cdot\exp(a\log\alpha_{\mathcal P}+a\log\log\log X)}{\exp(a\log(100\alpha_{\mathcal P}/99)+a\log\log\log X-a)}$$
$$=\frac{X}{\log X}\exp\bigl(a-a\log(100/99)\bigr)=\frac{X}{\log X}\bigl((\log X)^{\alpha_{\mathcal P}}\bigr)^{(100/99)[1-\log(100/99)]}$$
$$=\frac{X}{\log X}(\log X)^{\alpha_{\mathcal P}}\cdot\bigl((\log X)^{\alpha_{\mathcal P}}\bigr)^{(100/99)[1-\log(100/99)]-1}\ll\frac{\Phi^{\mathcal P}(X)}{(\log X)^{\alpha_{\mathcal P}(1/2)(1/100^2)}},$$

where we used that $(100/99)\bigl[1-\log(100/99)\bigr]-1\le-(1/2)(1/100^2)$. □

We recall $P_0=\exp\bigl((\log\log X)^{\eta_0}\bigr)$ and introduce $Q_1=\exp\exp\bigl((\log\log X)^{\eta_1}\bigr)$.

**Lemma 4.3.2.** *Suppose that $|\tilde r-\alpha_{\mathcal P}\log\log X|\le(\alpha_{\mathcal P}/99)\log\log X$.*

*For $\kappa_0>3$, the number of $\tilde d\in S^{\mathcal P}_{\tilde r}(X)$ that have more than $a_0=\kappa_0\alpha_{\mathcal P}\log\log P_0$ prime factors up to $P_0=\exp\bigl((\log\log X)^{\eta_0}\bigr)$ is*

$$\ll\frac{\Phi^{\mathcal P}_{\tilde r}(X)}{(\log P_0)^{\alpha_{\mathcal P}\kappa_0(\log\kappa_0-1-\theta)}}=\frac{\Phi^{\mathcal P}_{\tilde r}(X)}{(\log\log X)^{\alpha_{\mathcal P}\eta_0\kappa_0(\log\kappa_0-1-\theta)}}$$

*where $\theta=\log(100/99)$. We shall eventually take $\kappa_0=1/\eta_0\sqrt{\log 1/\eta_0}$ and $\eta_0\to0$, thus saving an arbitrarily large power of $(\log\log X)$.*

*The number of $\tilde d\in S^{\mathcal P}_{\tilde r}(X)$ with more than $a_1=3\alpha_{\mathcal P}\log\log Q_1$ prime factors up to $Q_1=\exp\exp\bigl((\log\log X)^{\eta_1}\bigr)$ is*

$$\ll\frac{\Phi^{\mathcal P}_{\tilde r}(X)}{(\log Q_1)^{\alpha_{\mathcal P}\cdot3(\log3-1-\theta)}}\ll\frac{\Phi^{\mathcal P}_{\tilde r}(X)}{(\log\log X)^{999}}.$$

*Proof.* For the first statement, we apply Tudesq's result (Lemma 3.3.1) with $E_0$ the set of all primes not in $\mathcal P$ and $e_0=0$, while $E_1$ is the set of primes in $\mathcal P$ up to $P_0$ with $e_1>a_0=\kappa_0\alpha_{\mathcal P}\log\log P_0$, and $E_2$ is the set of primes in $\mathcal P$ exceeding $P_0$ with $e_2=\tilde r-e_1$. This implies the number of exceptions is

$$\ll\frac{X}{\log X}\sum_{e_1>a_0}\frac{(\alpha_{\mathcal P}\log\log P_0+\tilde T_2)^{e_1}}{e_1!}\frac{(\alpha_{\mathcal P}\log\log X+\tilde T_2)^{\tilde r-e_1}}{(\tilde r-e_1)!}.$$

Since $\tilde r\le(100\alpha_{\mathcal P}/99)(\log\log X)$, by Sathe's asymptotic (2) we have

$$\frac{X}{\log X}\frac{(\alpha_{\mathcal P}\log\log X+\tilde T_2)^{\tilde r-e_1}}{(\tilde r-e_1)!}\ll\Phi^{\mathcal P}_{\tilde r+1-e_1}(X)\ll\Phi^{\mathcal P}_{\tilde r}(X)\cdot(100/99)^{e_1}.$$

[22]Here $\tilde T_2$ is Tudesq's $T_2$ plus a bound for $m_{\mathcal P}$ plus the error term in the Mertens asymptotic (3). It never plays a rôle because we only consider $(u+\tilde T_2)^e=u^e(1+\tilde T_2/u)^e$ with $e$ roughly of size $u$.

Then, since $e_1 > a_0 = \kappa_0\alpha_{\mathcal{P}}\log\log P_0 \geq 3\alpha_{\mathcal{P}}\log\log P_0$, the sum over $e_1$ is bounded (for sufficiently large $X$) by a decreasing geometric sequence in $e_1$, thus being dominated by $e_1 = a_0$. So the number of exceptions is

$$\begin{aligned}
&\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\frac{\big((100\alpha_{\mathcal{P}}/99)\log\log P_0 + (100/99)\tilde{T}_2\big)^{\lfloor a_0\rfloor}}{\lfloor a_0\rfloor!}\\
&\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\frac{\exp\big(a_0\log\log\log P_0 + a_0\log(100\alpha_{\mathcal{P}}/99)\big)}{\exp\big(a_0\log\log\log P_0 + a_0\log(\alpha_{\mathcal{P}}\kappa_0) - a_0\big)}\\
&= \frac{\Phi^{\mathcal{P}}_{\tilde{r}}(X)}{\exp\big(\alpha_{\mathcal{P}}(\log\log P_0)\cdot\kappa_0[\log\kappa_0 - 1 - \log(100/99)]\big)}
\end{aligned}$$

where we used Stirling's approximation and $a_0 = \kappa_0\alpha_{\mathcal{P}}\log\log P_0$.

The same argument for the second part of the Lemma readily gives a bound of $\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\cdot(100/99)^{a_1}\cdot(\alpha_{\mathcal{P}}\log\log Q_1 + \tilde{T}_2)^{\lfloor a_1\rfloor}/\lfloor a_1\rfloor!$ on the size of the exceptional set, and then using $a_1 = \kappa_1\alpha_{\mathcal{P}}\log\log Q_1$ with $\kappa_1 = 3$ in conjunction with Stirling's formula implies this is

$$\ll \frac{\Phi^{\mathcal{P}}_{\tilde{r}}(X)}{(\log Q_1)^{\alpha_{\mathcal{P}}\cdot 3(\log 3 - 1 - \log(100/99))}},$$

and the result then follows since $\log Q_1 = \exp\big((\log\log X)^{\eta_1}\big)$ exceeds any power of $(\log\log X)$ asymptotically (when $\eta_1 > 0$). □

4.3.3. Given a fixed parameter $\eta_1 > 0$, for a box $\bar{T}$ we let $k_1$ be the maximal index $t$ for which the basic interval $(A_t, B_t]$ has $B_t \leq Q_1 = \exp\exp\big((\log\log X)^{\eta_1}\big)$. We then accumulate our above analyses with the following result.

**Lemma 4.3.4.** *For $\eta_0, \eta_1 > 0$ and $\kappa_0 > 3$, the set of all $(X, \eta_0, \mathcal{P})$-boxes that have $|r - \alpha_{\mathcal{P}}\log\log X| \leq (\alpha_{\mathcal{P}}/99)(\log\log X)$ and additionally $k_0 \leq \kappa_0\alpha_{\mathcal{P}}\log\log P_0$ and $k_1 \leq 3\alpha_{\mathcal{P}}\log\log Q_1$ essentially covers $S^{\mathcal{P}}(X)$. The exceptional set has size*

$$\ll \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{\alpha_{\mathcal{P}}\eta_0\kappa_0(\log\kappa_0 - 1 - \theta)}} + \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{998}}$$

*where $\theta = \log(100/99)$.*

*Proof.* Apply Lemmata 4.3.1 and 4.3.2, taking the union over $\tilde{r}$ with the latter. □

4.4. We recollect our (likely empty) sequence of exceptional conductors of Dirichlet $L$-functions (§3.4), and show that the boxes that contain a multiple of one of them do not contribute many $\tilde{d}$. As Smith notes (Proposition 6.10), we can decontextualize the analysis from any discussion of Siegel zeros. We let $P_{\mathrm{s}} = \exp\big((\log\log X)^{\eta_{\mathrm{s}}}\big)$ for a given parameter $\eta_{\mathrm{s}} > 0$.

4.4.1. Consider a sequence of integers $\{w_l\}$ that has $w_0 \geq P_{\mathrm{s}}$ and $w_{l+1} \geq w_l^2$ for all $l \geq 0$. Write $w_l'$ for the part of $w_l$ coprime to $M_{\mathcal{P}}$, so that we have $w_l' \geq P_{\mathrm{s}}^{2^l}/M_{\mathcal{P}}$.

**Lemma 4.4.2.** *With the notation as given above, the set of $\tilde{d} \in S^{\mathcal{P}}_{\tilde{r}}(X)$ such that a $(X, \eta_0, \mathcal{P})$-box that represents $\tilde{d}$ also represents a multiple of some $w_l'$ has size*

$$\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X)\frac{1/C}{\log P_{\mathrm{s}}} \ll \frac{\Phi^{\mathcal{P}}_{\tilde{r}}(X)}{(\log\log X)^{99+\eta_{\mathrm{s}}}}.$$

*Proof.* Suppose $w_l' = p_1 \cdots p_m$ divides some squarefree integer represented by a box $\bar{T}$. Then for every $\tilde{d} \in \hat{T}$ there are (increasing) prime factors $q_1, \dots, q_m$ of $\tilde{d}$ such that $p_i = q_i$ if $p_i \le P_0$ and $p_i/(1+1/C) \le q_i \le p_i(1+1/C)$ otherwise (where $C$ is the compression factor $\lfloor (\log\log X)^{99} \rfloor$ as previously).

As with the proof of (b) in the above Lemma 4.2.1, we write $\lambda = 1 + 1/C$, and first note that the number of $\tilde{d}$ in such boxes is bounded as

$$\ll \prod_{\substack{p_i \le P_0 \\ p_i | w_l'}} \prod_{\substack{p_j > P_0 \\ p_j | w_l'}} \sum_{p_j/\lambda \le q_j \le p_j \lambda} \Phi^{\mathcal{P}}_{\tilde{r}-m}\Big(X \Big/ \prod_i p_i \prod_j q_j\Big).$$

For $w_l' \le \sqrt{X}$ we use Sathe's asymptotic (2) to bound $\Phi^{\mathcal{P}}_{\tilde{r}-m}$ in terms of $\Phi^{\mathcal{P}}_{\tilde{r}}$, getting

$$\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X) \prod_{\substack{p_i \le P_0 \\ p_i | w_l'}} \frac{1}{p_i} \prod_{\substack{p_j > P_0 \\ p_j | w_l'}} \sum_{p_j/\lambda \le q_j \le p_j \lambda} \frac{1}{q_j}.$$

The $q$-sums here are $\ll (1/C \log p_i) + \exp(-c\sqrt{\log p_i})$, and the considerations are dominated by the case when $w_l'$ is prime and exceeds $P_0$. A similar (cruder) argument works when $w_l' \ge \sqrt{X}$.

Since $w_l' \gg P_{\mathrm{s}}^{2^l}$ this gives a bound on the number of $\tilde{d}$ in such boxes as

$$\ll \Phi^{\mathcal{P}}_{\tilde{r}}(X) \cdot \Big[\frac{1/C}{2^l \log P_{\mathrm{s}}} + \exp\big(-c\sqrt{2^l \log P_{\mathrm{s}}}\big)\Big],$$

where the first term in brackets dominates since $\eta_{\mathrm{s}} > 0$. The sum over $l$ is convergent, and we conclude the statement of the Lemma. □

Note that Smith doesn't exploit the compression factors of the intervals here, and so his error estimate is $X/(\log\log X)^{\eta_{\mathrm{s}}}$ (or perhaps $X/(\log\log X)^{\eta_0}$), with this then to be balanced against the power-savings of $(\log\log X)$ that comes from the other parts of the argument. Contrarily, our version ensures that $\eta_{\mathrm{s}}$ does not directly affect our ultimate error bound (though we still will need $0 < \eta_{\mathrm{s}} < \eta_1 < \eta_0/2$).

4.5. Finally, we define *pleasant* boxes, and accumulate the results above to show that almost every squarefree $\tilde{d}$ is represented by such a box.

Let $\eta_0, \eta_1, \eta_{\mathrm{s}} > 0$ and $\kappa_0 > 3$ be given parameters. Then a $(\kappa_0, \eta_1, \eta_{\mathrm{s}})$-pleasant $(X, \eta_0, \mathcal{P})$-box of type $(\tilde{r}, k_0)$ is one with $|\tilde{r} - \alpha_{\mathcal{P}}(\log\log X)| \le (\alpha_{\mathcal{P}}/99)\log\log X$, and $k_0 \le \kappa_0 \alpha_{\mathcal{P}} \log\log P_0$ with $\log\log P_0 = \eta_0 \log\log\log X$, and $k_1 \le 3\alpha_{\mathcal{P}} \log\log Q_1$ where $\log\log Q_1 = (\log\log X)^{\eta_1}$ and $k_1$ is the largest index $t$ with $B_t \le Q_1$ for the basic interval $(A_t, B_t]$, and such that there is no $\tilde{d}$ represented by the box that is a multiple of the coprime-to-$\mathcal{P}$ part of some element of the sequence $\{\mathcal{M}_i\}$ of exceptional (Siegel) conductors that is $\ge P_{\mathrm{s}} = \exp\big((\log\log X)^{\eta_{\mathrm{s}}}\big)$.

**Lemma 4.5.1.** *The exceptional subset of $\tilde{d} \in S^{\mathcal{P}}(X)$ that are not represented by a $(\kappa_0, \eta_1, \eta_{\mathrm{s}})$-pleasant $(X, \eta_0, \mathcal{P})$-box has size (with $\theta = \log(100/99)$)*

$$\ll \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{\alpha_{\mathcal{P}}\eta_0\kappa_0(\log\kappa_0 - 1 - \theta)}} + \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{99}}.$$

*Proof.* Apply Lemmata 4.3.4 and 4.4.2. □

Recall also that every $\tilde{d} \in S^{\mathcal{P}}(X)$ is represented by at most one $(X, \eta_0, \mathcal{P})$-box.

## 5. Cutting up boxes

We recall $\mathcal{P}$ is the set of all primes lying in the residue classes $\mathcal{R}_{\mathcal{P}}$ modulo $M_{\mathcal{P}}$, and that a box $\bar{T}$ represents positive squarefree integers with prime factors from $\mathcal{P}$.

5.1. We now consider restricting boxes so that for all $l$ with $1 \le l \le \tilde{r}$ the $l$th prime factor is required to be in a specific residue class modulo $M_{\mathcal{P}}$. Note that this is somewhat different than the previous $\mathcal{P}$-restriction. For instance, we might have $\mathcal{P}$ contain all the primes that are 1 mod 8 and 3 mod 8, and now require that the first prime divisor is 1 mod 8, the second and third are 3 mod 8, while the fourth is 1 mod 8, etc.[23]

We thus define the $\mathcal{K}$-trimming of a box. For each $l$ with $1 \le l \le \tilde{r}$ we let $\mathcal{K}_l$ be a residue class modulo $M_{\mathcal{P}}$. Recall that the squarefree integers represented by a box $\bar{T}$ naturally lie in a Cartesian product $\prod_l T_l$, where each $T_l$ is a singleton set or the set of primes in a basic interval $(A_l, B_l]$ that are in $\mathcal{P}$. Assuming that each singleton set meets its requisite $\mathcal{K}$-condition (otherwise we just take $T(\mathcal{K})$ as empty), we define $T(\mathcal{K}) = \prod_l T_l(\mathcal{K})$ where $T_l(\mathcal{K})$ is the set of primes in the basic interval $(A_l, B_l]$ that are in the residue class specified by $\mathcal{K}_l$ (in other words, it is the subset of $T_l$ that meets said $\mathcal{K}_l$-condition).

This procedure of $\mathcal{K}$-trimming does not lose much in our estimates because we are simply taking progressions to a fixed modulus $M_{\mathcal{P}}$. (Note that we can specify $\tilde{d}$ to be in any desired coprime residue class to an auxiliary modulus $m$ via including $m$ in the modulus $M_{\mathcal{P}}$ (if necessary) and considering only the $\mathcal{K}$ that give the desired class. This does little more than induce an extra factor of $\varphi(m)$ in various estimates. We can similarly restrict to a non-coprime residue class by only considering boxes that have the common primes in the singleton sets. This then gives a method of handling non-twist-minimal curves, as suggested at the end of §1.1.4.)

5.2. We let $\mathcal{L}$ be a set of Legendre symbol specifications, meaning for $1 \le i < j \le \tilde{r}$ we take $\mathcal{L}_{ij} \in \{\pm 1\}$. We then define the $(\mathcal{K}, \mathcal{L})$-restriction of a box. This is the set of $\tilde{d} \in \hat{T}(\mathcal{K})$ with $\tilde{d} = p_1 \cdots p_{\tilde{r}}$ such that $(p_i|p_j) = \mathcal{L}_{ij}$ for all $1 \le i < j \le \tilde{r}$. This is more severe than the $\mathcal{K}$-trimming, as in general it will no longer be a Cartesian product. However, if we instead only specify Legendre symbol conditions for a suitable subset of the $(i, j)$, then we will indeed retain the Cartesian product aspect. Moreover, it is convenient to simultaneously limit the trimming effect of $\mathcal{K}$.

5.2.1. For a box $\bar{T}$, recall $k_0$ is the number of primes up to $P_0 = \exp\big((\log\log X)^{\eta_0}\big)$, with these primes associated to singleton sets in the Cartesian product, while $k_1$ is the largest $l$ such that the $l$th basic interval has $B_l \le Q_1 = \exp\exp\big((\log\log X)^{\eta_1}\big)$.

We define the $(\mathcal{K}, \mathcal{L}, [k_0, k_1])$-trimming of a box, which we denote as $\tilde{T}(\mathcal{K}, \mathcal{L})_{k_0}^{k_1}$. This corresponds to the set of $\tilde{d}$ represented by $\bar{T}$ with $\tilde{d} = p_1 \cdots p_{\tilde{r}}$ such that $p_i$ is in the residue class specified by $\mathcal{K}$ for $i \le k_1$, and $(p_i|p_j) = \mathcal{L}_{ij}$ for $i, j$ that satisfy $1 \le i < j \le k_1$ and $i \le k_0$. We let $\mathcal{I}(k_0, k_1)$ be this set of $(i, j)$-pairs.

The latter condition $i \le k_0$ ensures that the prime $p_i$ will be from a singleton set in the product for $T$. We thus have a Cartesian product

$$\tilde{T}(\mathcal{K}, \mathcal{L})_{k_0}^{k_1} = \prod_l \tilde{T}_l(\mathcal{K}, \mathcal{L})_{k_0}^{k_1}.$$

[23]Smith only requires that the $l$th prime have a specific Legendre symbol specification with each of the bad primes, which suffices when considering the 2-Selmer group.

For $l \le k_0$ this set $\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ is simply the singleton $T_l$ if the $(\mathcal{K},\mathcal{L})$-conditions are met and is empty otherwise. Meanwhile, for $l$ with $k_0 < l \le k_1$ the set $\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ is the subset of $T_l$ that is specified by: the Legendre symbol specifications from $\mathcal{L}$ for index pairs $(j,l) \in \mathcal{I}(k_0,k_1)$ with $j \le k_0$ (which are thus specifications with respect to singleton primes); and the residue condition from $\mathcal{K}_l$. These are somewhat arbitrary sets of primes over which we need not have much control. One expects that $\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ is $1/\xi_{\mathcal{P}}2^{k_0}$ as large as $T_l$, but this is not easy to show in general.[24]

Finally, for $l > k_1$ we simply have $\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1} = T_l$, which thus consists of all the primes in a basic interval that are in $\mathcal{P}$. The regularity of such primes will allow us to show that the partitioning of $T$ according to the index pairs not in $\mathcal{I}(k_0,k_1)$ is indeed fairly uniform, reducing the size roughly by the expected powers of 2 and $\xi_{\mathcal{P}}$.

5.3. We now turn to showing the main result about trimmed boxes, namely that when the trimmed box $\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ is further restricted by all the $(\mathcal{K},\mathcal{L})$-conditions, the size is therein reduced by the expected powers of 2 and $\xi_{\mathcal{P}}$.

Smith uses an inductive scheme in his Proposition 6.3, while we (similar to Kane in particular) instead consider the product over $\big[1+(p_i|p_j)\big]$, namely

$$\prod_{(i,j)\in\mathcal{I}_{\tilde{r}}\setminus\mathcal{I}(k_0,k_1)} \big[1+(p_i|p_j)\big]$$

where $\mathcal{I}_{\tilde{r}}$ is the set of all $(i,j)$-pairs with $1 \le i < j \le \tilde{r}$. At any rate, the crux of Smith's argument is that the exclusion of $\mathcal{I}(k_0,k_1)$ (which we handle later by a combinatorial argument involving the fixity of the 2-Selmer rank under permutations) allows suitable uniformity to be adduced.

5.3.1. As a motivation for the forthcoming proof, we will end up needing to estimate sums either of a bilinear form

$$\sum_{p_m}\sum_{p_n} a_m b_n (p_m|p_n) \tag{4}$$

where $p_m \in T_m$ and $p_n \in T_n$ with $m < n$ and $\{a_m\},\{b_n\}$ are arbitrary sequences bounded by 1, or the ostensibly simpler congruential form

$$\sum_{p_n} (M|p_n) \tag{5}$$

for some modulus $M$.

We have a natural ambient loss of

$$2^{k_0k_1} \le 2^{(\kappa_0\eta_0 \log\log\log X)\cdot(\kappa_1 \log\log Q_1)} = \exp\big(3\kappa_0\eta_0(\log 2)(\log\log\log X)(\log\log Q_1)\big)$$

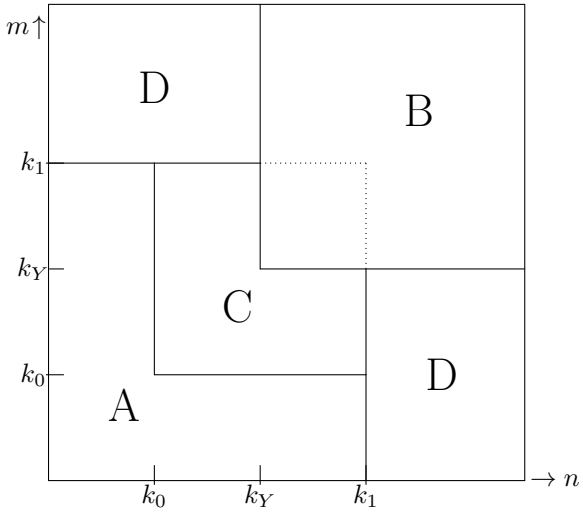
to overcome, corresponding to the amount we cannot directly access from the pairs in $\mathcal{I}(k_0,k_1)$. Moreover, our losses will have to take into account the number of relevant subsets of $\mathcal{I}_{\tilde{r}}\setminus\mathcal{I}(k_0,k_1)$ that occur when multiplying out $\prod\big[1+(p_i|p_j)\big]$: this number is bounded by $2^{\binom{\tilde{r}}{2}}$ – though this is quite large, it will suffices in ranges where one of the primes exceeds $\exp\big((\log\log X)^3\big)$; meanwhile the number of such subsets that do not involve an index exceeding $k_1$ is bounded by $2^{\binom{k_1}{2}}$, and to handle

[24] One can readily show this expectation upon assuming the Generalized Riemann Hypothesis, at least for $l$ that exceed $k_0$ by a sufficient amount (roughly that the primes in $T_l$ exceed the square of the modulus, the latter being the prime in $T_{k_0}$).

these it suffices to require $\eta_0 > 2\eta_1$. (Perhaps by re-arranging the argument one only needs $\eta_0 > \eta_1$, but such an improvement would ultimately be irrelevant).

We also lose a negligible $C^2 = (\log\log X)^{198}$ from interval splitting, and in the bilinear case $(\log p_m)(\log p_n)$ from the density of primes. Finally, there is a loss of $\xi_{\mathcal{P}}^{k_1}$ coming from the $\mathcal{K}$-splitting on the first $k_1$ primes.

5.3.2. In the Figure we divided the $(m,n)$-range (also reflecting to $m > n$ for symmetrical convenience) into 4 parts. This depends on a splitting at the parameter $Y = \exp\big((\log\log X)^{998}\big)$, and we write $k_Y$ for the largest index $l$ such that the right endpoint $B_l$ of the basic interval $\bar{T}_l$ has $B_l \le Y$.




The first region **A** has $m \le k_0$ and $n \le k_1$ (and its reflection) and corresponds to the indices in $\mathcal{I}(k_0, k_1)$. These will thus not appear in the current analysis, but rather in a permutation argument in the next section.

The second region **B** has $m, n \ge k_Y$ where from the bilinear sum (4) we will save $p_m^{1/9} \ge Y^{1/9}$, and since we have $Y = \exp\big((\log\log X)^{998}\big) \gg 2^{r^2}$ this easily suffices. The third region **C** has $k_0 < m, n \le k_1$ and here we save $\gg P_0^{1/9}$ in the bilinear estimation. Since we assume $\eta_0 > 2\eta_1$, while $\log\log Q_1 \ge k_1/3$ for pleasant boxes, this savings is

$$\begin{aligned}\exp\big((\log P_0)/9\big) &= \exp\big((\log\log X)^{\eta_0}/9\big)\\ &\gg \exp\big(9(\log\log X)^{2\eta_1}\big) = \exp\big(9(\log\log Q_1)^2\big) \ge \exp(k_1^2).\end{aligned}$$

This exceeds the losses from the number of relevant subsets.

The fourth region **D** has $m \le k_Y$ and $n > k_1$ and its reflection. For this region we can estimate the sum in (5) by results for primes in arithmetic progressions. When there is no exceptional zero, we save a dramatic amount $\exp\big(c\sqrt{\log Q_1}\big)$, which is asymptotically (much) larger than $\exp(\tilde{r}^2)$. Even when there is an exceptional zero, if the conductor has $M \le P_{\mathrm{s}}$ we can use the "trivial" bound on zeros and obtain an adequate result, namely a savings of size $Q_1^{1/M^\epsilon} \ge \exp\big((\log Q_1)/P_s^\epsilon\big)$. If we then write $P_{\mathrm{s}} = \exp\big((\log\log X)^{\eta_{\mathrm{s}}}\big)$, our savings requirement can be phrased as saying that $\log\log Q_1 = (\log\log X)^{\eta_1}$ should sufficiently exceed $\log P_{\mathrm{s}}^\epsilon = \epsilon(\log\log X)^{\eta_{\mathrm{s}}}$, and so it suffices to take $\eta_{\mathrm{s}} < \eta_1$ (note also that we can take $\epsilon = 1/2$ without any

difficulty).[25] We will still need to exclude exceptional conductors with $M \ge P_{\mathrm{s}}$, as codified above in our definition of pleasantness in §4.5.

Note that we can actually use prime distribution in arithmetic progressions for many more $(m,n)$-pairs than those in **D**. Indeed, we could typically expect to save something like $\exp\bigl(\frac{\log\exp(e^n)}{\log\exp(e^m)}\bigr) = \exp(e^{n-m})$ from primes of size $\exp(e^n)$ in arithmetic progressions to moduli $\exp(e^m)$, so when $n - m \gg (\log\log X)^{\eta_1}\log\log\log X$ (roughly) we could expect save the desired factor. Thus for $m, n \gg k_0k_1$ it is only rather close to the diagonal that usage of the bilinear bound is necessitated.

5.4. Let us now state and show the desired result.

**Proposition 5.4.1.** *Let $\bar{T}$ be a $(\kappa_0,\eta_1,\eta_{\mathrm{s}})$-pleasant $(X,\eta_0,\mathcal{P})$-box and $\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ its $(\mathcal{K},\mathcal{L},[k_0,k_1])$-trimming. Assume that $\eta_0/2 > \eta_1 > \eta_{\mathrm{s}} > 0$. Then*

$$\#T(\mathcal{K},\mathcal{L}) = \frac{\#\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}/\xi_{\mathcal{P}}^{\tilde{r}-k_1}}{2^{\binom{\tilde{r}}{2}-\binom{k_0}{2}-k_0(k_1-k_0)}} + O\Bigl(\frac{\#T(\mathcal{K})(\log\log X)^{198}}{2^{\binom{\tilde{r}}{2}-\binom{k_0}{2}-k_0(k_1-k_0)}}\cdot\frac{\exp\bigl(5(\log\log X)^{2\eta_1}\bigr)}{\exp\bigl((\log\log X)^{\eta_0}/9\bigr)}\Bigr).$$

*Proof.* We consider

$$V_{\mathcal{L}}(\vec{p}) = \prod_{(i,j)\in\mathcal{I}_{\tilde{r}}\setminus\mathcal{I}(k_0,k_1)} \bigl[1+\mathcal{L}_{ij}(p_i|p_j)\bigr],$$

which is used to pick off the Legendre conditions from the Cartesian product

$$\prod_{a\le k_1}\tilde{T}_a(\mathcal{K},\mathcal{L})_{k_0}^{k_1} \times \prod_{b>k_1} T_b(\mathcal{K}) \tag{6}$$

so that

$$\#T(\mathcal{K},\mathcal{L}) = \frac{1}{2^w}\prod_{a\le k_1}\sum_{p_a\in\tilde{T}_a(\mathcal{K},\mathcal{L})_{k_0}^{k_1}}\prod_{b>k_1}\sum_{p_b\in T_b(\mathcal{K})} V_{\mathcal{L}}(\vec{p})$$

where $w = \binom{\tilde{r}}{2} - \binom{k_0}{2} - k_0(k_1-k_0)$ is the number of the index pairs in the product.

Multiplying $V_{\mathcal{L}}(\vec{p})$ out, the component with all 1's gives the main term as

$$\begin{aligned}\frac{1}{2^w}\prod_{l\le k_1}\#\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}\prod_{l>k_1}\#T_l(\mathcal{K}) &= \frac{1}{2^w}\prod_{l\le k_1}\#\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1}\prod_{l>k_1}\#T_l\frac{\#T_l(\mathcal{K})}{\#T_l}\\ &= \frac{1}{2^w}\#\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}\prod_{l>k_1}\frac{\#T_l(\mathcal{K})}{\#T_l}.\end{aligned}$$

We can handle the latter product using estimates on primes in arithmetic progressions, here to the fixed modulus $M_{\mathcal{P}}$ for primes exceeding $Q_1$. Each term in the product is then $1/\xi_{\mathcal{P}} + O\bigl(\exp(-c\sqrt{\log Q_1})\bigr)$, yielding a negligible error.

[25] My impression concerning the reason that Smith's proof uses the ineffective Siegel bound is that he doesn't particularly exploit that $Q_1$ will be significantly larger than $M$. His usage of the Siegel bound occurs at the top of page 64, where he makes the savings estimate of $x^{\beta} \gg \exp(\frac{\log t_i}{t^c})$, where $c > 0$ is the Siegel exponent (having $c = 1/2$ would thus be an effective result). His $t_i$ here corresponds to our $Q_1$, so that its log is $\exp\bigl((\log\log X)^{\eta_1}\bigr)$, and his $t$ to our $\exp\bigl((\log\log X)^{\eta_{\mathrm{s}}}\bigr)$. He then merely exploits his condition (6) that $\log t_i \gg t^{c_6}$ (which implicitly requires him to later utilize small $c$), while in fact by taking $\eta_1 > \eta_{\mathrm{s}}$ these are on much different scales. This is then replicated in condition (iii) of [3, Proposition 5.7], and thus their result is also ineffective in the form given.

5.4.2. Otherwise, each component of $V_{\mathcal{L}}(\vec{p})$ when multiplied out corresponds to some nonempty subset $\mathcal{S}$ of $\mathcal{I}_{\tilde{r}} \backslash \mathcal{I}(k_0, k_1)$. Let $n$ be the largest index appearing in a pair in $\mathcal{S}$, and let $m$ be the largest index such that $(m,n) \in \mathcal{S}$.

When $(m,n)$ is in **B** or **C** (so in particular $m > k_0$) we will apply the bilinear bound. To uniformize notation over indicies, we write $U_l$ for the components in the Cartesian product (6). The underlying sets $U_m$ and $U_n$ need not be too regular in this case. The relevant sum is bounded as

$$\ll \frac{1}{2^w} \prod_{l \neq m,n} \sum_{p_l \in U_l} \Big| \sum_{p_m \in U_m} \sum_{p_n \in U_n} a_{p_m} b_{p_n} (p_m | p_n) \Big|$$

for some coefficients $a_{p_m}$ and $b_{p_n}$ that respectively take into account the other Legendre symbols involving $p_m$ and $p_n$ (thus depending on the $p_l$). In particular, these coefficients are bounded in size by 1.

By Lemma 3.4.3 the contribution for each such $\mathcal{S}$ is thus

$$\ll \frac{1}{2^w} \prod_{l \neq m,n} \#U_l \cdot \frac{B_m B_n}{B_m^{1/9}}$$

where the $B_m$ is the right endpoint of the $m$th basic interval for $\bar{T}_m$ and similarly for $n$. For the $l \neq m,n$ we apply the trivial bound $\#U_l \leq \#T_l(\mathcal{K})$, with this being an equality for $l > k_1$. Meanwhile, since $T_m(\mathcal{K})$ contains all the primes in the basic interval $\bar{T}_m$ in the $\mathcal{K}_m$ residue class, by estimates for primes in arithmetic progresions (to the fixed modulus $M_{\mathcal{P}}$) we have $B_m \ll \#T_m(\mathcal{K}) \cdot C \log \#T_m(\mathcal{K})$, and similarly for $n$ (here $C$ is the compression factor $\lfloor (\log\log X)^{99} \rfloor$).

The number of sets $\mathcal{S}$ such that $(m,n)$ is in **B** is trivially bounded as $2^{\binom{\tilde{r}}{2}}$, and their total contribution is thereby bounded as

$$\prod_l \#T_l(\mathcal{K}) \cdot \frac{2^{\binom{\tilde{r}}{2}} (C \log X)^2}{2^w Y^{1/9}} = \frac{\#T(\mathcal{K})}{2^w} \cdot \frac{2^{\binom{\tilde{r}}{2}} (C \log X)^2}{Y^{1/9}} \ll \frac{\#T(\mathcal{K})}{2^w (\log X)^{999}},$$

as $Y = \exp\big((\log\log X)^{998}\big)$ dominates the discussion. The number of sets $\mathcal{S}$ such that $(m,n)$ is in **C** is bounded as $2^{\binom{k_1}{2}}$, and their total contribution is bounded as

$$\ll \frac{\#T(\mathcal{K})}{2^w} \frac{2^{\binom{k_1}{2}} (C \log Q_1)^2}{P_0^{1/9}}$$

where $(\log Q_1)$ is $\exp\big((\log\log X)^{\eta_1}\big)$ and $P_0$ is $\exp\big((\log\log X)^{\eta_0}\big)$, and for pleasant boxes $k_1 \leq 3 \log\log Q_1$ implies $2^{k_1^2/2} \leq \exp(4(\log\log Q_1)^2) = \exp\big(4(\log\log X)^{2\eta_1}\big)$, so that $\eta_0 > 2\eta_1$ gives the desired bound.

5.4.3. Otherwise, the contribution from a set $\mathcal{S}$ is bounded as

$$\ll \frac{1}{2^w} \prod_{l \neq n} \sum_{p_l \in U_l} \Big| \sum_{p_n \in U_n} (M | p_n) \Big| \tag{7}$$

where here $M$ is the product of the primes taken from the $U_j$ with $(j,n) \in \mathcal{S}$. In particular, each such prime from $U_j$ is bounded by $Y$ (else we would be in case **B**), so the modulus is no more than $Y^{k_Y}$. Since $n > k_1$ we have that $U_n = T_n(\mathcal{K})$ is the set of primes in a basic interval that are in the residue class specified by $\mathcal{K}_n$.

Ergo, we can apply results about primes in arithmetic progressions. Our assumption that the box is $\eta_s$-pleasant implies that the inner sum in (7) is

$$\ll \frac{B_n}{B_n^{1/\sqrt{P_s}}} + B_n \exp\Big(-c\frac{\log B_n}{\sqrt{\log B_n} + 9\log M}\Big)(\log B_n M)^4.$$

Since $n > k_1$ we have $B_n \ge Q_1$ and so $\log B_n \gg \exp\big((\log\log X)^{\eta_1}\big)$, which much exceeds $\log M \le k_Y \log Y \ll (\log\log X)^{999}$. Meanwhile, we can similarly note the bound $(\log B_n)/\sqrt{P_s} \gg \exp\big((\log\log X)^{\eta_1}\big)\exp\big(-(\log\log X)^{\eta_s}/2\big)$, so that $\eta_1 > \eta_s$ implies the inner sum in (7) is

$$\ll B_n \exp\big[-c\exp\big((\log\log X)^{\eta_1}/2\big)\big].$$

We have that $B_n \ll \#T_n(\mathcal{K}) \cdot C\log \#T_n(\mathcal{K}) \ll \#T_n(\mathcal{K})(\log\log X)^{99}(\log X)$ and again use the trivial bound $\#U_l \le \#T_l(\mathcal{K})$ for $l \ne n$. Thus the above bound (7) is

$$\ll \frac{1}{2^w}\frac{\#T(\mathcal{K})}{\#T_n(\mathcal{K})} \cdot \#T_n(\mathcal{K})\frac{(\log X)(\log\log X)^{99}}{\exp\exp\big((\log\log X)^{\eta_1/2}\big)} \ll \frac{\#T(\mathcal{K})/2^w}{\exp\exp\big((\log\log X)^{\eta_1/3}\big)}$$

(indeed, there is an extra exponentiation in the denominator compared to the bound from the bilinear estimate, as due to $n \ge k_1$ we save $\exp\big(-c\sqrt{\log Q_1}\big)$, which is on a different exponential scale than the $Y$ or $P_0$ of before). Multiplying by $2^{\binom{\tilde r}{2}}$ for the number of sets $\mathcal{S}$ is harmless, and we conclude the Proposition. □

It will be slightly more convenient to rearrange the above formula, and we do so upon writing $v = \binom{k_0}{2} + k_0(k_1 - k_0)$.

**Corollary 5.4.4.** *Let $\bar T$ be an $(\kappa_0, \eta_1, \eta_s)$-pleasant $(X, \eta_0, \mathcal{P})$-box and $\tilde T(\mathcal{K}, \mathcal{L})_{k_0}^{k_1}$ its $(\mathcal{K}, \mathcal{L}, [k_0, k_1])$-trimming. Assume that $\eta_0/2 > \eta_1 > \eta_s > 0$. Then*

$$2^{\binom{\tilde r}{2}-v}\xi_{\mathcal{P}}^{\tilde r - k_1}\#T(\mathcal{K},\mathcal{L}) = \#\tilde T(\mathcal{K},\mathcal{L})_{k_0}^{k_1} + O\Big(\#T(\mathcal{K})\xi_{\mathcal{P}}^{\tilde r-k_1} \cdot \frac{\exp\big(6(\log\log X)^{2\eta_1}\big)}{\exp\big((\log\log X)^{\eta_0}/9\big)}\Big).$$

5.5. Although we will not need it until §7, it is also useful to record an upper bound on $\#T(\mathcal{K}, \mathcal{L})$.

**Lemma 5.5.1.** *Let $\bar T$ be a $(\kappa_0, \eta_1, \eta_s)$-pleasant $(X, \eta_0, \mathcal{P})$-box, and assume that the parameters satisfy $\eta_0/2 > \eta_1 > \eta_s > 0$. Then*

$$\#T(\mathcal{K},\mathcal{L}) \ll \frac{\#T(\mathcal{K})}{2^{\binom{\tilde r}{2}}}2^v$$

*where $v = \binom{k_0}{2} + k_0(k_1 - k_0)$. (This bound is $2^v$ more than the expected amount).*

*Proof.* From Proposition 5.4.1 we have

$$\#T(\mathcal{K},\mathcal{L}) = \frac{\#\tilde T(\mathcal{K},\mathcal{L})_{k_0}^{k_1}/\xi_{\mathcal{P}}^{\tilde r-k_1}}{2^{\binom{\tilde r}{2}-v}} + O\Big(\frac{\#T(\mathcal{K})}{2^{\binom{\tilde r}{2}-v}} \cdot \frac{\exp\big(6(\log\log X)^{2\eta_1}\big)}{\exp\big((\log\log X)^{\eta_0}/9\big)}\Big), \tag{8}$$

and since $\eta_0 > 2\eta_1$ the error term fits into our bound here.

Meanwhile, we have

$$\tilde T(\mathcal{K},\mathcal{L})_{k_0}^{k_1} = \prod_l \tilde T_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1} = \prod_{l\le k_1}\tilde T_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1} \times \prod_{l>k_1} T_l,$$

and using $\#\tilde{T}_l(\mathcal{K},\mathcal{L})_{k_0}^{k_1} \le \#T_l(\mathcal{K})$ this implies

$$\#\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1} \le \prod_{l\le k_1} \#T_l(\mathcal{K}) \prod_{l>k_1} \#T_l(\mathcal{K}) \frac{\#T_l}{\#T_l(\mathcal{K})} = \#T(\mathcal{K}) \prod_{l>k_1} \frac{\#T_l}{\#T_l(\mathcal{K})}.$$

We can again estimate the latter product by results on primes in arithmetic progressions to the fixed modulus $M_{\mathcal{P}}$ for primes exceeding $Q_1$, finding each term is $\xi_{\mathcal{P}} + O\big(\exp(-c\sqrt{\log Q_1})\big)$. We thus conclude $\#\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1} \ll \#T(\mathcal{K})\xi_{\mathcal{P}}^{\tilde{r}-k_1}$ and so

$$\frac{\#\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}/\xi_{\mathcal{P}}^{\tilde{r}-k_1}}{2^{\binom{\tilde{r}}{2}-v}} \ll \frac{\#T(\mathcal{K})\xi_{\mathcal{P}}^{\tilde{r}-k_1}/\xi_{\mathcal{P}}^{\tilde{r}-k_1}}{2^{\binom{\tilde{r}}{2}-v}} = \frac{\#T(\mathcal{K})}{2^{\binom{\tilde{r}}{2}}} 2^v,$$

with then the same asymptotic bound holding for $\#T(\mathcal{K},\mathcal{L})$ by (8). □

## 6. Averaging over permutations

As we exposit in §7, the 2-Selmer rank of $E_d$ is determined by the values of the Legendre symbols $(p_i|p_j)$ for primes $p_i, p_j$ dividing $d$, along with the values of the Legendre symbols $(q|p_i)$ for bad primes $q \in \Omega$, and the sign of $d$. The Legendre symbols are a weaker condition than our requirements from $\mathcal{K}$ that $p_i$ be in a given residue class modulo $M_{\mathcal{P}}$. Thus for any given $(\mathcal{K},\mathcal{L})$ the 2-Selmer rank is the same for every $d \in S_{\tilde{r}}^{\mathcal{P}}(\infty)$ that satisfies said $(\mathcal{K},\mathcal{L})$-conditions, and we write $s_{\tilde{r}}^{+}(\mathcal{K},\mathcal{L})$ for it. The 2-Selmer rank is similarly the same for every $-d \in S_{\tilde{r}}^{\mathcal{P}}(\infty)$ for which $|d|$ satisfies the $(\mathcal{K},\mathcal{L})$-conditions, and we write $s_{\tilde{r}}^{-}(\mathcal{K},\mathcal{L})$ for it.

More explicitly, the 2-Selmer rank is the dimension of the kernel of a square matrix of size $2(\tilde{r}+\#\tilde{\Omega})$ with $\mathbf{F}_2$ entries corresponding to Legendre symbols, with 2 rows/columns for each odd prime involved (and 3 for $p=2$ and 1 for the infinite place). As such, to compute the 2-Selmer rank it does not matter how we permute the rows and columns. This latter observation is keenly exploited by Smith.

6.1. We let $\mathcal{D}(\tilde{r},\mathcal{P})$ be the set of all possible choices $(\mathcal{K},\mathcal{L})$ for given $\tilde{r}$ and $\mathcal{P}$. This has size $2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}$ where $\xi_{\mathcal{P}}$ is the number of residue classes in $\mathcal{R}_{\mathcal{P}}$.

6.1.1. Given a pleasant box $\bar{T}$, a fixed sign $\varepsilon$ for $d$, and a fixed value of $s$, we want to estimate

$$\sum_{\substack{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}).$$

However, we do not have good control over the individual $\#T(\mathcal{K},\mathcal{L})$, as Proposition 5.4.1 only describes their relative size in $\tilde{T}(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$. To circumvent this difficulty, we consider the effect of permuting $(\mathcal{K},\mathcal{L})$ by $\sigma \in \mathrm{Sym}_{\tilde{r}}$, where this set is the symmetric group on the indices $j$ with $1 \le j \le \tilde{r}$. (Note that many of the $T(\mathcal{K}^\sigma,\mathcal{L}^\sigma)$ will be empty, for in particular the permutation $\sigma$ must respect the $\mathcal{K}$-conditions for the singleton sets $T_l$ for $l \le k_0$).

Since the 2-Selmer rank is fixed under such permutations, we then have

$$\tilde{r}! \sum_{\substack{(\mathcal{K},\mathcal{L})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) = \sum_{\substack{(\mathcal{K},\mathcal{L})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \sum_{\sigma\in\mathrm{Sym}_{\tilde{r}}} \#T(\mathcal{K}^\sigma,\mathcal{L}^\sigma),$$

and will be able to demonstrate adequate control over the sizes of $T$ averaged over $\sigma$.

6.2. We start by noting a purely combinatorial Lemma. For a given $\tilde{d}$ represented by a box $\bar{T}$ we define

$$W_{\tilde{d}}(\mathcal{K},\mathcal{L}) = \{\sigma \in \mathrm{Sym}_{\tilde{r}} \mid \tilde{d} \in \tilde{T}(\mathcal{K}^\sigma, \mathcal{L}^\sigma)_{k_0}^{k_1}\}.$$

Smith's combinatorial result shows that most $(\mathcal{K},\mathcal{L})$ have $W_{\tilde{d}}$ of nearly its expected size, saving a factor of essentially $2^{k_0}k_1^2/\tilde{r}$ in a mean-square estimate for it. As a convenience, we write $v = \binom{k_0}{2} + k_0(k_1-k_0)$.

**Lemma 6.2.1.** *Let $\bar{T}$ be a box and assume $2^{k_0+1}\xi_{\mathcal{P}}k_1^2 \le \tilde{r}$. For any $\tilde{d} \in \hat{T}$ we have*

$$\sum_{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})} \Big[\frac{\tilde{r}!/\xi_{\mathcal{P}}^{k_1}}{2^v} - \#W_{\tilde{d}}(\mathcal{K},\mathcal{L})\Big]^2 \le 2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}} \cdot \frac{\tilde{r}!^2/\xi_{\mathcal{P}}^{2k_1}}{2^{2v}} \frac{2^{k_0+1}\xi_{\mathcal{P}}k_1^2}{\tilde{r}}.$$

This is essentially Smith's Proposition 6.7, or the content of [3, Proposition 5.8] (they apply Cauchy's inequality in their statement of the result). Note that $\tilde{r}$ is typically of size $(\log\log X)$, while $k_1 \le 3(\log\log X)^{\eta_1}$ and $k_0 \le \kappa_0\eta_0 \log\log\log X$ for pleasant boxes. Taking $\eta_1 \to 0$ and $\kappa_0\eta_0 \to 0$ (in an appropriate way) will then ensure that the $\tilde{r}$-factor dominates.

*Proof.* For $\sigma \in \mathrm{Sym}_{\tilde{r}}$ we write $S(\sigma)$ for the set of $(\mathcal{K},\mathcal{L})$ with $\tilde{d}$ in $\tilde{T}(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1}$. We can note that the mean value of $W_{\tilde{d}}(\mathcal{K},\mathcal{L})$ is as suggested, namely

$$\sum_{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})} \#W_{\tilde{d}}(\mathcal{K},\mathcal{L}) = \sum_{\sigma\in\mathrm{Sym}_{\tilde{r}}} \#S(\sigma) = \frac{\tilde{r}!/\xi_{\mathcal{P}}^{k_1}}{2^{\binom{k_0}{2}+k_0(k_1-k_0)}} \cdot 2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}},$$

as either sum in question is the size of

$$\big\{(\sigma,\mathcal{K},\mathcal{L}) \mid \tilde{d} \in \tilde{T}(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1}\big\},$$

and each $\sigma$-section has the same size, as the allowable $(\mathcal{K},\mathcal{L})$-specifications are given by $v = \binom{k_0}{2}+k_0(k_1-k_0)$ Legendre conditions and $k_1$ residue conditions modulo $M_{\mathcal{P}}$.

6.2.2. We proceed to compute the mean value of $\#W_{\tilde{d}}(\mathcal{K},\mathcal{L})^2$. This is the number of pairs $(\sigma_1,\sigma_2)$ with $\tilde{d}$ in both $\tilde{T}(\mathcal{K}^{\sigma_1},\mathcal{L}^{\sigma_1})_{k_0}^{k_1}$ and $\tilde{T}(\mathcal{K}^{\sigma_2},\mathcal{L}^{\sigma_2})_{k_0}^{k_1}$. We can invert the problem by starting with the two permutations $\sigma_1,\sigma_2$ and writing $S(\sigma_1,\sigma_2)$ for the set of $(\mathcal{K},\mathcal{L})$ with $\tilde{d}$ in both $\tilde{T}(\mathcal{K}^{\sigma_1},\mathcal{L}^{\sigma_1})_{k_0}^{k_1}$ and $\tilde{T}(\mathcal{K}^{\sigma_2},\mathcal{L}^{\sigma_2})_{k_0}^{k_1}$. Indeed, we then have

$$\sum_{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})} \#W_{\tilde{d}}(\mathcal{K},\mathcal{L})^2 = \sum_{\sigma_1,\sigma_2\in\mathrm{Sym}_{\tilde{r}}}\sum \#S(\sigma_1,\sigma_2).$$

corresponding to two different partitions of

$$\big\{(\sigma_1,\sigma_2,\mathcal{K},\mathcal{L}) \mid \tilde{d} \in \tilde{T}(\mathcal{K}^{\sigma_1},\mathcal{L}^{\sigma_1})_{k_0}^{k_1} \cap \tilde{T}(\mathcal{K}^{\sigma_2},\mathcal{L}^{\sigma_2})_{k_0}^{k_1}\big\}.$$

Contrary to the $S(\sigma)$, the $S(\sigma_1,\sigma_2)$ can be of different sizes, depending on how many independent conditions are specified. We proceed to bound $S(\sigma_1,\sigma_2)$ based upon how many indices the permutations simultaneously map to $\le k_1$.

We write $U(\sigma_1,\sigma_2) = \{i : 1 \le i \le \tilde{r} \mid \sigma_1(i) \le k_1, \sigma_2(i) \le k_1\}$ and $u(\sigma_1,\sigma_2)$ for its size, and note when $u(\sigma_1,\sigma_2) = u$ there are at least $(2v-uk_0)$ independent conditions for $\mathcal{L}$ and at least $(2k_1-u)$ independent conditions for $\mathcal{K}$. For $(\sigma_1,\sigma_2)$ such that $u(\sigma_1,\sigma_2) = u$, we thus have the bound

$$\#S(\sigma_1,\sigma_2) \le 2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}/2^{2v-uk_0}\xi_{\mathcal{P}}^{2k_1-u}.$$

We then bound the number $A(u)$ of pairs of permutations with $u(\sigma_1,\sigma_2)=u$. There are $\binom{\tilde r}{u}$ subsets of $1\le i\le \tilde r$ of size $u$. Fixing such a subset $V$, we then bound the number of $\sigma_1$ that have $U(\sigma_1,\sigma_2)=V$ for some $\sigma_2$. There are $\binom{k_1}{u}$ ways to choose the image set of $V$ under $\sigma_1$, and $u!$ ways to permute the images. Meanwhile, the other $(\tilde r-u)$ elements of the domain can be chosen to have arbitrary images outside of $\sigma_1(V)$, giving a factor of $(\tilde r-u)!$. Symmetrically, the same argument for $\sigma_2$ yields the same result. Summing over $V$, this gives a bound for $A(u)$ of

$$\le \binom{\tilde r}{u}\Big[\binom{k_1}{u}u!(\tilde r-u)!\Big]^2 = \frac{\tilde r!}{u!(\tilde r-u)!}\Big(\frac{k_1!}{(k_1-u)!}(\tilde r-u)!\Big)^2$$
$$= \tilde r!^2\frac{(\tilde r-u)!}{u!\tilde r!}\frac{k_1!^2}{(k_1-u)!^2} \le \tilde r!^2\frac{(\tilde r-u)!}{\tilde r!}\frac{k_1!^2}{(k_1-u)!^2} \le \tilde r!^2\Big(\frac{k_1^2}{\tilde r}\Big)^u,$$

where we used that $k_1^2\le \tilde r$ in the final step.

Summing over $u$ then gives the mean-square bound

$$\sum_{(\mathcal{K},\mathcal{L})}\#W_{\tilde d}(\mathcal{K},\mathcal{L})^2 \le 2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}\cdot\tilde r!^2\sum_{u=0}^{\infty}\Big(\frac{k_1^2}{\tilde r}\Big)^u\frac{2^{uk_0}\xi_{\mathcal{P}}^u}{2^{2v}\xi_{\mathcal{P}}^{2k_1}} = \tilde r!^2\frac{2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}}{2^{2v}\xi_{\mathcal{P}}^{2k_1}}\Big(1-\frac{2^{k_0}\xi_{\mathcal{P}}k_1^2}{\tilde r}\Big)^{-1}$$

so that with $\mu=(\tilde r!/\xi_{\mathcal{P}}^{k_1})/2^v$ and $\lambda=2^{k_0}\xi_{\mathcal{P}}k_1^2/\tilde r$ we have

$$\sum_{(\mathcal{K},\mathcal{L})}\Big[\frac{\tilde r!/\xi_{\mathcal{P}}^{k_1}}{2^{\binom{k_0}{2}+k_0(k_1-k_0)}}-\#W_{\tilde d}(\mathcal{K},\mathcal{L})\Big]^2 = \sum_{(\mathcal{K},\mathcal{L})}\big[\mu^2-2\mu\#W_{\tilde d}(\mathcal{K},\mathcal{L})+\#W_{\tilde d}(\mathcal{K},\mathcal{L})^2\big]$$
$$\le 2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}\cdot\frac{\tilde r!^2/\xi_{\mathcal{P}}^{2k_1}}{2^{2v}}\big[1-2+1/(1-\lambda)\big].$$

As we assume $\lambda\le 1/2$, the final bracketed term is $\le 2\lambda$, and the result follows. □

We then have the indicated Corollary coming from applying Cauchy's inequality.

**Corollary 6.2.3.** *Assume $2^{k_0+1}\xi_{\mathcal{P}}k_1^2\le\tilde r$. Then for any $\tilde d\in\hat T$ we have*

$$\sum_{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde r,\mathcal{P})}\Big|\frac{\tilde r!/\xi_{\mathcal{P}}^{k_1}}{2^v}-\#W_{\tilde d}(\mathcal{K},\mathcal{L})\Big| \le 2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}\cdot\frac{\tilde r!/\xi_{\mathcal{P}}^{k_1}}{2^v}\Big(\frac{2^{k_0+1}\xi_{\mathcal{P}}k_1^2}{\tilde r}\Big)^{1/2}.$$

6.3. We next show that pleasant boxes restricted by $(\mathcal{K},\mathcal{L})$ have the expected size when averaged over $\mathrm{Sym}_{\tilde r}$.

**Proposition 6.3.1.** *Let $\bar T$ be a $(\kappa_0,\eta_1,\eta_{\mathrm{s}})$-pleasant $(X,\eta_0,\mathcal{P})$-box and assume that $\eta_0/2>\eta_1>\eta_{\mathrm{s}}>0$. Then*

$$\sum_{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde r,\mathcal{P})}\Big|\frac{\tilde r!}{\xi_{\mathcal{P}}^{\tilde r}}\frac{\#T}{2^{\binom{\tilde r}{2}}}-\sum_{\sigma\in\mathrm{Sym}_{\tilde r}}\#T(\mathcal{K}^\sigma,\mathcal{L}^\sigma)\Big| \ll \tilde r!\#T\cdot\Big(\frac{2^{k_0+1}\xi_{\mathcal{P}}k_1^2}{\tilde r}\Big)^{1/2}.$$

This is Smith's Theorem 6.4, or Theorem 5.9 of [3]. We use the above estimate from our Proposition 5.4.1 on $\tilde T(\mathcal{K},\mathcal{L})_{k_0}^{k_1}$ as the fulcrum in a triangle inequality.

*Proof.* The quantity of interest here is bounded as

$$\le \frac{1}{2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}}\sum_{(\mathcal{K},\mathcal{L})}\Big|\tilde r!\#T-2^v\xi_{\mathcal{P}}^{k_1}\sum_{\sigma\in\mathrm{Sym}_{\tilde r}}\#\tilde T(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1}\Big|$$
$$+\frac{2^v\xi_{\mathcal{P}}^{k_1}}{2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}}\sum_{\sigma\in\mathrm{Sym}_{\tilde r}}\sum_{(\mathcal{K},\mathcal{L})}\Big|\#\tilde T(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1}-\frac{2^{\binom{\tilde r}{2}}\xi_{\mathcal{P}}^{\tilde r}}{2^v\xi_{\mathcal{P}}^{k_1}}\#T(\mathcal{K}^\sigma,\mathcal{L}^\sigma)\Big|.$$

For the first sum we use $\sum_\sigma \#\tilde{T}(\mathcal{K}^\sigma,\mathcal{L}^\sigma)_{k_0}^{k_1} = \sum_{\tilde{d}} \#W_{\tilde{d}}(\mathcal{K},\mathcal{L})$ and get a bound of

$$\frac{1/\xi_{\mathcal{P}}^{\tilde{r}}}{2^{\binom{\tilde{r}}{2}}} \sum_{(\mathcal{K},\mathcal{L})} \Big|\tilde{r}!\#T - 2^v\xi_{\mathcal{P}}^{k_1} \sum_{\tilde{d}\in\hat{T}} \#W_{\tilde{d}}(\mathcal{K},\mathcal{L})\Big| \le \frac{1/\xi_{\mathcal{P}}^{\tilde{r}}}{2^{\binom{\tilde{r}}{2}}} \sum_{\tilde{d}\in\hat{T}} \sum_{(\mathcal{K},\mathcal{L})} \Big|\tilde{r}! - 2^v\xi_{\mathcal{P}}^{k_1}\#W_{\tilde{d}}(\mathcal{K},\mathcal{L})\Big|$$
$$\le \frac{1/\xi_{\mathcal{P}}^{\tilde{r}}}{2^{\binom{\tilde{r}}{2}}} \sum_{\tilde{d}\in\hat{T}} 2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}\cdot\tilde{r}!\Big(\frac{2^{k_0+1}\xi_{\mathcal{P}}k_1^2}{\tilde{r}}\Big)^{1/2} = \#T\cdot\tilde{r}!\Big(\frac{2^{k_0+1}\xi_{\mathcal{P}}k_1^2}{\tilde{r}}\Big)^{1/2}$$

where between the lines we used the previous Corollary 6.2.3.

The second term is bounded by Corollary 5.4.4 as

$$\ll \frac{2^v\xi_{\mathcal{P}}^{k_1}}{2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}} \sum_{\sigma\in\mathrm{Sym}_{\tilde{r}}} \sum_{(\mathcal{K},\mathcal{L})} \Big(\#T(\mathcal{K})\xi_{\mathcal{P}}^{\tilde{r}-k_1}\cdot\frac{\exp\big(6(\log\log X)^{2\eta_1}\big)}{\exp\big((\log\log X)^{\eta_0}/9\big)}\Big) \ll \frac{\tilde{r}!\#T\cdot 2^v}{\exp\big(c(\log\log X)^{\eta_0}\big)}$$

and as $v\log 2 \ll k_0k_1 \ll (\log\log\log X)(\log\log X)^{\eta_1}$ is dominated by $c(\log\log X)^{\eta_0}$ asymptotically, the $2^v$ can be ignored. Also $\exp\big(c(\log\log X)^{\eta_0}\big) \gg \log\log X \gg \tilde{r}^{1/2}$, implying this error is smaller than the previous, giving the Proposition. □

6.4. Using the above Proposition we can then compute (cf. above with §6.1.1)

$$\tilde{r}! \sum_{\substack{(\mathcal{K},\mathcal{L})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) = \sum_{\substack{(\mathcal{K},\mathcal{L})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \sum_{\sigma\in\mathrm{Sym}_{\tilde{r}}} \#T(\mathcal{K}^\sigma,\mathcal{L}^\sigma) = \sum_{\substack{(\mathcal{K},\mathcal{L})\\ s_{\tilde{r}}^{\varepsilon}(\mathcal{K},\mathcal{L})=s}} \tilde{r}!\frac{\#T}{2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}} + O\Big(\frac{\tilde{r}!\#T\cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big)$$

By the analysis given in the next section we shall find that (as $\tilde{r}\to\infty$) the number of $(\varepsilon,\mathcal{K},\mathcal{L})$ with 2-Selmer rank $(s+2)$ is $\sim \rho_s\cdot 2^{\binom{\tilde{r}}{2}}\xi_{\mathcal{P}}^{\tilde{r}}$ for some constant $\rho_s$, and will thus we will be able to conclude that for a pleasant box $\bar{T}$ we have

$$\#\{d\in\hat{T}^{\pm} : s(E_d) = s+2\} = \sum_{\substack{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})\\ s_{\tilde{r}}^{+}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) + \sum_{\substack{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P})\\ s_{\tilde{r}}^{-}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) \sim \rho_s\cdot\#T.$$

Summing over all such pleasant boxes will then give the main Theorem 1.1.2.

## 7. The 2-Selmer group, in brief

We return to the situation we introduced in §1.1, namely considering an elliptic curve $E : y^2 = (x-c_1)(x-c_2)(x-c_3)$ with $\Omega$ its set of bad primes,[26] and $\tilde{\Omega}$ when appending the infinite place. We can assume the $c_i$ are integral with no common nontrivial square factor. We consider twists $E_d : y^2 = (x-dc_1)(x-dc_2)(x-dc_3)$ of $E$ by squarefree $d$ that are coprime to $\Omega$. We write $\delta_{ij} = (c_i - c_j)$, and the bad primes will be those that divide at least one of the $\delta_{ij}$ (with $i\ne j$). We let $M_{\mathcal{P}}$ be 4 times the product of the primes in $\Omega$ (which necessarily contains 2), and take $\mathcal{R}_{\mathcal{P}}$ to contain all the coprime residue classes, so that $\alpha_{\mathcal{P}} = 1$.

We will be somewhat sketchy about the 2-Selmer group and genericity here, postponing more details to the next section. Our exposition will mostly follow Swinnerton-Dyer [50, §3] (see also his [49] for some extra details).

[26]Note that the given model is always bad at 2 (and thus $2\in\Omega$), though a minimal model for $E$ need not be: *e. g.*, the elliptic curve of conductor 15 with $(c_1,c_2,c_3) = (-9,0,16)$.

7.1. For the 2-Selmer group, we are interested in finding everywhere locally soluble quadric intersections $m_i y_i^2 = x - dc_i$ where $m_1 m_2 m_3$ is a nonzero square, denoting such a curve by $\mathcal{C}(\vec{m})$. As elements of $\mathbf{Q}^\star/(\mathbf{Q}^\star)^2$ the triples $\vec{m} = (m_1, m_2, m_3)$ form an abelian group as $\vec{m} \times \vec{m}' = (m_1 m_1', m_2 m_2', m_3 m_3')$ under component-wise multiplication, and we can restrict attention to $\vec{m}$ where the $m_i$ are units at all primes outside $\tilde{\Omega}$ that do not divide $d$.

7.1.1. Thus we can re-interpret the situation locally. Following Swinnerton-Dyer's notation we write $\mathcal{B}$ for the union of $\tilde{\Omega}$ with the primes dividing $d$.

We let $Y_l = \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$ for $l \in \mathcal{B}$, which is naturally a vector space over $\mathbf{F}_2$, being of dimension 2 for $l \geq 3$ (and dimension 3 for $l = 2$ and 1 for $l = \infty$). We let $V_l$ be the space of 3-tuples $(m_1, m_2, m_3) \in Y_l^3$ with $m_1 m_2 m_3 = 1$; this is again naturally a vector space over $\mathbf{F}_2$, of twice the dimension of $Y_l$. We then let $V_{\mathcal{B}} = \sum_{l \in \mathcal{B}} V_l$.

We then define $U_{\mathcal{B}}$ as the subspace of $V_{\mathcal{B}}$ generated by the diagonally embedded elements $(1, l, l)$ and $(l, l, 1)$ for all $l \in \mathcal{B}$ (with $l = -1$ corresponding to the infinite place). Meanwhile, we define $W_l \subset V_l$ as the image generated by the points on $E_d(\mathbf{Q}_l)$ under the Kummer map, defined away from 2-torsion points as $(X, Y) \to (X - dc_1, X - dc_2, X - dc_3)$, and for 2-torsion points by continuity. With $W_{\mathcal{B}} = \sum_{l \in \mathcal{B}} W_l$, the 2-Selmer group is then the intersection of $U_{\mathcal{B}}$ and $W_{\mathcal{B}}$ (both of these vector spaces have half the dimension of $V_{\mathcal{B}}$).

There is also Tate's pairing-based interpretation. We define $e_l$ on $V_l \times V_l$ by

$$e_l\big((m_1, m_2, m_3) \times (m_1', m_2', m_3')\big) = (m_1, m_1')_l + (m_2, m_2')_l + (m_3, m_3')_l$$

where $(u, v)_l$ is the Hilbert symbol (with values in $\mathbf{F}_2$) defined as 0 if $z^2 = ux^2 + vy^2$ has a nontrivial solution $(x, y, z) \in \mathbf{Q}_l^3$ and 1 otherwise. We extend $\oplus_l e_l$ to $V_{\mathcal{B}} \times V_{\mathcal{B}}$ by additivity. Our desired matrix is then $e_{\mathcal{B}}$ on $U_{\mathcal{B}} \times W_{\mathcal{B}}$ (or more precisely on bases therein), and the dimension of the kernel of this matrix is the 2-Selmer rank.

In particular, the Hilbert symbols in question are determined (with an obvious identification of $\pm 1$ with $\mathbf{F}_2$) by: the Legendre symbols $(p_i|p_j)$ for primes $p_i$ and $p_j$ that divide $d$ (so associated to $\mathcal{L}$), by $(q|p_j)$ for $q \in \tilde{\Omega}$, and by the sign of $d$. The conditions for $(q|p_j)$ are weaker than our congruential $\mathcal{K}$-conditions, at least assuming 8 divides $M_{\mathcal{P}}$ as we have ensured. We refer to these weaker Legendre conditions for bad primes as $\tilde{\mathcal{K}}$-conditions.

Thus the entries of the pairing matrix $\mathbf{M}(E_d)$ (though not the underlying interpretation in the terms of 2-covers) are the same for all $d \in S_{\vec{r}}^{\mathcal{P}}(\infty)$ that meet given $(\tilde{\mathcal{K}}, \mathcal{L})$-specifications. It will be notationally convenient to place the sign-condition on $d$ as a subscript on $\tilde{\mathcal{K}}$. In particular, all $d$ as above have the same 2-Selmer rank that we denote by $s(\tilde{\mathcal{K}}_+, \mathcal{L})$. Similarly all $-d \in S_{\vec{r}}^{\mathcal{P}}(\infty)$ with $|d|$ satisfying the $(\tilde{\mathcal{K}}, \mathcal{L})$-specifications have the same 2-Selmer rank, denoted by $s(\tilde{\mathcal{K}}_-, \mathcal{L})$.

7.1.2. There is still the issue of choosing a decent basis for $U_{\mathcal{B}}$ and the $W_l$ so that this pairing has nice properties, such as being (skew-)symmetric or even alternating. As Swinnerton-Dyer explains, one must identify $U_{\mathcal{B}}$ with $W_{\mathcal{B}}$ via an isomorphism, and for this can use an abstract theory of maximal isotropic subspaces to ameliorate the situation in general. For $p|d$ we can be more concrete, noting that $W_p$ will have have three nontrivial elements (from the 2-torsion under the Kummer map)[27]

$$\vec{w}_p^1 = (\delta_{12}\delta_{13}, d\delta_{12}, d\delta_{13}), \ \ \vec{w}_p^2 = (d\delta_{21}, \delta_{21}\delta_{23}, d\delta_{23}), \ \text{ and } \ \vec{w}_p^3 = (d\delta_{31}, d\delta_{32}, \delta_{31}\delta_{32}).$$

[27]In our description here we take (following Swinnerton-Dyer [50, (12)]) the second and third as basis elements; but in the next section we use the first and second (which seems more natural).

We shall discuss the choice of isomorphism with $U_{\mathcal{B}}$ more in §8.3.

7.2. Additionally, one can "estimate" the 2-Selmer rank using only the places in $\tilde{\Omega}$ and the first $c$ primes dividing $d$ (for both local solubility and in defining suitable 2-covers). In terms of the pairing matrix, this corresponds to taking the submatrix of size $2(\#\Omega + c)$-by-$2(\#\Omega + c)$ associated to $\tilde{\Omega}$ and the first $c$ primes.[28]

This submatrix is then the same for all $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ that have the same $c$-restriction, which indeed have the same $c$th 2-Selmer rank estimation $s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$.

Now this "estimation" is merely taking the rank of a minor of the full matrix, and indeed for any solitary $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ the principal mathematical feature it has is that the $\tilde{r}$th-estimation is indeed equal to the 2-Selmer rank. However, as described in §7.2.2 below, when averaging over $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ one can relate the *distribution* of $s_c$ to that of $s_{c+1}$, at least in the generic case.

7.2.1. With regards to $U_{\mathcal{B}}$ restricted to the first $c$ primes dividing $d$, which we denote by $U_{\mathcal{B}}^c$, this is then generated by $(1, l, l)$ and $(l, 1, l)$ where $l$ runs over the places in $\tilde{\Omega}$ and the first $c$ primes dividing $d$.

On the other hand, there is some subtlety when interpreting the restriction of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ to the first $c$ primes. Namely, defining $\mathcal{L}_{ji}$ from $\mathcal{L}_{ij}$ by quadratic reciprocity (feasible since the $\tilde{\mathcal{K}}$-conditions fix each prime modulo 4), we then define $\mathcal{L}_{jj} = \prod_{i \neq j} \mathcal{L}_{ij}$. Considering the first $c$ conditions for $\mathcal{L}$ in the sense of the 2-Selmer group then means to require all the prescribed $\mathcal{L}$-conditions to be met for $\mathcal{L}_{ij}$ for $1 \le i, j \le c$. This is now $\binom{c}{2} + c$ conditions for $0 \le c \le \tilde{r} - 1$. (In other words, the diagonal entries depend on all the primes dividing $d$, and this information should be retained even when considering the restriction – it may be more trenchant to include $\varepsilon$ in the definition of $\mathcal{L}_{jj}$ somehow, but this is unneeded).

Similarly, we can define $\tilde{\mathcal{K}}_{q0} = \prod_j \tilde{\mathcal{K}}_{qj}$, and restricting the $\tilde{\mathcal{K}}_\varepsilon$-conditions to the first $c$ primes will mean considering $\varepsilon$ and all $\tilde{\mathcal{K}}_{qj}$ with $0 \le j \le c$. In particular, the 0th conditions are always present, and the ultimate prime is not independent information. (This is equivalent to fixing the class of $d$ in $\mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$ for all $l \in \tilde{\Omega}$).

Note that the 0th 2-Selmer estimation already depends on $d$. Indeed, we shall show below that all the $s_c$ have the same parity, and thus $s_0$ contains whether the twist has even or odd parity. The global root numbers for $E$ and $E_d$ differ[29] by a factor of $(d|-N_E)$ when twisting by fundamental discriminants $d$ coprime to the conductor $N_E$; if this transferred to 2-Selmer parities, our specification of $d$ in $\mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$ for all $l \in \tilde{\Omega}$ would be a stronger formulation therein. It turns out to be rather nontrivial to make this transference, though Monsky [36] combines computations of Kramer [29] with Kolyvagin's theorem [26] to do so. For our purposes we can avoid the latter, and in §13.2 we show the weaker result that exactly half of the $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ have $s_0$ with even parity.[30]

[28] One need not take the primes in order, but there is also no reason not to do so.

[29] This is typically noted in a context of a functional equation for the $L$-functions, thus using modularity, and then follows from the work of Atkin and Lehner [1] with modular forms. I am not sure of the history, but I think Kramer's work [29] is the first to delve into the issue for elliptic curves as such, though as noted, it seems more pertinent to then consider the Selmer parities rather than the global root number *per se*.

[30] Kane asserts this in [23, Lemma 1], though his claim that $E$ can be twisted to have good or multiplicative reduction everywhere is false (*e. g.* for the congruent number curve, of conductor $2^5$). This prevents Kramer's Corollary 1 (to Proposition 6) from being used in such situations. Also,

7.2.2. The purpose of the above estimation is that we can often relate $s_c$ to $s_{c+1}$, at least in a distributional sense. First let us note that $s_{\tilde{r}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is indeed the 2-Selmer rank, and it is always 2 more than $s_{\tilde{r}-1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, with this addition of 2 corresponding to the 2-torsion points (see §8.4.1).

When the $c$-restriction of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is suitably generic, there is a distributional relation between $s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ and $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ for $c < \tilde{r} - 1$.

The essential reason for this is a consideration of the appending of the information from the $(c+1)$st prime (call it $p$) to the pairing matrix, which is discussed by Swinnerton-Dyer in [50, §5]. We first consider the kernel of the pairing submatrix $\mathbf{M}_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, and the pairing restricted to said kernel is clearly 0. Upon appending the information for the $(c+1)$st prime, the essential part of the pairing matrix for the first $(c+1)$ primes can be written in the form

$$G = \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & C \end{pmatrix}$$

where $A$ is 2-by-$s$ $\big($with $s = s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})\big)$ and $C$ is 2-by-2 and alternating. Given the $c$-restriction $\mathbf{M}_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, there are thus $(2s+1)$ conditions that determine $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$.

With the above beneficial choice of basis for $W_p$ for $p$ dividing $d$, given an element $\vec{u} = (u_1, u_2, u_3) \in \mathcal{U}_{\mathcal{B}}^c$ we can compute that the pairing between $u$ and $\vec{w}_p^2$ is given by $(d\delta_{21}, u_1)_p + (\delta_{21}\delta_{23}, u_2)_p + (d\delta_{23}, u_3)_p$. Here the middle term is 0 as both components are units modulo $p$, and similarly the first term becomes $(p, u_1)_p$ and the third $(p, u_3)_p$. Since $u_1u_2u_3 = 1$, this is the same as $(p, u_2)_p$. Similarly, the pairing for $\vec{u}$ with $\vec{w}_p^3$ is $(p, u_3)_p$. This then gives the matrix $A$.

Meanwhile, a beneficial choice of $\vec{\alpha}_p^2$ and $\vec{\alpha}_p^3$ as basis vectors for $U_p$ ensures that $C$ is alternating (and that the lower part of the matrix is indeed $A^{\mathrm{T}}$), and with the off-diagonal entry depending on $(d, p)_p$ for $c < \tilde{r} - 1$; thus this off-diagonal entry is independent, in Swinnerton-Dyer's sense, of the entries in $A$ (which only depend on Legendre symbols involving $p$ and the first $c$ primes).

7.2.3. Swinnerton-Dyer gives genericity conditions that ensure that the entries of $A$ are suitably independent, by ensuring they involve differing primes in Legendre symbols. Moreover, the nondiagonal entry of $C$ depends on all primes dividing $d$, and thus is always independent (for $c \neq \tilde{r} - 1$) of the other considerations. Upon showing that non-genericity occurs for at most $O\big(2^{\binom{c+1}{2}} 2^{(c+1)\#\tilde{\Omega}} \cdot (15/16)^c\big)$ of the possible $(\mathcal{K}'_\varepsilon, \mathcal{L}')$-restrictions to the first $c$ primes, he then reduces the problem to the generic case. Thereupon we note the pairing matrix $G$ has rank 0 when $A = C = 0$, rank 4 when $A$ has rank 2, and rank 2 otherwise. In the first case the 2-Selmer rank estimate (which is the dimension of the kernel) increases by 2, in the second case it decreases by 2, and in the final case it remains constant. It is then an exercise to show that of the $2^{2s+1}$ choices for $(A, C)$, one of them induces $G$ of rank 0, and that $2 \cdot 3(2^s - 1)$ yield $G$ with rank 2 (the nonzero entries in $A$ either need to lie all in the same column, or be replicated the same in each column).

We sum this up as follows, postponing more details (mostly about genericity) to the next section. We let $\mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ be the set of all $2 \cdot 2^{\binom{\tilde{r}}{2}} 2^{\tilde{r}\#\tilde{\Omega}}$ choices of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, and $\mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ the set of $c$-restrictions of them.

---

given that Kane moves to such a minimal twist, this would seem to inhibit the replacement of $E$ by $E'$ (to handle twists by bad primes) as discussed in our Introduction (cf. Footnote 2).

**Lemma 7.2.4.** *Suppose that no element of $\{\delta_{12}\delta_{13}, \delta_{21}\delta_{23}, \delta_{31}\delta_{32}\}$ is square (which is equivalent to $E$ having no rational 4-torsion). For $c < \tilde{r} - 1$, the number of $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ with $\mathbf{M}_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$ nongeneric is $O\big(2^{\binom{c+1}{2}} 2^{(c+1)\#\tilde{\Omega}} \cdot (15/16)^c\big)$.*

**Lemma 7.2.5.** *Let $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ be generic and consider the subset of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ whose $c$-restriction is $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$. (Note here that genericity necessarily implies $c < \tilde{r} - 1$).*

*The $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ are then distributed from $s = s_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$ in the following manner:*

- *a proportion $1/2^{2s+1}$ of them have $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) = s + 2 = s_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') + 2$,*
- *a proportion $3/2^s - 5/2^{2s+1}$ have $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) = s = s_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$,*
- *a proportion $1 - 3/2^s + 2/4^s$ have $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) = s - 2 = s_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') - 2$.*

7.3. Finally we sketch the Markov chain analysis as given by Swinnerton-Dyer, again postponing more details (and indeed, a somewhat different formulation) to the next section.

7.3.1. Firstly we note the $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ that are non-generic at some $j \ge J$ are not very problematic. Indeed, the number of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ that are non-generic at $j$ is $\ll (15/16)^j 2^{\binom{\tilde{r}}{2}} 2^{\tilde{r}\#\tilde{\Omega}}$, and by Lemma 5.5.1 we have $\#T(\tilde{\mathcal{K}}, \mathcal{L}) \ll \#T(\tilde{\mathcal{K}}) 2^v / 2^{\binom{\tilde{r}}{2}}$ where $v \le k_0 k_1 \le \kappa_0 \eta_0 \log\log\log X \cdot 3(\log\log X)^{\eta_1}$. Summing over $j$ and $\tilde{\mathcal{K}}_\varepsilon$ then gives that the total number of $d \in \hat{T}$ that are in some $T(\tilde{\mathcal{K}}, \mathcal{L})$ that is nongeneric for some choice of $\varepsilon$ and some $j$ with $J \le j < \tilde{r} - 1$ is bounded as $\ll \#T \cdot 2^v (15/16)^J$. This is then adequately small with $J \sim (\log\log X)/99$ (say).

7.3.2. The essential idea of the Markov chain is to consider a starting distribution $\vec{h}^J$ whose $s$th component is the number (or proportion) of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\Omega)$ whose 2-Selmer $J$-estimation $s_J(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $s$. Ignoring genericity, this will then lead to a distribution $\vec{h}^j$ for $J \le j < \tilde{r} - 1$ with

$$h_s^{j+1} = h_{s-2}^j \cdot \frac{1}{2^{2s+1}} + h_s^j \cdot \Big(\frac{3}{2^s} - \frac{5}{2^{2s+1}}\Big) + h_{s+2}^j \cdot \Big(1 - \frac{3}{2^s} + \frac{4}{2^{2s+1}}\Big).$$

(One can, and we shall in the next section, interpret this in terms of matrices). The question is then what the distribution $\vec{h}^{\tilde{r}-1}$ looks like, with our expectation that it should be reasonably close to a linear combination of the two stable distributions (one for each parity of $s_J$). These are given in terms of

$$\rho_s = \frac{2^s}{\prod_{j=1}^s (2^j - 1)} \prod_{n=0}^{\infty} (1 - 1/2^{2n+1}),$$

for which we have $\rho_0 + \rho_2 + \rho_4 + \cdots = \rho_1 + \rho_3 + \rho_5 + \cdots = 1$ as expected from a probability distribution, and indeed for $s \ge 0$ (with $\rho_{-2} = \rho_{-1} = 0$) we have

$$\rho_s = \rho_{s-2} \cdot \frac{1}{2^{2(s-2)+1}} + \rho_s \cdot \Big(\frac{3}{2^s} - \frac{5}{2^{2s+1}}\Big) + \rho_{s+2} \cdot \Big(1 - \frac{3}{2^{s+2}} + \frac{4}{2^{2(s+2)+1}}\Big).$$

The main theorem of Markov chains[31] then implies that the distribution tends exponentially quickly (with rate determined by the second largest eigenvalue $\lambda_2$) to

[31]For those who have not seen the argument before, one first determines the distribution after $u$ steps when starting in the ground state (which is either 0 or 1 in our case). This gives that the probability of being in state $\mathcal{S}$ is $\rho_{\mathcal{S}}(u) = \alpha_{\mathcal{S}} + O(e^{-\kappa u})$ for some $\kappa > 0$. One then considers starting in state $\mathcal{I}$, and computes the probability $\beta(m)$ that the first time to reach the ground state will be at the $m$th step. The probability of being in state $S$ at the $v$th step when starting in state $\mathcal{I}$ is then $\sum_m \rho_{\mathcal{S}}(v-m)\beta(m)$. In our case, we reach the ground state for the first time almost surely

the stable distribution(s). As our walking length is $\tilde{r} - J \geq 97(\log\log X)/99$, this gives an error estimate of $O(\lambda_2^{\tilde{r}-J}) = O\big(1/(\log X)^c\big)$.

There are minor issues with including non-genericity into this analysis, which we regard more in the next section. We sum this up as follows, recalling that $s_{\tilde{r}}$ is 2 more than $s_{\tilde{r}-1}$, and that half of the $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ yield 2-Selmer rank of each parity.

**Lemma 7.3.3.** *Consider the set of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ that are $j$-generic for all $j$ with $J \leq j < \tilde{r}-1$, where $J = \lfloor(\log\log X)/99\rfloor$. Then the proportion of such $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ whose $s_{\tilde{r}}$-value is a given value $(s+2)$ is $\rho_s/2 + O(1/(\log X)^c)$ for some $c > 0$.*

This then allows us to complete the proof of Theorem 1.1.2 as put forth in §6.4. (We will recapitulate the entire proof structure in §9).

7.4. We now give an extended example, to try to illuminate the notion of genericity. This example is a bit contrived, as in practice the number of primes dividing $d$ tends to infinity (thus much exceeds $\#\tilde{\Omega}$), while we only have 5 primes dividing $d$.

7.4.1. We consider the congruent number curve $E : y^2 = x^3 - x$, so $\tilde{\Omega} = \{2, \infty\}$. We will consider twisting $E$ by a product $d = \varepsilon p_1 p_2 p_3 p_4 p_5$ of a sign and five primes.

First we consider the 0th estimation which (as in §7.2.1), specifies the sign of $d$ and its class in $\mathbf{Q}_2^\star/(\mathbf{Q}_2^\star)^2$ (indeed, the $\mathcal{K}$-condition for $\infty$ specifies $(-1|d)$ and the $\mathcal{K}$-condition at 2 specifies $(2|d)$, while $\varepsilon$ gives the sign). In our example we will consider $d \equiv 1\,(8)$ with $d$ positive.

We have $(c_1, c_2, c_3) = (-1, 0, 1)$ so $(\delta_{12}, \delta_{13}, \delta_{23}) = (-1, -2, -1)$. We immediately see that the 0th estimation is in fact the $\tilde{r}$th estimation for $E$ itself (with $d = 1$), for which the 2-Selmer rank is 2, with the elements given by $(1, 1, 1)$, $(2, -1, -2)$, $(1, -1, -1)$, and $(2, 1, 2)$, corresponding to global torsion elements for $E_1$.

In fact, let us derive this a bit more concretely, taking the opportunity to note a subtlety with $W_2$. We have a 4-by-4 pairing matrix corresponding to bases for $U^0_{\mathcal{B}}$, and $W_l$ for $l \in \{\infty, 2\}$. The former is generated by the elements $(1, 2, 2)$, $(2, 1, 2)$, $(1, -1, -1)$ and $(-1, 1, -1)$, while $W_\infty$ is generated by $(1, -1, -1)$. Now, two independent elements of $W_2$ can be determined by 2-torsion points, namely $(2, -d, -2d)$ and $(d, -1, -d)$, associated respectively to the torsion points $X = -d$ and $X = 0$, under the Kummer map $(X, Y) \to (X - dc_1, X - dc_2, X - dc_3) = (X + d, X, X - d)$ extended by continuity. A third independent 2-adic point can be seen to come from $X = 28$, as then $X^3 - d^2X = X(X^2 - d^2)$ has both $X$ and $X^2 - d^2 \equiv X^2 - 1\,(8)$ in the 7 mod 8 square class, so that their product is square. The Kummer image $(28 + d, 28, 28 - d)$ is then 2-adically equivalent to $(5, 7, 3)$ for our $d \equiv 1\,(8)$. We can then compute the pairing matrix on such bases as in Table 1, where in the computations we repeatedly used multiplicativity of Hilbert symbols combined with the 3 facts: $(2, 2)_2 = 0$, $(2, u)_2 = 0$ for $u \equiv 1, 7\,(8)$ while $(2, u) = 1$ for $u \equiv 3, 5\,(8)$, and $(u, v)_2 = 0$ for odd $u, v$ unless $u \equiv v \equiv 3\,(4)$. For instance, we have the calculation $\langle(-1, 1, -1), (2, -d, -2d)\rangle_2 = (-1, 2)_2 + (1, -d)_2 + (-1, -2d)_2 = (-1, -d)_2 = 1$ since $d \equiv 1\,(8)$. As seen, the kernel is indeed generated by $(2, 1, 2)$ and $(1, -1, -1)$. (We later make the pairing matrix be symmetric by modifying the bases).

We give our definition of genericity for the $c$th estimation in §8.4.3 below, and for $c \neq \tilde{r} - 1$ it is that the kernel $U^c_{\mathcal{B}}(\mathcal{K}_\varepsilon, \mathcal{L})$ of the $c$th pairing matrix $\mathbf{M}_c(\mathcal{K}_\varepsilon, \mathcal{L})$

---

within a bit more than $J$ steps, say $J + \sqrt{J}(\log J)^2$, with then nearly $\tilde{r} - 2J \geq 97(\log\log X)/99$ steps remaining. This then gives the probability of being in state $\mathcal{S}$ as $\rho_{\mathcal{S}} + O(1/(\log X)^c)$ for some $c > 0$, and this is suitably uniform over starting states $\mathcal{I}$ by taking $J$ appropriately small.

| | $(1,-1,-1)_\infty$ | $(5,7,3)_2$ | $(2,-d,-2d)_2$ | $(d,-1,-d)_2$ |
|---|---|---|---|---|
| $(1,2,2)$ | 0 | $(2,21)_2=1$ | $(2,2)_2=0$ | $(2,d)_2=0$ |
| $(2,1,2)$ | 0 | $(2,15)_2=0$ | $(2,-d)_2=0$ | $(2,-1)_2=0$ |
| $(1,-1,-1)$ | 0 | $(-1,21)_2=0$ | $(-1,2)_2=0$ | $(-1,d)_2=0$ |
| $(-1,1,-1)$ | 1 | $(-1,15)_2=1$ | $(-1,-d)_2=1$ | $(-1,-1)_2=1$ |

TABLE 1. Initial pairing matrix for the congruent number curve

should not have have any element of the form $(1,u,u)$ with $u\neq 1$, and should also not have any two elements $(u_1,u_2,u_3)$ and $(v_1,v_2,v_3)$ with $u_1=v_2$. From this we find, somewhat trivially, that the 0-restriction of every $(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ with $d\equiv 1\,(8)$ and $d$ positive is nongeneric. However, this doesn't stop us from carrying out the appending process in §7.2.2 – it simply means that the proportions given in Lemma 7.2.5 need not apply.

7.4.2. We proceed from the 0-restrictions to 1-restrictions by appending $p_1$. This introduces only the information from $\mathcal{L}_{11}=(d/\varepsilon p_1|p_1)$ for Legendre conditions, while the $\mathcal{K}$-conditions again specify $p_1$ modulo 8. Thus we find there are 8 different 1-restrictions, specified by $(p_1 \bmod 8,\mathcal{L}_{11})\in(\mathbf{Z}/8\mathbf{Z})^\star\times\{\pm1\}$.

The description in §8.3 gives a way of selecting local images of basis elements for $U_{\mathcal{B}}^{c+1}/U_{\mathcal{B}}^c$, (this being valid for $c<\tilde{r}-1$). We describe the result in Table 2. Here $\delta_{12}^{13}=\delta_{12}\delta_{13}$ is a typographical shorthand, as is $\delta_{21}^{23}$ for $\delta_{21}\delta_{23}$, while we write $p=p_1$, and $\nu_p$ for a quadratic non-residue modulo $p$. In reading the left half of the table, the rows correspond to a condition on $(\delta_{12}\delta_{13}|p)$ and the columns to one on $(\delta_{12}|p)$, with the entry in the table then giving the applicable local image.

| $(\delta_{12}^{13}\|p)$ | $(\delta_{12}\|p)=+1$ | $(\delta_{12}\|p)=-1$ |
|---|---|---|
| $+1$ | $(1,p,p)$ | $(1,p\nu_p,p\nu_p)$ |
| $-1$ | $(\nu_p,p,p\nu_p)$ | $(\nu_p,p\nu_p,p)$ |

| $(\delta_{21}^{23}\|p)$ | $(\delta_{21}\|p)=+1$ | $(\delta_{21}\|p)=-1$ |
|---|---|---|
| $+1$ | $(p,1,p)$ | $(p\nu_p,1,p\nu_p)$ |
| $-1$ | $(p,\nu_p,p\nu_p)$ | $(p\nu_p,\nu_p,p)$ |

TABLE 2. Local images for basis elements for $p$ in the pairing matrix

For the congruent number curve $\delta_{12}\delta_{13}=2$ while $\delta_{21}\delta_{23}=-1$, and $\delta_{12}=1$ so $\delta_{21}=-1$ (in particular the $(\delta_{12}|p)=(1|p)=-1$ column is irrelevant). The appended local images of the basis elements with the matrix $G$ of §7.2.2 are: $(1,p,p)$ when $(2|p)=+1$ and $(\nu_p,p,p\nu_p)$ when $(2|p)=-1$; and $(p,1,p)$ when $(-1|p)=+1$ and $(p\nu_p,\nu_p,p)$ when $(-1|p)=-1$. We can take $\nu_p=-1$ when $(-1|p)=-1$ and $\nu_p=2$ when $(2|p)=-1$ (and either works for $p\equiv 3\,(8)$). For such a local image $s$ there is $\tilde{s}\in U_{\mathcal{B}}^{c+1}$ that maps to it such that the pairing of $\tilde{s}$ with respect to the $W_l$ for $l\in\tilde{\Omega}$ and to the $W_p$ for the first $c$ primes is always 0 (this $\tilde{s}$ is unique, up to translates by elements of $U_{\mathcal{B}}^c(\mathcal{K}_\varepsilon,\mathcal{L})$, and possibly the other appended element).

We summarize the $G$-matrices in the various cases in Table 3. Therein we have written $l=(d,p_1)_{p_1}=(d/\varepsilon p_1,p_1)_{p_1}+(\varepsilon p_1,p_1)_{p_1}$, where the first term corresponds to $\mathcal{L}_{11}$, and the second is $(p_1,p_1)_{p_1}=(-1,p_1)_{p_1}$. Note that the upper-right 2-by-2 corner $A$ has rank 2 for $p\equiv 3\,(8)$, and thus $U_{\mathcal{B}}^1(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ will be trivial in this case, as the kernel of $G$ is trivial and the first 2-Selmer estimation is 0. Meanwhile $A$ has rank 1 for $p\equiv 5,7\,(8)$, and here $G$ has a kernel of dimension 2, giving the first 2-Selmer estimation. Finally, when $p\equiv 1\,(8)$ we see $A=0$, and thus whether the first 2-Selmer estimation is 4 or 2 depends on whether $l=0$.

The computation of $A$ is similar to the recipe given in §7.2.2 (though using the first and second elements as a basis rather than the second and third as Swinnerton-Dyer did), namely for a row $(u_1,u_2,u_3)$ the left column corresponds to $(u_1,p)_p$ and the right column to $(u_2,p)_p$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p \equiv 1\,(8)$ | $(1,-1,-1)$<br>$(2,1,2)$<br>$(1,p,p)$<br>$(p,1,p)$ | $\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&l\\0&0&l&0\end{pmatrix}$ | $(1,-1,-1)$<br>$(2,1,2)$<br>$(\nu_p,p,p\nu_p)$<br>$(p\nu_p,\nu_p,p)$ | $\begin{pmatrix}0&0&0&1\\0&0&1&0\\0&1&0&l\\1&0&l&0\end{pmatrix}$ | $p \equiv 3\,(8)$ |
| $p \equiv 5\,(8)$ | $(1,-1,-1)$<br>$(2,1,2)$<br>$(\nu_p,p,p\nu_p)$<br>$(p,1,p)$ | $\begin{pmatrix}0&0&0&0\\0&0&1&0\\0&1&0&l\\0&0&l&0\end{pmatrix}$ | $(1,-1,-1)$<br>$(2,1,2)$<br>$(1,p,p)$<br>$(p\nu_p,\nu_p,p)$ | $\begin{pmatrix}0&0&0&1\\0&0&0&0\\0&0&0&l\\1&0&l&0\end{pmatrix}$ | $p \equiv 7\,(8)$ |

TABLE 3. Basis elements (local images) and matrices $G$ in various cases

The final accounting is that 2 of the 8 selections of $(p_1 \bmod 8, \mathcal{L}_{11})$ give a first 2-Selmer estimation of 0, while 5 of them give 2, and 1 gives 4. Indeed, instead of obtaining five independent conditions from the entries of $A$ and the off-diagonal entry of $C$, we find that the upper-left and lower-right entries of $A$ are forced to be 0, as they correspond to $(1,p)_p$. Thus we have three independent conditions, with indeed $2^3$ possibilities. The estimation increases (to 4) only when $A = C = 0$

For our running example we will take $p_1 \equiv 5\,(8)$ with $l = 1$ (which is $\mathcal{L}_{11} = -1$). Then a basis of $U^1_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is given by $(1,-1,-1)$ and $(2p,1,2p)$. Indeed, the latter is not only the local image at $p$, as we readily compute that

$$\langle (2p,1,2p),(5,7,3)\rangle_2 = (2p,15)_2 = (2,15)_2 + (p,15)_2 = 0+0,$$
$$\langle (2p,1,2p),(2,-d,-2d)\rangle_2 = (2p,-d)_2 = (2,-d)_2 + (p,-d)_2 = 0+0,$$
$$\text{and } \langle (2p,1,2p),(d,-1,-d)\rangle_2 = (2p,-1)_2 = (2,-1)_2 + (p,-1)_2 = 0+0.$$

7.4.3. In passing from a 1-restriction to a 2-restriction, we will have 2 new Legendre conditions and again a congruence condition mod 8. Thus each 1-restriction will yield 16 different 2-restrictions, via $(p_2 \bmod 8, \mathcal{L}_{12}, \mathcal{L}_{22}) \in (\mathbf{Z}/8\mathbf{Z})^\star \times \{\pm 1\} \times \{\pm 1\}$.

In some cases there are simplifications. For instance for $p_1 \equiv 3\,(8)$, whether or not the second 2-Selmer estimation is 0 or 2 depends solely on $\mathcal{L}_{22}$, so is independent of $p_2\,(8)$ and $\mathcal{L}_{12}$ (indeed, this case is generic, and we get the expected split).

Continuing our example with $p_1 \equiv 5\,(8)$ and $l = 1$, we can then produce the matrices $G$ seen in Table 4, where we wrote $p = p_2$. Here we wrote $e = (p_1,p)_p$ and $\bar{e} = e+1$; however we need not simply have $l = (d,p)_p$, as this is only valid for the local image – in general $l$ *depends* on $(d,p)_p$ but could also involve Legendre symbols with other primes and $p$. In any event, specifiying $e$ is equivalent to specifying $\mathcal{L}_{12}$, and similarly with $l$ and $\mathcal{L}_{22}$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p \equiv 1\,(8)$ | $(1,-1,-1)$<br>$(2p_1,1,2p_1)$<br>$(1,p,p)$<br>$(p,1,p)$ | $\begin{pmatrix}0&0&0&0\\0&0&e&0\\0&e&0&l\\0&0&l&0\end{pmatrix}$ | $(1,-1,-1)$<br>$(2p_1,1,2p_1)$<br>$(\nu_p,p,p\nu_p)$<br>$(p\nu_p,\nu_p,p)$ | $\begin{pmatrix}0&0&0&1\\0&0&\bar{e}&0\\0&\bar{e}&0&l\\1&0&l&0\end{pmatrix}$ | $p \equiv 3\,(8)$ |
| $p \equiv 5\,(8)$ | $(1,-1,-1)$<br>$(2p_1,1,2p_1)$<br>$(\nu_p,p,p\nu_p)$<br>$(p,1,p)$ | $\begin{pmatrix}0&0&0&0\\0&0&\bar{e}&0\\0&\bar{e}&0&l\\0&0&l&0\end{pmatrix}$ | $(1,-1,-1)$<br>$(2p_1,1,2p_1)$<br>$(1,p,p)$<br>$(p\nu_p,\nu_p,p)$ | $\begin{pmatrix}0&0&0&1\\0&0&e&0\\0&e&0&l\\1&0&l&0\end{pmatrix}$ | $p \equiv 7\,(8)$ |

TABLE 4. Matrices $G$ in various cases for second estimation

For our running example we take $p_2 \equiv 7\,(8)$ and $e = 1$ (so $\mathcal{L}_{12} = (p_1|p_2) = -1$), which gives us a 2-estimation of 0. For the purposes of this 2-estimation it doesn't matter what $\mathcal{L}_{22}$ is, though when considering later basis elements it will.

7.4.4. Upon passing from a 2-restriction to a 3-restriction we see there are three new Legendre conditions and again a congruence condition mod 8. Thus each 2-restriction corresponds to 32 different 3-restrictions.

In our running example, since we had a 2-estimation of 0, the only consideration is whether $(d,p_3)_{p_3}$ is 0 or not, as this determines the off-diagonal entry of $C$ (though again we stress that, following Swinnerton-Dyer's setup, this determination involves lifting a local image). When $C=0$ the 3-estimation is 2, and else $C$ has trivial kernel and the 3-estimation is 0. We take $p_3 \equiv 3\,(8)$, with $(d,p_3)_{p_3}$ so that $C=0$.

It is now a chore to calculate the elements in $U^3_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ corresponding to the local images. It is probably easier, and fitting more with our pedagogical aims in this example, to return to the full 10-by-10 pairing matrix and work from there.

In Table 5 we give the relevant matrix, writing $\bar{x}=x+1$, with $l_{ij}=(p_i,p_j)_{p_j}$ for $i\neq j$ and $l_{jj}=(p_j,d)_{p_j}$.

| | $W_\infty$ | $\tilde{W}_2^0$ | $\tilde{W}_2^1$ | $\tilde{W}_2^2$ | $W_{p_1}^1$ | $W_{p_1}^2$ | $W_{p_2}^1$ | $W_{p_2}^2$ | $W_{p_3}^1$ | $W_{p_3}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $(1,2,2)$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| $(-1,1,-1)$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| $(1,-1,-1)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| $(2,1,2)$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| $(1,p_1,p_1)$ | 0 | 0 | 0 | 1 | 1 | $l_{11}$ | 0 | $l_{12}$ | 0 | $l_{13}$ |
| $(p_1,1,p_1)$ | 0 | 0 | 0 | 0 | $l_{11}$ | 0 | $l_{12}$ | 0 | $l_{13}$ | 0 |
| $(1,p_2,p_2)$ | 0 | 0 | 0 | 0 | 0 | $l_{12}$ | 0 | $l_{22}$ | 0 | $l_{23}$ |
| $(p_2,1,p_2)$ | 0 | 1 | 1 | 0 | $l_{12}$ | 0 | $\bar{l}_{22}$ | 1 | $l_{23}$ | 0 |
| $(1,p_3,p_3)$ | 0 | 0 | 0 | 1 | 0 | $l_{13}$ | 0 | $\bar{l}_{23}$ | 1 | $l_{33}$ |
| $(p_3,1,p_3)$ | 0 | 1 | 1 | 0 | $l_{13}$ | 0 | $\bar{l}_{23}$ | 0 | $\bar{l}_{33}$ | 1 |

TABLE 5. Pairing matrix after appending $(p_1,p_2,p_3)\equiv(5,7,3)$ mod 8

Note that we have symmetrised the initial pairing matrix (seen in Table 1), which is the 4-by-4 upper-left corner, by swapping the second and fourth rows, and adding the first column to the other columns. Thus the first column label $W_\infty$ here is the pairing with $(-1,-1,-1)$ at $\infty$, the column label $\tilde{W}_2^0$ is $W_\infty+W_2^0$ where the latter is the pairing with $(5,7,3)$ at 2, the column label $\tilde{W}_2^1$ is $W_\infty+W_2^1$ where the latter is with $T_1=(2,-d,-2d)$ at 2, and the column label $\tilde{W}_2^2$ is $W_\infty+W_2^2$ where the latter is with $T_2=(d,-1,-d)$ at 2. In general $W_p^1$ is the pairing with $T_1$ at $p$ and $W_p^2$ is the pairing with $T_2$ at $p$.

Each 2-by-2 block in the matrix has a specific pattern. For instance, the diagonal 2-by-2 blocks in the lower-right are determined by the congruence class of $p$ modulo 8; the given array displays 3 of the 4 possibilities therein. These are readily calculated as $\langle(2,-d,-2d),(1,p,p)\rangle_p=(2,p)_p$ and $\langle(d,-1,-d),(p,1,p)\rangle_p=(-1,p)_p$ in the diagonal entries, while $\langle(2,-d,-2d),(p,1,p)\rangle_p=(-d,p)_p=(-1,p)_p+(d,p)_p$ and $\langle(d,-1,-d),(1,p,p)\rangle_p=(d,p)_p$. The off-diagonal blocks in the lower-right section are computed in a similar manner, as $\langle(2,-d,-2d),(1,p_i,p_i)\rangle_{p_j}=(2,p_i)_{p_j}=0$ while $\langle(2,-d,-2d),(p_i,1,p_i)\rangle_{p_j}=(-d,p_i)_{p_j}=(p_j,p_i)_{p_j}$, and the behaviours switch for $T_2=(d,-1,-d)$.

The entries in the 4-by-6 upper-right block are determined by the congruence classes mod 8 for $(p_1,p_2,p_3)$, as similarly are those in the 6-by-4 lower-left block. We summarise the computations therein in Table 6.

| | $W_p^1$ | $W_p^2$ |
|---|---|---|
| $(1,2,2)$ | $0$ | $(2,p)_p$ |
| $(-1,1,-1)$ | $(-1,p)_p$ | $0$ |
| $(1,-1,-1)$ | $0$ | $(-1,p)_p$ |
| $(2,1,2)$ | $(2,p)_p$ | $0$ |

| | $W_\infty$ | $\tilde{W}_2^0$ | $\tilde{W}_2^1$ | $\tilde{W}_2^2$ |
|---|---|---|---|---|
| $(1,p,p)$ | $0$ | $0$ | $0$ | $(2,p)_2$ |
| $(p,1,p)$ | $0$ | $(-1,p)_2$ | $(-1,p)_2$ | $0$ |

TABLE 6. Pairing computations

To achieve symmetry, and indeed make the whole matrix be alternating, we can replace the last 6 rows $(\vec{r}_5,\dots,\vec{r}_{10})$ by

$$(\vec{r}_5+\vec{r}_4,\vec{r}_6+\vec{r}_2+\vec{r}_3,\vec{r}_7+\vec{r}_1+\vec{r}_3,\vec{r}_8+\vec{r}_1+\vec{r}_3,\vec{r}_9+\vec{r}_1+\vec{r}_3+\vec{r}_4,\vec{r}_{10}+\vec{r}_1+\vec{r}_2)$$

In any case, we want to determine a basis for the 2-dimensional kernel; so far we have specified $(l_{11},l_{12})=(1,1)$ and that $l_{33}$ yield a 2-dimensional kernel – for our example we take $(l_{13},l_{22},l_{23},l_{33})=(1,0,0,1)$, and the nontrivial kernel elements are $(p_2p_3,-1,-p_2p_3)$, $(2,p_1p_3,2p_1p_3)$, and $(2p_2p_3,-p_1p_3,-2p_1p_2)$.

What is then important for us is that $U^3_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ is now generic, given the definition[32] we cited in the last paragraph of §7.4.1. Thus when passing from this generic 3-restriction to various 4-restrictions we should find that $1/32$ of the possibilities yield a 4-estimation of 4, while $19/32$ of them yield 2, and $3/8$ of them yield 0.

7.4.5. We then pass from our 3-restriction to various 4-restrictions. The 2-by-2 matrix $A$ is simply[33]

$$\begin{pmatrix}(p_2p_3,p_4)_{p_4} & (-1,p_4)_{p_4}\\ (2,p_4)_{p_4} & (p_1p_3,p_4)_{p_4}\end{pmatrix}$$

To have $A=0$ we need $p_4$ to be 1 mod 8 while all the $(p_i,p_4)_{p_4}$ for $1\le i\le 3$ need to be the same; this is $1/16$ of the possibilities, and then one of the 2 values $(d,p_4)_{p_4}$ yields a kernel of rank 4. A calculation shows that the 2 Legendre specifications that achieve this are $(l_{14},l_{24},l_{34},l_{44})$ as $(0,0,0,0)$ and $(1,1,1,1)$. In the former case the 2 new basis elements can be taken as $(1,p_4,p_4)$ and $(p_4,1,p_4)$, while in the latter $(p_1p_4,1,p_1p_4)$ and $(p_2p_3,p_2p_4,p_3p_4)$ are suitable lifts for them.

7.4.6. Finally, every 4-restriction is necessarily nongeneric (since $c=\tilde{r}-1$), and indeed the 2-Selmer rank is simply the fourth 2-Selmer estimation plus 2 for the global torsion.

In our running example, where $(p_1,p_2,p_3,p_4)\equiv(5,7,3,1)$ modulo 8, we now must have $p_5\equiv 1\,(8)$ for $d\equiv 1\,(8)$ to hold, and all the Legendre symbols with $p_5$ have already been specified by the values of $(d,p_i)_{p_i}$ for $1\le i\le 4$.

The above two cases that have 4-estimation of 4 do indeed in turn have 2-Selmer rank 6, with a basis given by the stated four elements and the global 2-torsion elements $(2,-d,-2d)$ and $(d,-1,-d)$ appended.

---

[32]Note here the existence of co-ordinates (namely $-1$ and 2) which are not dependent on the $p_i$, while Swinnerton-Dyer's Lemma 6 (our Lemma 8.5.5) removes such instances (as a convenience) when bounding the proportion of nongeneric specifications. See Footnote 39 for more.

I think one needs to have 4 primes involved to meet the wider notion of genericity, for which an example (with $d\equiv 1\,(8)$ and positive) is $(p_1,p_2,p_3,p_4)\equiv(3,3,3,5)$ modulo 8 and $(l_{11},l_{12},l_{13},l_{14},l_{22},l_{23},l_{24},l_{33},l_{34},l_{44})$ as $(0,1,1,1,0,0,0,0,0,0)$, with basis elements given by $(2p_2p_3p_4,-p_2p_4,-2p_3)$, $(p_1p_2p_4,p_3p_4,p_1p_2p_3)$, and $(2p_1p_3,-p_2p_3,-2p_1p_2)$.

[33]As with the previous Footnote 32, in a more "generic" example *all* of the entries of $A$ would depend on $\mathcal{L}$-information with the appended prime, and not merely on its $\mathcal{K}$-information.

## 8. Genericity and the 2-Selmer group

We next give a description of genericity, which turns out to be of some import when trying to generalize Smith's later results to the class group case.[34]

8.1. In the previous section we noted that the 2-Selmer rank of $E_d$ and its various estimations were determined by $E$ and the conditions $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ on $d$. However, a description of the actual elements of $U_{\mathcal{B}}$ (and ostensibly to determine if we are in a generic case) would therein involve writing things in terms of prime divisors of $d$.

As Kane [23, Proposition 18] alludes to, we can circumvent this by operating on purely formal symbols, thereby removing prime divisors of $d$ from the picture (and thus eviscerating any notion of a "random prime" from the terminology). While this is somewhat of an exercise in pedantry, it will hopefully clarify some matters.

We write $(a|b)^\star$ for the value of a nontrivial Legendre symbol in $\mathbf{F}_2$, with similarly $\tilde{\mathcal{K}}^\star$ and $\mathcal{L}^\star$ taking values in $\mathbf{F}_2$ rather than $\{\pm 1\}$.

8.1.1. We fix an elliptic curve $E : y^2 = (x - c_1)(x - c_2)(x - c_3)$, and thus the set of bad places $\tilde{\Omega}$ and the differences $\delta_{ij} = c_i - c_j$ between the roots of the cubic.

Given $u$, we define the set $\mathbf{P}_u$ of $u$ formal symbols $\dot{p}_i$ for $1 \le i \le u$. We operate on such symbols by multiplication/concatenation, and assume this is commutative. We define $\nu_{\dot{p}}$ to be a "quadratic non-residue" for the formal symbol $\dot{p}$, so that we have $(\nu_{\dot{p}}|\dot{p}) = (\dot{p}, \nu_{\dot{p}})_{\dot{p}} = 1 \in \mathbf{F}_2$. As such, we then consider $\mathbf{Q}^\star_{\dot{p}}/(\mathbf{Q}^\star_{\dot{p}})^2$ to have four elements, which we take to be $\{1, \nu_{\dot{p}}, \dot{p}, \nu_{\dot{p}}\dot{p}\}$.

8.1.2. We will also make use of basic formal variables $\dot{\mathcal{K}}^\star_{qj}$ for $q \in \tilde{\Omega}$ and $1 \le j \le \tilde{r}$, and similarly $\dot{\mathcal{L}}^\star_{ij}$ for $1 \le i < j \le \tilde{r}$. (The moniker of "variables" here is simply to distinguish them from the above formal symbols). We also have the derived formal variables $\dot{\mathcal{K}}^\star_{q0} = \sum_j \dot{\mathcal{K}}^\star_{qj}$ for $q \in \tilde{\Omega}$, and for $j > i$ we have (corresponding to quadratic reciprocity) $\dot{\mathcal{L}}^\star_{ji} = \dot{\mathcal{L}}^\star_{ij} + \dot{\mathcal{K}}^\star_{\infty i}\dot{\mathcal{K}}^\star_{\infty j}$, while $\dot{\mathcal{L}}^\star_{jj} = \sum_{i \ne j} \dot{\mathcal{L}}^\star_{ij}$.

We define Hilbert symbols for the formal symbols in terms of the formal variables, namely as $(q|\dot{p}_j)^\star = (q, \dot{p}_j)_{\dot{p}_j} = \dot{\mathcal{K}}^\star_{qj}$ for $q \in \tilde{\Omega}$ (with $q = -1$ for $\infty$) and $1 \le j \le \tilde{r}$, and $(\dot{p}_i|\dot{p}_j)^\star = (\dot{p}_i, \dot{p}_j)_{\dot{p}_j} = \dot{\mathcal{L}}^\star_{ij}$ for $1 \le i \ne j \le \tilde{r}$, and $(\dot{p}_j, \dot{p}_j)_{\dot{p}_j} = (-1, \dot{p}_j)_{\dot{p}_j} = \dot{\mathcal{K}}^\star_{\infty j}$. We also define the convenient $\dot{d} = \dot{\varepsilon}\dot{p}_1 \cdots \dot{p}_{\tilde{r}}$ so that we have $\dot{\mathcal{L}}^\star_{jj} = (\dot{d}/\dot{\varepsilon}\dot{p}_j, \dot{p}_j)_{\dot{p}_j}$. Here $\dot{\varepsilon}$ corresponds to the sign of $d$, and has either $(\dot{\varepsilon}, \dot{p}_j)_{\dot{p}_j} = 0$ for all $j$ when it is positive, while $(\dot{\varepsilon}, \dot{p}_j)_{\dot{p}_j} = (-1, \dot{p}_j)_{\dot{p}_j} = \dot{\mathcal{K}}^\star_{\infty j}$ otherwise.

We note $(x, y)_{\dot{p}} = 0$ when neither $x$ nor $y$ is divisible by $\dot{p}$; also, we often simplify via multiplicativity, for instance if $\dot{p} \| x$ then $(x, y)_{\dot{p}} = (x/\dot{p}, y)_{\dot{p}} + (\dot{p}, y)_{\dot{p}} = 0 + (\dot{p}, y)_{\dot{p}}$.

8.1.3. Given a linear combination in the formal variables $\dot{\mathcal{K}}^\star$ and $\dot{\mathcal{L}}^\star$, one idea is to write it in a canonical form in terms of the basic formal variables. This need no longer be a linear combination, though the only non-monomials will be terms like $\dot{\mathcal{K}}^\star_{\infty i}\dot{\mathcal{K}}^\star_{\infty j}$ that arise when replacing $\dot{\mathcal{L}}^\star_{ji}$ by $\dot{\mathcal{L}}^\star_{ij}$. On the other hand, we do not particularly want to consider $\dot{\mathcal{L}}^\star_{ij}$ and $\dot{\mathcal{L}}^\star_{ji}$ to be independent in any event.

We thus consider a set of linear combinations in the formal variables to be *independent* if they are independent in the free algebra (over $\mathbf{F}_2$) modulo: the relations $\dot{\mathcal{L}}^\star_{ij} = \dot{\mathcal{L}}^\star_{ji}$ for $i \ne j$; the relations from $\dot{\mathcal{K}}^\star_{q0} = \sum_j \dot{\mathcal{K}}^\star_{qj}$ for $q \in \tilde{\Omega}$; and the relations $\dot{\mathcal{L}}^\star_{jj} = \sum_{i \ne j} \dot{\mathcal{L}}^\star_{ij}$ for $1 \le j \le \tilde{r}$.

[34]The substitute in [3, Definition 6.4] requires every nonobvious vector in the kernel of the $\mathcal{L}$-matrix (interpreted over $\mathbf{F}_2$) to have approximately $\tilde{r}/2$ ones and $\tilde{r}/2$ zeros in its co-ordinates.

The point of independence comes about when we apply the evaluation homomorphism at $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$ to the formal variables $(\dot{\mathcal{K}}^\star, \dot{\mathcal{L}}^\star)$. We call a linear combination in the formal variables an *expression*. If a set of $n$ expressions is independent, then exactly $1/2^n$ of the possible $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$ have the evaluation of all the expressions be 0.

Indeed we should stress that this conclusion is precisely why we formalised the notion of "random primes" in the first place.

8.1.4. Given an expression, we can thus write it in a canonical form as a linear combination of basic formal variables, using the above quotient-algebra relations to do so. If this canonical form includes a specific basic formal variable, we say that the expression *depends* on said variable.

Our plan to show independence of a set of expressions will be a form of triangularisation; namely, we will find a basic formal variable such that exactly one expression depends on it – said expression is then independent of the others in the set, and we can recurse on the complement after removing it. It also useful to note that the "diagonal" formal variables $\dot{\mathcal{L}}^\star_{jj}$ have a useful property herein. To wit, if there is some $i$ such that none of the expressions depends[35] on $\dot{\mathcal{L}}^\star_{ij}$ (for $i < j$) or $\dot{\mathcal{L}}^\star_{ji}$ (for $j < i$), then a unique expression with $\dot{\mathcal{L}}^\star_{jj}$ appearing will be independent of the others; indeed, either $\dot{\mathcal{L}}^\star_{ij}$ or $\dot{\mathcal{L}}^\star_{ji}$ will appear in its canonical form, and it will be the only expression to depend on such. Moreover, this situation commonly occurs in our setup of restricting the Legendre conditions to the first $c$ formal variables, with consideration of appending the $(c+1)$st; when $c < \tilde{r} - 1$ we will see that none of the relevant expressions will depend on $\dot{\mathcal{L}}^\star_{j\tilde{r}}$ for any $j$, and the above commentary then implies that a unique expression involving $\dot{\mathcal{L}}^\star_{jj}$ will be independent of the others.

8.2. We then define $V_{\mathcal{B}}$ analogously to the schema of the previous section (again following [50, §3]), though now in terms of formal symbols rather than primes. We write $\mathcal{B} = \tilde{\Omega} \cup \mathbf{P}_{\tilde{r}}$, and given $l \in \mathcal{B}$ we define $Y_l = \mathbf{Q}_l^\star / (\mathbf{Q}_l^\star)^2$, and $V_l$ as the vector space of triples $(\mu_1, \mu_2, \mu_3) \in Y_l^3$ with $\mu_1\mu_2\mu_3 = 1$.

The pairing of Tate on $V_l \times V_l$ is then given as the sum $\sum_i (m_i, m_i')_l$ over the 3 coordinates. Here the Hilbert symbol $(a, b)_l$ is well-defined for $l \in \mathcal{B}$ for $a, b \in \bigoplus_{l \in \mathcal{B}} Y_l$ in terms of various formal variables $\dot{\mathcal{K}}^\star$ and $\dot{\mathcal{L}}^\star$. We then extend this pairing by additivity to $V_{\mathcal{B}} \times V_{\mathcal{B}}$ where $V_{\mathcal{B}} = \bigoplus_{l \in \mathcal{B}} V_l$.

8.2.1. We then define $U_{\mathcal{B}}$ to be the subspace generated by the diagonally embedded elements $(1, l, l)$ and $(l, 1, l)$ (and $(l, l, 1)$ if desired) over $l \in \mathcal{B}$, where again $l = -1$ is taken for the infinite place. We further define $U_{\mathcal{B}}^c$ as the subspace generated by such elements for $l \in \tilde{\Omega} \cup \mathbf{P}_c$. We have $\dim V_{\mathcal{B}} = 4(\tilde{r} + \#\tilde{\Omega})$ while $\dim U_{\mathcal{B}}^c = 2(c + \#\tilde{\Omega})$. Indeed, by class field theory one can show $U_{\mathcal{B}}$ is a maximal isotropic subspace for $e_{\mathcal{B}}$.

There is a homomorphism (the Kummer map) from $E_{\dot{d}}$ to $V_{\mathcal{B}}$ that is given by $(X, Y) \to (X - \dot{d}c_1, X - \dot{d}c_2, X - \dot{d}c_3)$ away from 2-torsion points, and by continuity we find the respective images of the 2-torsion points $(\dot{d}c_i, 0)$ to be

$$\vec{w}^1 = (\delta_{12}\delta_{13}, \dot{d}\delta_{12}, \dot{d}\delta_{13}), \ \vec{w}^2 = (\dot{d}\delta_{21}, \delta_{21}\delta_{23}, \dot{d}\delta_{23}), \text{ and } \vec{w}^3 = (\dot{d}\delta_{31}, \dot{d}\delta_{32}, \delta_{31}\delta_{32}),$$

These are globally defined, and for $l \in \mathcal{B}$ we take $W_l$ to be the subspace of $V_l$ generated by their projections. For $l \notin \{2, \infty\}$ this $W_l$ is a maximal isotropic

[35]There is a important point to be aware of here, namely that if $\dot{\mathcal{L}}^\star_{ii}$ appears in an expression, such an expression will then depend on either $\dot{\mathcal{L}}^\star_{ij}$ or $\dot{\mathcal{L}}^\star_{ji}$.

subspace with respect to $e_l$ (as Tate showed), and upon considering $l \in \{2, \infty\}$ separately one finds that $\dim W_{\mathcal{B}} = 2(\tilde{r} + \#\tilde{\Omega})$.

8.2.2. Upon choosing bases for $U_{\mathcal{B}}$ and $W_{\mathcal{B}}$ and applying the evaluation homomorphism at a given $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$ to the formal symbols $(\dot{\mathcal{K}}^\star, \dot{\mathcal{L}}^\star)$, the pairing matrix $\mathbf{M}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is then well-defined (with $\mathbf{F}_2$-entries) on $U_{\mathcal{B}} \times W_{\mathcal{B}} \subset V_{\mathcal{B}} \times V_{\mathcal{B}}$, and the dimension of its kernel is independent of a choice of basis. We denote this rank by $s_{\tilde{r}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$. The left kernel of the pairing matrix corresponds to elements of $U_{\mathcal{B}}$ that are "everywhere locally soluble" when interpreted in terms of elliptic curves.

Moreover, we can consider the restriction of the pairing matrix to the symbols in $\tilde{\Omega} \cup \mathbf{P}_c$, thus considering only $U^c_{\mathcal{B}}$ and $W^c_{\mathcal{B}}$ (the latter being the direct sum of $W_l$ for $l \in \tilde{\Omega} \cup \mathbf{P}_c$) and we denote the dimension of its kernel by $s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$.

8.3. We now describe how to obtain a relatively nice isomorphism $\tau_{\mathcal{B}}$ between $U_{\mathcal{B}}$ and $W_{\mathcal{B}}$, or at least on the relevant subspaces therein. As Swinnerton-Dyer notes, this can be done by specifying maximal isotropic subspaces $K_l \subset V_l$ for each $l$ in such a way that $V_{\mathcal{B}} = U_{\mathcal{B}} \bigoplus \oplus_l K_l$. We then define

$$U'_{\mathcal{B}} = U_B \oplus (W_{\mathcal{B}} + K_{\mathcal{B}}) \text{ and } W'_B = W_{\mathcal{B}}/(W_{\mathcal{B}} \cap K_{\mathcal{B}}) = \bigoplus_{l \in \mathcal{B}} W_l/(W_l \cap K_l),$$

and upon taking $t_{\mathcal{B}} : V_{\mathcal{B}} \to U_{\mathcal{B}}$ to be the projection along $\oplus_l K_l$, this induces an isomorphism $\tau_{\mathcal{B}}$ from $W'_{\mathcal{B}}$ to $U'_{\mathcal{B}}$, which suffices for our purposes; in particular the induced pairing $e'_{\mathcal{B}}$ is then symmetric. One aims to select $K_l$ in such a way to ensure $U'_{\mathcal{B}}$ is as small as possible, namely equal to $U_{\mathcal{B}} \cap W_{\mathcal{B}}$.

For $l \in \mathbf{P}_{\tilde{r}}$ we can be more direct and indeed select the isotropic subspace $K_{\dot{p}}$ explicitly as being generated by $(1, \nu_{\dot{p}}, \nu_{\dot{p}})$ and $(\nu_{\dot{p}}, 1, \nu_{\dot{p}})$, with the resulting $\dot{p}$-part then being orthogonal, and we can ensure the pairing is alternating. For $l \in \tilde{\Omega}$ the situation is more complex,[36] and we will simply be happy with a symmetric pairing.

8.3.1. We now give an explicit choice of isomorphism that yields an alternating pairing for formal symbols. We do this locally (as given by a local isomorphism $\tau_{\dot{p}}$) for every formal symbol $\dot{p}$, and then note how to glue them together. The global torsion gives $\vec{w}^1_{\dot{p}} = (\delta_{12}\delta_{13}, \dot{d}\delta_{12}, \dot{d}\delta_{13})$ and $\vec{w}^2_{\dot{p}} = (\dot{d}\delta_{21}, \delta_{21}\delta_{23}, \dot{d}\delta_{23})$ as a basis for $W_{\dot{p}}$. We want to determine $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$ and $\tau_{\dot{p}}\vec{\alpha}^2_{\dot{p}}$ such that the 2-by-2 matrix of their pairings with $\vec{w}^1_{\dot{p}}$ and $\vec{w}^2_{\dot{p}}$ is alternating, moreover with the off-diagonal term being $(\dot{d}, \dot{p})_{\dot{p}}$.

By rotely computing the pairing,[37] we find there are four elements $x \in V_{\dot{p}}$ that have $e_{\dot{p}}(x, \vec{w}^1_{\dot{p}})$ not depending on $(\dot{d}, \dot{p})_{\dot{p}}$. First there is $(1, \dot{p}, \dot{p})$ which has the pairing as $(\dot{p}, \dot{d}\delta_{12})_{\dot{p}} + (\dot{p}, \dot{d}\delta_{13})_{\dot{p}} = (\dot{p}, \delta_{12}\delta_{13})_{\dot{p}}$, and $(1, \nu_{\dot{p}}\dot{p}, \nu_{\dot{p}}\dot{p})$ for which it is similarly given by $(\nu_{\dot{p}}\dot{p}, \delta_{12}\delta_{13})_{\dot{p}} = (\dot{p}, \delta_{12}\delta_{13})_{\dot{p}}$; and $(\nu_{\dot{p}}, \dot{p}, \nu_{\dot{p}}\dot{p})$ which yields

$$\begin{aligned}(\nu_{\dot{p}}, \delta_{12}\delta_{13})_{\dot{p}} + (\dot{p}, \dot{d}\delta_{12})_{\dot{p}} + (\nu_{\dot{p}}\dot{p}, \dot{d}\delta_{13})_{\dot{p}} &= 0 + (\dot{p}, \dot{d}\delta_{12})_{\dot{p}} + (\dot{p}, \dot{d}\delta_{13})_{\dot{p}} + (\nu_{\dot{p}}, \dot{d}\delta_{13})_{\dot{p}} \\ &= (\dot{p}, \delta_{12}\delta_{13})_{\dot{p}} + (\nu_{\dot{p}}, \dot{p})_{\dot{p}} = (\dot{p}, \delta_{12}\delta_{13})_{\dot{p}} + 1,\end{aligned}$$

with a similar computation giving the same result for $(\nu_{\dot{p}}, \nu_{\dot{p}}\dot{p}, \dot{p})$. Thus two of these four possibilities for $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$ will have a pairing equal to 0 with $\vec{w}^1_{\dot{p}}$, with which two depending on whether $\delta_{12}\delta_{13}$ is a square modulo $\dot{p}$ (thus determined by $\tilde{\mathcal{K}}$).

[36] As Swinnerton-Dyer notes in his Theorem 2, one can achieve the alternating aspect for all $l \in \tilde{\Omega}$ other than 2, the infinite place, and those primes that have even valuation at all the $\delta_{ij}$.

[37] Swinnerton-Dyer achieves the same conclusion at the bottom of page 524, via his Lemma 2.

We then want $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$ to have a pairing with $\vec{w}^2_{\dot{p}}$ that only depends on $(\dot{d},\dot{p})_{\dot{p}}$. The pairing of $(1,\dot{p},\dot{p})$ with $\vec{w}^2_{\dot{p}}$ is $(\dot{p},\dot{d}\delta_{21})_{\dot{p}} = (\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{21})_{\dot{p}}$ while the pairing of $(1,\nu_{\dot{p}}\dot{p},\nu_{\dot{p}}\dot{p})$ with it is

$$(\nu_{\dot{p}}\dot{p},\dot{d}\delta_{21})_{\dot{p}} = (\nu_{\dot{p}},\dot{d}\delta_{21})_{\dot{p}} + (\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{21})_{\dot{p}} = 1 + (\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{21})_{\dot{p}}.$$

Thus one of these two will have a pairing-value of $(\dot{p},\dot{d})_{\dot{p}}$ with $\vec{w}^2_{\dot{p}}$, depending on whether $\delta_{21}$ is square modulo $\dot{p}$. When $\delta_{12}\delta_{13}$ is square modulo $\dot{p}$, this then gives us our $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$. Similarly, we have that the pairing of $(\nu_{\dot{p}},\dot{p},\nu_{\dot{p}}\dot{p})$ with $\vec{w}^2_{\dot{p}}$ is

$$\begin{aligned}(\nu_{\dot{p}},\dot{d}\delta_{21})_{\dot{p}} + (\dot{p},\delta_{21}\delta_{23})_{\dot{p}} + (\nu_{\dot{p}}\dot{p},\dot{d}\delta_{23})_{\dot{p}} &= 1 + (\dot{p},\delta_{21}\delta_{23})_{\dot{p}} + [(\nu_{\dot{p}},\dot{d}\delta_{23})_{\dot{p}} + (\dot{p},\dot{d}\delta_{23})_{\dot{p}}]\\ &= (\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{21})_{\dot{p}},\end{aligned}$$

and that with $(\nu_{\dot{p}},\nu_{\dot{p}}\dot{p},\dot{p})$ is

$$\begin{aligned}(\nu_{\dot{p}},\dot{d}\delta_{21})_{\dot{p}} + (\nu_{\dot{p}}\dot{p},\delta_{21}\delta_{23})_{\dot{p}} + (\dot{p},\dot{d}\delta_{23})_{\dot{p}} &= 1 + (\dot{p},\delta_{21}\delta_{23})_{\dot{p}} + [(\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{23})_{\dot{p}}]\\ &= 1 + (\dot{p},\dot{d})_{\dot{p}} + (\dot{p},\delta_{21})_{\dot{p}},\end{aligned}$$

so again one of these pairing-values will be $(\dot{p},\dot{d})_{\dot{p}}$, giving us our choice of $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$ in the case where $\delta_{12}\delta_{13}$ is not a square modulo $\dot{p}$.

This then gives an adequate choice of $\tau_{\dot{p}}\vec{\alpha}^1_{\dot{p}}$, and we can repeat the process to choose $\tau_{\dot{p}}\vec{\alpha}^2_{\dot{p}}$ with respect to $\vec{w}^2_{\dot{p}}$, with the result being alternating as desired.

8.3.2. The above gives a method for selecting $\tau_{\dot{p}}$ locally for each $\dot{p}$; however, there is still the task of glueing these together into a global $\tau$-map, which need not be trivial, particularly due to the occurrence of $\nu_{\dot{p}}$. It turns out (as an application of quadratic reciprocity and/or class field theory) that we can select $\nu_{\dot{p}}$ freely in $V_{\mathcal{B}}$ for all but one of the formal symbols (which we take to be $\dot{p}_{\tilde{r}}$). The local isomorphism at the final formal symbol is then fixed by the condition with the global torsion (rather than by the above setup), and indeed in this case we have $(\tau_{\dot{p}}\vec{\alpha}^i_{\dot{p}},\vec{w}^j_{\dot{p}})_{\dot{p}} = 0$ for $i,j \in \{1,2\}$ (equivalently, the matrix $G$ below has $C = 0$).

Writing $\dot{p}_k$ for $\dot{p}$ now, note also that the off-diagonal entry of $(\dot{p}_k,\dot{d})_{\dot{p}_k}$ is only the *local* image, and as in the extended example in §7.4 one needs to lift it, which can cause various $(\dot{p}_k,\dot{p}_i)_{\dot{p}_k}$ for $i<k$ to also appear. However, the off-diagonal entry of $C$ will always *depend* on $(\dot{p}_k,\dot{d})_{\dot{p}_k}$.

8.4. We can then consider the process of passing from $\mathbf{M}_c(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ to $\mathbf{M}_{c+1}(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ for $0 \le c < \tilde{r}$, writing $\dot{p}$ for the $(c+1)$st formal symbol (it is fixed for our analysis). This follows [50, §5] of Swinnerton-Dyer.

We restrict the pairing matrix to the kernels of $\mathbf{M}_c$, writing $\tilde{U}_c = U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon,\mathcal{L})$ for the relevant subset of $U^c_{\mathcal{B}}$, and similarly with $\tilde{W}_c$ for $W^c_{\mathcal{B}}$. Upon appropriate choice of bases the pairing matrix on $(\tilde{U}_c \oplus \tau_{\dot{p}}U_{\dot{p}}) \times (\tilde{W}_c \oplus W_{\dot{p}})$ is then of the form

$$G = \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & C \end{pmatrix}$$

where $C$ is alternating with non-diagonal entry depending on $(\dot{p},\dot{d})_{\dot{p}}$ when $c < \tilde{r}-1$.

8.4.1. We can then compute $s_{c+1}$ in terms of $s_c$ and the ranks of $A$ and $C$. Indeed, when $A = C = 0$ the rank of $G$ is 0 and thus $s_{c+1} = s_c + 2$. When $A$ itself has rank 2 we see that $G$ has rank 4 and so $s_{c+1} = s_c - 2$, while if $A$ has rank 1 we find that $G$ has rank 2 (independently of what $C$ is) and $s_{c+1} = s_c$.

Note that when $\tilde{U}_{c+1}$ contains two elements that depend on $\dot{p}_{c+1}$ in an independent manner, we must then have $s_{c+1} = s_c + 2$, and so $A = C = 0$.

This will certainly be the case when $c = \tilde{r} - 1$, since then $\tilde{U}_{\tilde{r}}$ contains the images under $\tau$ of the global elements $\vec{w}^1 = (\delta_{12}\delta_{13}, \dot{d}\delta_{12}, \dot{d}\delta_{13})$ and $\vec{w}^2 = (\dot{d}\delta_{21}, \delta_{21}\delta_{23}, \dot{d}\delta_{23})$, and these are not in $\tilde{U}_{\tilde{r}-1}$ since they depend on $\dot{d}$ (and thus $\dot{p}_{\tilde{r}}$). From this we find that $s_{\tilde{r}} = 2 + s_{\tilde{r}-1}$.

8.4.2. In general, we will need to know more about the entries of $A$. We write $Y^0_{\mathcal{B}}$ for the direct sum of the $Y_l = \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$ over $l \in \tilde{\Omega}$.

For a given element $(u_1, u_2, u_3) \in \tilde{U}_c$, each $u_i$ can be written as $\xi \prod_a \dot{p}_a^{e_a}$ for some $\xi \in Y^0_{\mathcal{B}}$ and $\vec{e}$ with $e_a \in \{0, 1\}$ for $1 \le a \le c$. In particular $\dot{p} = \dot{p}_{c+1}$ will not divide $u_i$. Thus the pairing of $(u_1, u_2, u_3)$ with $\vec{w}^1_{\dot{p}}$ is given by

$$(\delta_{12}\delta_{13}, u_1)_{\dot{p}} + (\dot{d}\delta_{12}, u_2)_{\dot{p}} + (\dot{d}\delta_{13}, u_3)_{\dot{p}} = 0 + (\dot{p}, u_2)_{\dot{p}} + (\dot{p}, u_3)_{\dot{p}} = (\dot{p}, u_1)_{\dot{p}},$$

the last step since $u_1u_2u_3 = 1$. The pairing with $\vec{w}^2_{\dot{p}}$ similarly gives $(\dot{p}, u_2)_{\dot{p}}$.

The entries of $A$ thus consist of $(\dot{p}, u_1)_{\dot{p}}$ and $(\dot{p}, u_2)_{\dot{p}}$ for various $(u_1, u_2, u_3) \in U^c_{\mathcal{B}}$, and each $(\dot{p}, u)_{\dot{p}} = (\dot{p}_{c+1}, u)_{\dot{p}}$ is some expression in the $\dot{\mathcal{K}}^\star_{q,(c+1)}$ for $q \in \tilde{\Omega}$ and the $\dot{\mathcal{L}}^\star_{(c+1),j}$ with $1 \le j \le c$. Thus the expression $(\dot{p}, u)_{\dot{p}}$ is nontrivial unless $u = 1$.

Meanwhile, for $(\dot{p}, \dot{d})_{\dot{p}}$ from $C$, this is equal to $(\dot{p}, \dot{d}/\dot{\varepsilon}\dot{p})_{\dot{p}} + (\dot{p}, \dot{\varepsilon}\dot{p})_{\dot{p}}$, where the former is $\dot{\mathcal{L}}^\star_{(c+1),(c+1)}$ and the latter is either $\dot{\mathcal{K}}^\star_{\infty,(c+1)}$ or 0 (depending on $\dot{\varepsilon}$). In particular, when $c < \tilde{r} - 1$ we see that this expression is independent of the others (in the sense of §8.1.3), due to the presence of the diagonal entry $\dot{\mathcal{L}}^\star_{(c+1),(c+1)}$.

As for the entries of $A$, we write $\{u^n\}$ for a basis of of $U_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ and consider the $2s_c$ expressions given by $(\dot{p}, u_1^n)_{\dot{p}}$ and $(\dot{p}, u_2^n)_{\dot{p}}$ for $1 \le n \le s_c$. A dependency between these expressions would in particular imply there is some nonempty minimal subset $\bigcup_{a\in Z_1}\{(\dot{p}, u_1^a)_{\dot{p}}\} \cup \bigcup_{a\in Z_2}\{(\dot{p}, u_2^b)_{\dot{p}}\}$ of said expressions for which the product $\prod_a(u_1^a)\prod_b(u_2^b)$ is equal to 1. If one of the products is empty, say the second, then we can just add the elements $(u^a)$ themselves together and get something whose first component is 1. Otherwise, we can add the elements $(u^a)$ together to get $(U_1, U_2, U_3)$ and similarly the $(u^b)$ to get $(U_1', U_2', U_3')$ with $U_1 = U_2'$. Thus if $\tilde{U}_c$ contains no nontrivial element of the form $(1, U_2, U_2)$ or $(U_1, 1, U_1)$, and no 2 nontrivial elements of the form $(U_1, U_2, U_3)$ and $(U_1', U_2', U_3')$ with $U_1 = U_2'$, the entries of $A$ then determine $2s_c$ independent expressions in the formal variables (and thus are independent in the sense of specifying a coset of $\mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$).

8.4.3. This then gives the definition: a specification $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ is *generic* at $c$ if $U_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ has no nontrivial element of the form $(1, u_0, u_0)$ or $(u_0, 1, u_0)$, and also doesn't contain 2 nontrivial elements $(u_1, u_2, u_3)$ and $(v_1, v_2, v_3)$ with $u_1 = v_2$. We similarly define genericity of $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$, and sum up as follows.

**Lemma 8.4.4.** *Let $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ and suppose it is generic. Consider the set of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c+1]$ whose $c$-restriction is $(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$. The $s_{c+1}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ are then distributed from $s = s_c(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}')$ in the following manner:*

- *a proportion $1/2^{2s+1}$ of them have $s_{c+1}(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') = s + 2 = s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) + 2$,*

- *a proportion* $3/2^s - 5/2^{2s+1}$ *have* $s_{c+1}(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') = s = s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$,
- *a proportion* $1 - 3/2^s + 2/4^s$ *have* $s_{c+1}(\tilde{\mathcal{K}}'_\varepsilon, \mathcal{L}') = s - 2 = s_c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) - 2$.

*Proof.* Given the above description of $s_{c+1}$ in terms of $s_c$ and the assumption of genericity, we need only note that the given proportions correspond to the number of choices of $(A, C)$ with the requisite ranks of $G$ as given above. □

8.5. Next we will follow Swinnerton-Dyer's analysis [50, §4] to bound the size of the set of the nongeneric $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$. This is perhaps little more than 5+ pages of re-presenting his (lengthy) arguments, though we make some corrections and minor improvements, and handle the notion of "probability" more robustly.

8.5.1. We begin by noting some basic calculations for the pairings. Let $(u_1, u_2, u_3)$ be an element of $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, so that it is in $U^c_{\mathcal{B}}$ and meets the $c$-restriction of the $(\mathcal{K}_\varepsilon, \mathcal{L})$ conditions.

Let $\dot{p} = \dot{p}_a$ for some $1 \le a \le c$ and suppose it does not divide any of the components $u_i$. The pairing of $(u_1, u_2, u_3)$ with $\vec{w}^1_{\dot{p}} = (\delta_{12}\delta_{13}, \dot{d}\delta_{12}, \dot{d}\delta_{13})$ is then

$$(\delta_{12}\delta_{13}, u_1)_{\dot{p}} + (\dot{d}\delta_{12}, u_2)_{\dot{p}} + (\dot{d}\delta_{13}, u_3)_{\dot{p}} = 0 + (\dot{p}, u_2)_{\dot{p}} + (\dot{p}, u_3)_{\dot{p}} = (\dot{p}, u_2u_3)_{\dot{p}} = (\dot{p}, u_1)_{\dot{p}},$$

and similarly the pairings with $\vec{w}^2_{\dot{p}}$ and $\vec{w}^3_{\dot{p}}$ are respectively $(\dot{p}, u_2)_{\dot{p}}$ and $(\dot{p}, u_3)_{\dot{p}}$. The conditions that $(u_1, u_2, u_3)$ be locally soluble at $\dot{p}$ (and so be in the space $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ of everywhere locally soluble elements) are thus that each of the expressions

$$(\dot{p}, u_1)_{\dot{p}}, \ \ (\dot{p}, u_2)_{\dot{p}}, \ \text{ and } \ (\dot{p}, u_3)_{\dot{p}} \tag{9}$$

should be zero when evaluating the formal variables $(\dot{\mathcal{K}}^\star, \dot{\mathcal{L}}^\star)$ therein at $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$. (This is at most two independent conditions, since $u_3 = u_1u_2$).

When $\dot{p}$ divides one of the components of $u$, then it divides exactly two of them. When $u_1$ is a unit at $\dot{p}$, then we find that the pairing with $\vec{w}^1_{\dot{p}}$ is (writing $u_2 = \dot{p}\xi_2$)

$$\begin{aligned} 0 + (\dot{d}\delta_{12}, u_2)_{\dot{p}} + (\dot{d}\delta_{13}, u_3)_{\dot{p}} &= (\dot{d}\delta_{12}, p)_{\dot{p}} + (\dot{d}\delta_{12}, \xi_2)_{\dot{p}} + (\dot{d}\delta_{13}, p)_{\dot{p}} + (\dot{d}\delta_{13}, \xi_3)_{\dot{p}} \\ &= (\delta_{12}\delta_{13}, p)_{\dot{p}} + (\dot{d}, \xi_2\xi_3)_{\dot{p}} = (\delta_{12}\delta_{13}, p)_{\dot{p}} + (\dot{p}, u_1)_{\dot{p}}, \end{aligned}$$

where we used $(\dot{d}, \xi_2\xi_3)_{\dot{p}} = (\dot{p}, \xi_2\xi_3)_{\dot{p}} = (\dot{p}, \dot{p}\xi_2 \cdot \dot{p}\xi_3)_{\dot{p}} = (\dot{p}, u_2u_3)_{\dot{p}} = (\dot{p}, u_1)_{\dot{p}}$ in the last step. Again this must evaluate to 0 for $(u_1, u_2, u_3)$ to be in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$. The other two computations take a different form, with the pairing with $\vec{w}^2_{\dot{p}}$ being

$$\begin{aligned} &(\dot{d}\delta_{21}, u_1)_{\dot{p}} + (\delta_{21}\delta_{23}, u_2)_{\dot{p}} + (\dot{d}\delta_{23}, u_3)_{\dot{p}} \\ &\qquad = (\dot{p}, u_1)_{\dot{p}} + (\delta_{21}\delta_{23}, \dot{p})_{\dot{p}} + (\dot{d}\delta_{23}, \dot{p})_{\dot{p}} + (\dot{d}\delta_{23}, \xi_3)_{\dot{p}} \\ &\qquad = (\dot{p}, u_1)_{\dot{p}} + (\delta_{21}\delta_{23}, \dot{p})_{\dot{p}} + (\dot{d}, \dot{p})_{\dot{p}} + (\delta_{23}, \dot{p})_{\dot{p}} + (\dot{p}, \xi_3)_{\dot{p}} \\ &\qquad = (\delta_{21}, \dot{p})_{\dot{p}} + (\dot{p}, u_1u_3\dot{d}/\dot{p})_{\dot{p}} = (\delta_{21}, \dot{p})_{\dot{p}} + (\dot{p}, u_2)_{\dot{p}} + (\dot{p}, \dot{d}/\dot{p})_{\dot{p}}, \end{aligned}$$

and the pairing with $\vec{w}^3_{\dot{p}}$ is then $(\delta_{31}, \dot{p})_{\dot{p}} + (\dot{p}, u_3)_{\dot{p}} + (\dot{p}, \dot{d}/\dot{p})_{\dot{p}}$. Upon rotating the indices, we can then catalogue the expressions from the nine computations as:

$$\begin{array}{lll} (\delta_{12}\delta_{13}, \dot{p}) + (\dot{p}, u_1) & (\delta_{21}, \dot{p}) + (\dot{p}, u_2) + (\dot{p}, \dot{d}/\dot{p}) & (\delta_{31}, \dot{p}) + (\dot{p}, u_3) + (\dot{p}, \dot{d}/\dot{p}) \\ (\delta_{21}\delta_{23}, \dot{p}) + (\dot{p}, u_2) & (\delta_{12}, \dot{p}) + (\dot{p}, u_1) + (\dot{p}, \dot{d}/\dot{p}) & (\delta_{32}, \dot{p}) + (\dot{p}, u_3) + (\dot{p}, \dot{d}/\dot{p}) \\ (\delta_{31}\delta_{32}, \dot{p}) + (\dot{p}, u_3) & (\delta_{13}, \dot{p}) + (\dot{p}, u_1) + (\dot{p}, \dot{d}/\dot{p}) & (\delta_{23}, \dot{p}) + (\dot{p}, u_2) + (\dot{p}, \dot{d}/\dot{p}) \end{array} \tag{10}$$

where the first line corresponds to $u_1$ being a unit, the second line to $u_2$ being such, etc., and we have suppressed the $\dot{p}$-subscript on the Hilbert symbols. In each case, for $(u_1, u_2, u_3)$ to be in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ the given three expressions must be 0 (for

every $\dot{p}$) when evaluating the formal variables $(\dot{\mathcal{K}}^\star, \dot{\mathcal{L}}^\star)$ at $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$, though of course at most two of the three expressions are independent.

8.5.2. We now give some rudimentary bounds on the number of various nongeneric elements. However, the main (and most lengthy) demonstration will be left to the next subsection. We write $Y_{\mathcal{B}}^c$ for the sum of the $Y_l = \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$ over $l \in \tilde{\Omega} \cup \mathbf{P}_c$.

**Lemma 8.5.3.** [50, Lemma 4]. *The proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ with a nontrivial element in $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ of the form $(1, \xi, \xi)$ with $\xi \in Y_{\mathcal{B}}^0$ is $\ll (1/2)^c$ for $0 \le c < \tilde{r} - 1$.*

Clearly we can handle $(\xi, 1, \xi)$ and $(\xi, \xi, 1)$ in the same manner by symmetry.

*Proof.* Since $\xi$ is a unit for every formal symbol $\dot{p}_j$, from (9) we have expressions $(\dot{p}_j, \xi)_{\dot{p}_j}$ for $1 \le j \le c$, all of which must be zero when the formal variables $\dot{\mathcal{K}}^\star$ and $\dot{\mathcal{L}}^\star$ are evaluated. We wish to show that these expressions are all independent,[38] and this readily follows (when $c < \tilde{r} - 1$) as they involve $\dot{\mathcal{K}}_{qj}^\star$ for different $j$. In other words, we have $(\dot{p}_j, \xi)_{\dot{p}_j} = \sum_{q|\xi} \dot{\mathcal{K}}_{qj}^\star$ where the nonempty $q$-sum is over those $q \in \tilde{\Omega}$ with $q|\xi$, and we can take any such $q$ and note that $(\dot{p}_j, \xi)_{\dot{p}_j}$ is the unique expression that depends on the basic formal variable $\dot{\mathcal{K}}_{qj}^\star$.

The independence of expressions implies each reduces the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ by a factor of 2, so for any $\xi$ the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ with $(1, \xi, \xi)$ in $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $1/2^c$, and as the total number of $\xi$ is $(2^{\#\tilde{\Omega}} - 1) \ll 1$, the Lemma follows. □

**Lemma 8.5.4.** [50, Lemma 5]. *Suppose that $\delta_{12}\delta_{13}$ is not a square. The proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ with a nontrivial element in $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ of the form $(1, u_2, u_2)$ with $u_2 \in Y_{\mathcal{B}}^c$ is $\ll (3/4)^c$ for $0 \le c < \tilde{r} - 1$.*

Again we can also handle $(u_2, 1, u_2)$ and $(u_2, u_2, 1)$ analogously by symmetry.

*Proof.* Let $A$ be the set of formal symbols in $\mathbf{P}_c$ that divide $u_2$, and $B$ be the set of those that don't. For the latter, we again have the expressions $(\dot{p}_j, u_2)_{\dot{p}_j}$ from (9) that must evaluate to zero at $(\tilde{\mathcal{K}}^\star, \mathcal{L}^\star)$. On the other hand, for formal symbols in $A$, from the first line of (10) we have the expressions $(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j}$ and $(\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_2)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$.

Our tactic to show these expressions are independent shall be to show that each depends on some $\dot{\mathcal{K}}$- or $\dot{\mathcal{L}}$-variable that appears in no other expression. This is easiest for the last type of expression, as they involve $\dot{d}/\dot{p}_j$ and thus $\dot{\mathcal{L}}_{jj}^\star$ (and thus $\dot{\mathcal{L}}_{j\tilde{r}}^\star$ if one prefers), and as $c < \tilde{r} - 1$ these ensure independence; so these expressions are independent of each other, and also independent of those remaining.

We can assume that $u_2 \notin Y_{\mathcal{B}}^0$ (else the previous Lemma applies), so the expressions from $B$ with $(\dot{p}_j, u_2)_{\dot{p}_j}$ each involve at least one $\dot{\mathcal{L}}_{jk}^\star$ for some $k$. The only way such an expression could fail to be independent with the others is if some other expression included $\dot{\mathcal{L}}_{kj}^\star$. However, every $k$ for which $\dot{\mathcal{L}}_{jk}^\star$ appears in $(\dot{p}_j, u_2)_{\dot{p}_j} = \sum_{q|\xi_2} \dot{\mathcal{K}}_{qj}^\star + \sum_{k:\dot{p}_k|u_2} \dot{\mathcal{L}}_{jk}^\star$ corresponds to a formal symbol $\dot{p}_k \in A$, so in particular there is no occurrence of $\dot{\mathcal{L}}_{kj}^\star$ in the $B$-expressions. We conclude that

[38] Already in this case we can begin to see the nuances of extending the notion of genericity to a case where $\alpha_{\mathcal{P}} \ne 1$. For instance, when $\xi = -1$ and the formal symbols $\dot{p}_j$ correspond to primes that are restricted to be 1 mod 4, all the conditions are trivial.

these $B$-expressions are independent of each other (since each involves a different $j$), and independent of those remaining.

Finally, since $\delta_{12}\delta_{13}$ is not square (in $\mathbf{Q}$) the expressions $(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j}$ with $A$ are nontrivial. These indeed only involve $\dot{\mathcal{K}}$, and as each involves a different $j$, they are independent of each other.

Summing up, we have $\#B+2\#A = c+\#A$ expressions, all of which are independent in terms of basic formal variables, so for a given $u_2$ the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ with $(1, u_2, u_2)$ in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $1/2^{c+\#A}$. We then aggregate over possible choices of $u_2$, of which there are $2^{\#\tilde{\Omega}}\binom{c}{\#A}$ possibilities for a given value of $\#A$. This gives a total proportion of

$$\sum_{a=1}^{c} \binom{c}{a} \frac{2^{\#\tilde{\Omega}}}{2^{c+a}} \ll (3/4)^c$$

as the $a$-part of sum (with $a = 0$ included) is $(3/2)^c$ by the binomial expansion. □

We can generalize Lemma 8.5.4 slightly,[39] to further restrict the types of $\vec{u}$.

**Lemma 8.5.5.** [50, Lemma 6]. *The proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ for which there is a nontrivial element in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ of the form $(\xi, u_2, u_2\xi)$ (and again similarly by symmetry) with $\xi \in Y^0_{\mathcal{B}}$ is $\ll (3/4)^c$ for $0 \le c < \tilde{r} - 1$.*

*Proof.* We let $A$ be set of formal symbols in $\mathbf{P}_c$ that divide $u_2$ and $B$ be the set of those that do not. We can assume that $A$ is nonempty. For $\dot{p}_j \in B$ we have the expressions $(\xi, \dot{p}_j)_{\dot{p}_j}$ and $(u_2, \dot{p}_j)_{\dot{p}_j}$ from (9). For $\dot{p}_j \in A$ we take the second expression from (10), namely that $(\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_2)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$. (Note here $\xi$ could be $\delta_{12}\delta_{13}$, when the other expressions from (10) are not independent).

Again the latter expressions involve $\dot{\mathcal{L}}^\star_{jj}$ and thus are independent of each other and those remaining, while the second type of expression for formal symbols in $B$ involves $\dot{\mathcal{L}}^\star_{jk}$ for some $k$ that are all associated to a formal symbol in $A$, with thus $\dot{\mathcal{L}}^\star_{kj}$ not occurring and the expressions hence being independent of each other and those remaining. Finally, we can assume that $\xi \neq 1$, so that the expressions of the first type with $B$ are all independent. Thus we have $\#A + 2\#B = c + \#B$ independent expressions, and as before conclude the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $O\big((3/4)^c\big)$. □

Swinnerton-Dyer then launches into the main result, though he only gives details for a special case.[40] I think it is superior to split off this part of the proof explicitly.

**Lemma 8.5.6.** [50, Lemma 7]. *The proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ for which there are two elements in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ of the form $(u_1, u_2, u_3)$ and $(v_1, v_2, v_3)$ with $u_1 = v_2$ and $u_2 = v_3\xi$ for some $\xi \in Y^0_{\mathcal{B}}$ is $\ll (7/8)^c$ for $0 \le c < \tilde{r} - 1$.*

*Proof.* Let $(u_1, u_2, u_3)$ and $(u_3\xi, u_1, u_2\xi)$ be elements as specified. There are then four types of formal symbols $\dot{p}_j$ to consider; each will give at least two expressions that are mutually independent, and some types will yield a third. Then we will show that by removing at most 2 formal symbols (thus at most 6 expressions) from consideration the totality of the expressions remaining will be independent.

[39] As noted with Footnote 32, both this and Lemma 8.5.6 handle specific subcases that are convenient to avoid in the proof of Lemma 8.6.1, though if either fails one can still have genericity.

[40] Note that his second special case is superfluous, as we then have $(u_1, u_2, u_3)$ and $(v_1, v_2, v_3)$ with $u_3v_3 \in Y^0_{\mathcal{B}}$, and so can add $\vec{u}$ and $\vec{v}$ and be in the case of the previous Lemma.

We let $B_1$ be the set of formal symbols in $\mathbf{P}_c$ that divide $u_1$ and $u_2$ (and thus do not divide $u_3 = u_1u_2$). For $\dot{p}_j \in B_1$ we have (from the third line of (10))

$$(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_3)_{\dot{p}_j} \text{ and } (\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_1)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$$

from $(u_1, u_2, u_3)$, and from $(u_3\xi, u_1, u_2\xi)$ also have (second entry of first line)[41]

$$(\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_1)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j},$$

whereupon adding the latter two yields the expression $(\delta_{21}\delta_{13}, \dot{p}_j)_{\dot{p}_j}$.

With $B_2$ as the set of formal symbols that divide $u_1$ and $u_3$, for $\dot{p}_j \in B_2$ we have (second line)

$$(\delta_{21}\delta_{23}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_2)_{\dot{p}_j} \text{ and } (\delta_{12}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_1)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$$

from $(u_1, u_2, u_3)$, and from $(u_3\xi, u_1, u_2\xi)$ also have (last entry of third line)

$$(\delta_{23}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_1)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j},$$

so that adding the latter two yields the expression $(\delta_{12}\delta_{23}, \dot{p}_j)_{\dot{p}_j}$.

Letting $B_3$ be the set of formal symbols that divide $u_2$ and $u_3$, for $\dot{p}_j \in B_3$ we have (first line)

$$(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_1)_{\dot{p}_j} \text{ and } (\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_2)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$$

and this will suffice (for $(u_3\xi, u_1, u_2\xi)$, the relevant entries on the second line will involve $\xi$, and so it seems easier to just ignore them).

Lastly, with $B_4$ as the set of formal symbols that do not divide any $u_i$, for $\dot{p}_j \in B_4$ we get expressions from (9) as

$$(\dot{p}_j, u_1)_{\dot{p}_j}, \ (\dot{p}_j, u_2)_{\dot{p}_j}, \text{ and } (\dot{p}_j, u_3\xi)_{\dot{p}_j}.$$

Since $u_3 = u_1u_2$ the latter can be replaced by $(\dot{p}_j, \xi)_{\dot{p}_j}$, which is nontrivial for $\xi \neq 1$. (Swinnerton-Dyer has the third expressions from $B_1$ and $B_2$ in terms of $\xi$ (perhaps incorrectly), and then juxtaposes $\xi = 1$ into this).

In Table 7, we give the relevant expressions for each case (reading across the rows), where we suppressed the $j$-subscript on $\dot{p}_j$.

| | | | |
|---|---|---|---|
| $B_1$ | $(\delta_{31}\delta_{32}, \dot{p})_{\dot{p}} + (\dot{p}, u_3)_{\dot{p}}$ | $(\delta_{13}, \dot{p})_{\dot{p}} + (\dot{p}, u_1)_{\dot{p}} + (\dot{p}, \dot{d}/\dot{p})_{\dot{p}}$ | $(\delta_{21}\delta_{13}, \dot{p})_{\dot{p}}$ |
| $B_2$ | $(\delta_{21}\delta_{23}, \dot{p})_{\dot{p}} + (\dot{p}, u_2)_{\dot{p}}$ | $(\delta_{12}, \dot{p})_{\dot{p}} + (\dot{p}, u_1)_{\dot{p}} + (\dot{p}, \dot{d}/\dot{p})_{\dot{p}}$ | $(\delta_{12}\delta_{23}, \dot{p})_{\dot{p}}$ |
| $B_3$ | $(\delta_{12}\delta_{13}, \dot{p})_{\dot{p}} + (\dot{p}, u_1)_{\dot{p}}$ | $(\delta_{21}, \dot{p})_{\dot{p}} + (\dot{p}, u_2)_{\dot{p}} + (\dot{p}, \dot{d}/\dot{p})_{\dot{p}}$ | |
| $B_4$ | $(\dot{p}, u_1)_{\dot{p}}$ | $(\dot{p}, u_2)_{\dot{p}}$ | $(\dot{p}, \xi)_{\dot{p}}$ |

TABLE 7. Expressions in the various cases

[41] It's not completely clear to me, but I think Swinnerton-Dyer implicitly uses the first entry of the first line, whence $(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} = (\dot{p}_j, u_3\xi)_{\dot{p}_j}$ here, and $(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} = (\dot{p}_j, u_2\xi)_{\dot{p}_j}$ below, though I must admit I don't see how the computation then concludes in terms of merely $\xi$.

8.5.7. We thus have at least 2 and sometimes 3 expressions for each $\dot{p}_j$, which are at least independent among themselves. Similar to previously, the expressions that involve $\dot{d}/\dot{p}_j$ can all be shown to independent of each other and all the others, as they uniquely involve $\dot{\mathcal{L}}^\star_{jj}$. (Note that $B_4$ has no such expressions).

It is more subtle to consider the expressions that are of the form

$$(\delta, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u_i)_{\dot{p}_j} \text{ or } (\dot{p}_j, u_i)_{\dot{p}_j}, \tag{11}$$

the latter type coming from $B_4$ (which indeed gives two such expressions), and the former from the other three cases. Here we note $u_1u_2 \notin Y^0_{\mathcal{B}}$, as else the previous Lemma 8.5.5 applies with $u_3 = u_1u_2$. Thus there is some formal symbol dividing $u_1u_2$, which we call $\dot{p}^\star_1$. Moreover, since $u_1, u_2 \notin Y^0_{\mathcal{B}}$, there is some other formal symbol $\dot{p}^\star_2$ that divides one of $u_1$ or $u_2$ but not the other. We then exclude from consideration the (at most 6) expressions arising from these special formal symbols.

Letting $k_1$ and $k_2$ be the indices of these formal symbols, the above expressions in (11) depend on $\dot{\mathcal{L}}^\star_{jk_1}$ but not $\dot{\mathcal{L}}^\star_{jk_2}$ for $u_3 = u_1u_2$, on $\dot{\mathcal{L}}^\star_{jk_2}$ but not $\dot{\mathcal{L}}^\star_{jk_1}$ for whichever of $u_1$ or $u_2$ is divisible by $\dot{p}^\star_2$, and on both $\dot{\mathcal{L}}^\star_{jk_1}$ and $\dot{\mathcal{L}}^\star_{jk_2}$ for the other.[42]

Moreover, by construction (or destruction?) we have excluded $\dot{\mathcal{L}}^\star_{k_1j}$ and $\dot{\mathcal{L}}^\star_{k_2j}$ from occurring in the expressions, having removed those corresponding to $\dot{p}_{k_1}$ and $\dot{p}_{k_2}$. Thus these expressions are independent of each other, and also the remaining ones (which only depend on $\dot{\mathcal{K}}$).

Finally, we have the expressions $(\delta_{21}\delta_{13}, \dot{p}_j)_{\dot{p}_j}$ for $B_1$ and $(\delta_{12}\delta_{23}, \dot{p}_j)_{\dot{p}_j}$ for $B_2$; when the relevant $\delta$ is not square, they will be independent of each other (coming from different $j$) and from all the other expressions (not depending on $\dot{\mathcal{L}}$). We then conclude by noting

$$\delta_{21}\delta_{13} + \delta_{12}\delta_{23} = (c_2 - c_1)(c_1 - c_3) - (c_2 - c_1)(c_2 - c_3) = -(c_2 - c_1)^2 < 0,$$

so that at least one of the initial summands is not a square.

Writing $b_i$ for the size of $B_i$ (so that their sum is $c$), we deduce that there at least $2(b_1 + b_2 + b_3 + b_4) - 6 + \min(b_1, b_2)$ independent expressions. (We can simply ignore $(\dot{p}_j, \xi)_{\dot{p}_j}$ from $B_4$, though Swinnerton-Dyer does include it in his analysis).

8.5.8. Given two elements of the specified type, the above then gives that the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ with both elements in $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $\ll 1/2^{2c-6+\min(b_1,b_2)}$. Upon aggregating over possible choices of $(u_1, u_2, \xi)$, of which there are $2^{\#\tilde{\Omega}}\binom{c}{b_1,b_2,b_3,b_4}$ in terms of the vector $\vec{b}$ (of divisibilities) associated to $(u_1, u_2)$, summing over all such possibilities of $\vec{b}$ then gives a proportion of

$$\ll \frac{1}{2^{2c}} \sum_{\sum_i b_i = c} \binom{c}{b_1, b_2, b_3, b_4} \Big(\frac{1}{2^{b_1}} + \frac{1}{2^{b_2}}\Big) = 2\frac{(1+1+1+1/2)^c}{4^c} \ll (7/8)^c,$$

where we applied the 4-fold multinomial expansion. □

[42]Here, when using the term "depends" for an expression, we should pedantically be careful to refer to whichever of $\dot{\mathcal{L}}^\star_{jk_1}$ or $\dot{\mathcal{L}}^\star_{k_1j}$ is a basic formal variable (that is, whether $j < k_1$ or not); however, we employ a bit of economy in language.

8.6. Finally we turn to the most difficult case (indeed, the most generic) of Swinnerton-Dyer's non-genericity analysis. Here we will have two elements that have a total three independent co-ordinates, and 8 possibilities for the types of primes. By excluding three formal symbols involving the products of the independent co-ordinatess, we will be able to show that every other formal symbol contributes 3 independent expressions, and at least one of the 8 classes yields 4.

**Lemma 8.6.1.** [50, Lemma 7]. *The proportion of* $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})[c]$ *for which there are two elements in* $U^c_{\mathcal{B}}(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ *of the form* $(u_1, u_2, u_3)$ *and* $(v_1, v_2, v_3)$ *with* $u_1 = v_2$ *is* $\ll (15/16)^c$ *for* $0 \le c < \tilde{r} - 1$.

Smith notes (page 77) the same bound when $u_1 = v_2\xi$ for some $\xi \in Y^0_{\mathcal{B}}$ – the proof follows in the same manner, with some of the apperances of $u$ being $u\xi$ instead.

*Proof.* We write the elements as $(u, v, uv)$ and $(w, u, uw)$.

The most special set $B_1$ of formal symbols are those that divide $u$ but neither $v$ nor $w$ (note that these did not occur in the previous Lemma 8.5.6). The expressions for $(u, v, uv)$ thus come from the second line of (10) and are

$$(\delta_{21}\delta_{23}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, v)_{\dot{p}_j} \text{ and } (\delta_{12}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$$

while those for $(w, u, uw)$ come from the first line and are

$$(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, w)_{\dot{p}_j} \text{ and } (\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}.$$

In particular,[43] the expressions with $(\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}$ can be added to get $(-1, \dot{p}_j)_{\dot{p}_j}$. The other two expressions will be independent of each other since $vw \notin Y^0_{\mathcal{B}}$.

Next, the set of formal symbols $\dot{p}_j$ that divide none of $u, v, w$ yield the expressions

$$(\dot{p}_j, u)_{\dot{p}_j}, \ (\dot{p}_j, v)_{\dot{p}_j}, \text{ and } (\dot{p}_j, w)_{\dot{p}_j}.$$

For each such $j$ these will be independent of each other since $uv, uw, vw, uvw \notin Y^0_{\mathcal{B}}$.

8.6.2. For the rest of the divisibility possibilities, we will take one expression involving $\dot{d}/\dot{p}_j$ and two independent expressions that do not. The cases where $\dot{p}_j$ does not divide $u$ and not both $v$ and $w$ will have expressions from (9).

When $\dot{p}_j$ divides $u$ we can always take the expression with $\dot{d}/\dot{p}_j$ to be of the form

$$(\delta, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}.$$

When $\dot{p}_j$ divides all of $u, v, w$ the other conditions are

$$(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uv)_{\dot{p}_j} \text{ and } (\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uw)_{\dot{p}_j},$$

and when $\dot{p}_j$ divides $u$ and $v$ but not $w$ they are

$$(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uv)_{\dot{p}_j} \text{ and } (\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, w)_{\dot{p}_j},$$

while when $\dot{p}_j$ divides $u$ and $w$ but not $v$ they are

$$(\delta_{21}\delta_{23}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, v)_{\dot{p}_j} \text{ and } (\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uw)_{\dot{p}_j}.$$

When $\dot{p}_j$ doesn't divide $u$, we can always take one of the expressions as

$$(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, u)_{\dot{p}_j}.$$

[43] As with Footnote 38 (and indeed, more dramatic in impact), these extra conditions are trivial when the formal symbols correspond to primes that are restricted to be 1 mod 4, making the notion of genericity more delicate. Similar comments apply to other cases.

When $\dot{p}_j$ divides $v$ but not $w$ we take the other two expressions as

$$(\delta_{12}\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, w)_{\dot{p}_j}, \text{ and } (\delta_{21}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, v)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j},$$

while when it divides $w$ but not $v$ we take

$$(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uw)_{\dot{p}_j}, \text{ and } (\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, w)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}.$$

Finally, when $\dot{p}_j$ divides $v$ and $w$ we consider $(u, v, uv) + (w, u, uw) = (uw, uv, vw)$ and apply the third line of (10) to get

$$(\delta_{31}\delta_{32}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, vw)_{\dot{p}_j} \text{ and } (\delta_{13}, \dot{p}_j)_{\dot{p}_j} + (\dot{p}_j, uw)_{\dot{p}_j} + (\dot{p}_j, \dot{d}/\dot{p}_j)_{\dot{p}_j}.$$

8.6.3. We now turn to analysis of independence of these expressions. Again those that involve $\dot{d}/\dot{p}_j$ are the easiest, as the expressions uniquely involve $\dot{\mathcal{L}}^\star_{jj}$, ensuring they are independent of each other, and of all the others.

We then select three formal symbols that distinguish $u, v, w$. This is possible since $uv, uw, vw, uvw \notin Y_{\mathcal{B}}^0$ (when $uvw = \xi \in Y_{\mathcal{B}}^0$ our elements in $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ are $(u, v, uv)$ and $(w, u, v\xi)$, and Lemma 8.5.6 applies). With $k_1, k_2, k_3$ as the indices of said symbols, the 7 nontrivial products $x$ from $u, v, w$ each have $(\dot{p}_j, x)_{\dot{p}_j}$ depending on a different nontrivial combination of $\dot{\mathcal{L}}^\star_{jk_1}$, $\dot{\mathcal{L}}^\star_{jk_2}$, and $\dot{\mathcal{L}}^\star_{jk_3}$. Thus such expressions will be independent from all the others if no $\dot{\mathcal{L}}^\star_{k_ij}$ appears elsewhere, and indeed we ensure so by excluding the three formal symbols $\dot{p}_{k_1}$, $\dot{p}_{k_2}$, and $\dot{p}_{k_3}$ from consideration in our set of expressions.

We are then left with the expressions $(-1, \dot{p}_j)_{\dot{p}_j}$ from $B_1$, which don't depend on $\dot{\mathcal{L}}$, and are thus independent of the others (and of each other as the $j$ differ).

Writing $b_i$ for the size of $B_i$ (so that their sum is $c$), we conclude there are at least $3\sum_i b_i - 12 + b_1$ independent expressions.

8.6.4. Given two elements of the specified type, the above then gives that the proportion of $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ with both elements in $U_{\mathcal{B}}^c(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ is $\ll 1/2^{3c-12+b_1}$. We can then aggregate over possible choices of $(u, v, w)$, of which there $2^{\#\tilde{\Omega}}\binom{c}{\vec{b}}$ in terms of the vector $\vec{b}$ (of divisibilities) corresponding to $(u, v, w)$. Summing over all such possibilities of $\vec{b}$ then gives a proportion of

$$\ll \frac{1}{2^{3c}} \sum_{\sum_i b_i = c} \binom{c}{\vec{b}} \frac{1}{2^{b_1}} = \frac{(1+1+1+1+1+1+1+1/2)^c}{8^c} = (15/16)^c$$

where we applied the 8-fold multinomial expansion. □

8.7. Finally we make some comments about the Markov chain analysis.

As with §7.3.2, we let

$$\rho_s = \frac{2^s}{\prod_{j=1}^s (2^j - 1)} \prod_{n=0}^{\infty} (1 - 1/2^{2n+1}),$$

for which we have $\rho_0 + \rho_2 + \rho_4 + \cdots = \rho_1 + \rho_3 + \rho_5 + \cdots = 1$, while for $s \geq 0$ we have the recurrence relation (with $\rho_{-1} = \rho_{-2} = 0$)

$$\rho_s = \rho_{s-2} \cdot \frac{1}{2^{2(s-2)+1}} + \rho_s \cdot \Big(\frac{3}{2^s} - \frac{5}{2^{2s+1}}\Big) + \rho_{s+2} \cdot \Big(1 - \frac{3}{2^{s+2}} + \frac{4}{2^{2(s+2)+1}}\Big).$$

Numerically we have $\rho_0 \approx 0.4194224418$ and

$$\Big(\frac{\rho_0}{\rho_0}, \frac{\rho_2}{\rho_0}, \frac{\rho_4}{\rho_0}, \frac{\rho_6}{\rho_0}, \frac{\rho_8}{\rho_0}, \ldots\Big) = \Big(1, \frac{4}{3}, \frac{16}{315}, \frac{64}{615195}, \frac{256}{19923090075}, \ldots\Big),$$

while $\rho_1 = 2\rho_0 \approx 0.8388448836$ and

$$\Big(\frac{\rho_1}{\rho_1}, \frac{\rho_3}{\rho_1}, \frac{\rho_5}{\rho_1}, \frac{\rho_7}{\rho_1}, \frac{\rho_9}{\rho_1}, \dots\Big) = \Big(1, \frac{4}{21}, \frac{16}{9765}, \frac{64}{78129765}, \frac{256}{10180699028325}, \dots\Big).$$

Rather than directly utilizing the $\rho_s$, Smith phrases his result (Corollary 6.11) in terms of the proportion of alternating matrices over $\mathbf{F}_2$ of size $(2m+s)$ with kernel of dimension $s$, and indeed it turns out that this proportion is $\rho_s$ as $m \to \infty$. However, as far as I can tell,[44] this is something that one can only conclude *ex post facto* after computing the Markov stable state, and simply matching the obtained $\rho_s$ with the known limiting proportion of alternating matrices.[45] This is thus unlike the case of 4-ranks of narrow class groups of Gaussian discriminants (no prime factors are 3 mod 4) in §12, where the relevant Rédei matrix that yields the 4-rank is directly a symmetric matrix involving just the $(p_i|p_j)$, with indeed no restriction from any $\widetilde{\mathcal{K}}$-information; so the proportion of $\mathcal{L}$ that give a specific rank of this Rédei matrix is thus the proportion of associated symmetric matrices over $\mathbf{F}_2$. Thereby we avoid the Markov chain completely in this case.[46]

8.7.1. We let $M$ be the infinite matrix (indexed starting at 0) whose nonzero entries are the "transition probabilities" given by

$$M_{s,s+2} = \frac{1}{2^{2s+1}}, \;\; M_{s,s} = \Big(\frac{3}{2^s} - \frac{5}{2^{2s+1}}\Big), \;\text{ and }\; M_{s,s-2} = \Big(1 - \frac{3}{2^s} + \frac{4}{2^{2s+1}}\Big)$$

(the latter for $s \ge 2$), noting the row-sums are all equal to 1. This naturally splits into even/odd numbered rows and columns, and for $s \le 7$ the entries are given by[47]

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 3/8 & 19/32 & 1/32 & 0 \\ 0 & 105/128 & 91/512 & 1/2^9 \\ 0 & 0 & 1953/2^{11} & 379/2^{13} \end{pmatrix}, \begin{pmatrix} 7/8 & 1/8 & 0 & 0 \\ 21/32 & 43/128 & 1/128 & 0 \\ 0 & 465/512 & 187/2^{11} & 1/2^{11} \\ 0 & 0 & 8001/2^{13} & 763/2^{15} \end{pmatrix},$$

where the left/right matrix has the even/odd-indexed entries. We write $M_{\mathrm{e}}$ and $M_{\mathrm{o}}$ for the respective restrictions of $M$ to the even/odd rows and columns, retaining the indexing of these, and write $M_\star$ when a statement applies to both restrictions.

Since $M_\star$ has row-sums of 1 it has 1 as an eigenvalue (with $(1,1,1,\dots)$ as the eigenvector). Thus the transpose $M_\star^{\mathrm{T}}$ also has 1 as an eigenvalue, with the associated eigenvector being the stable vector $\vec{\rho}_\star$.

---

[44] Smith is terse, but I think he directly re-interprets Swinnerton-Dyer's $P(d,M)$ in [50, (21)] as this proportion. However, this $P(d,M)$ is for the specific alternating matrices induced by congruence and Legendre symbol conditions, and it is unobvious to me (though true) they are proportionally rank-distributed in the whole. In any case, Smith's ultimate statement is correct.

[45] Said proportion, which is related to Delaunay's elliptic curve adaptation [5] of the (number fields) heuristic of Cohen and Lenstra, is discussed more by Bhargava, Kane, Lenstra, Poonen, and Rains [2, §1.4, §3.6ff].

[46] On the other hand, Gerth's analysis [12] of the distribution of 4-ranks of narrow quadratic class groups in the general case was similar to our current situation, and indeed computing the actual proportions of ranks for (*e. g.*) matrices of size $m$ over $\mathbf{F}_2$ with an anti-symmetric corner of size roughly $m/2$ with the rest being symmetric (this corresponds to quadratic reciprocity) was done by a Markov analysis (moreover, more extensive than the one here); the fact that said proportions were asymptotically the same as those from selecting a random $\mathbf{F}_2$ matrix (with no conditions whatsoever) again seemed to appear in a rather *ex post facto* manner.

[47] From the columns of the $M_\star$ one finds the ratios in the stable distributions, for instance $\rho_1 = (7/8)\rho_1 + (21/32)\rho_3$ so that $\rho_3 = (4/21)\rho_1$, and $\rho_3(1 - 43/128) = (1/8)\rho_1 + (465/512)\rho_5$ so that $\rho_5 = (512/465)\rho_1\big[(85/128)(4/21) - (1/8)\big] = (16/9765)\rho_1$, etc.

8.7.2. Let $\vec{w}_\star^c$ be the vector whose $s$th component gives the proportion of elements $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L}) \in \mathcal{Y}(\tilde{r}, \#\tilde{\Omega})$ whose $c$-restriction is generic (as defined in §8.4.3) with $s_c = s$. We can rephrase Lemma 8.4.4 as saying that the $s$th component of $M_\star \vec{w}_\star^c$ is the proportion of such elements with $s_{c+1} = s$.

Meanwhile, if we write $\vec{\epsilon}_\star^c$ for the similar vector containing the proportion of elements whose $c$-restriction is nongeneric with a given $s_c$, there is a transition matrix $W_\star^c$ (with at most three nonzero entries per row and column) with row-sums of 1 such that the entries of $W_\star^c \vec{\epsilon}_\star^c$ have the proportion of such elements with $s_{c+1} = s$. By the combination of Lemmata 8.5.3, 8.5.4, 8.5.5, 8.5.6, and 8.6.1, we see the 1-norm of the nongeneric $\vec{\epsilon}_\star^c$ is $\ll (15/16)^c$ for $E$ with no rational 4-torsion point (only Lemma 8.5.4 uses this torsion condition). The following Lemma shows such errors do not accumulate when $M_\star$ and $W_\star^c$ are applied.

**Lemma 8.7.3.** *Suppose that the distribution of 2-Selmer estimations at $J$ is given by $\vec{h}_\star^J$ (including $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ that are nongeneric). Then the distribution of 2-Selmer estimations at $(\tilde{r} - 1)$ is given by $M_\star^{\tilde{r}-1-J}\vec{h}_\star^J + \vec{\beta}_\star$ for some $\vec{\beta}_\star$ whose 1-norm is $\ll (15/16)^J$ when $E$ has no rational 4-torsion point.*

*Proof.* We have that $\vec{h}_\star^{c+1} = M_\star(\vec{h}_\star^c - \vec{\epsilon}_\star^c) + W_\star^c \vec{\epsilon}_\star^c$ for each $c$, and by iteratively applying this for $J \le c < \tilde{r} - 1$ we find that

$$\vec{h}_\star^{\tilde{r}-1} = M_\star^{\tilde{r}-1-J}\vec{h}_\star^J - \sum_{u=J}^{\tilde{r}-2} M_\star^{\tilde{r}-1-u}\vec{\epsilon}_\star^u + \sum_{u=J}^{\tilde{r}-2} M_\star^{\tilde{r}-2-u} W_\star^u \vec{\epsilon}_\star^u.$$

Since $M_\star$ and the $W_\star^u$ have row-sums of 1 while the 1-norm of $\vec{\epsilon}_\star^u$ is $\ll (15/16)^u$, the latter two sums give an error whose 1-norm is also thus bounded (the dominant term comes from $u = J$). □

8.8. This previous Lemma 8.7.3 shows that the distribution of 2-Selmer estimations $s_{\tilde{r}-1}$ (and thus the 2-Selmer ranks themselves) has a main term of $M_\star^{\tilde{r}-1-J}\vec{h}_\star^J$, and thus we are in a situation where Markov chain analysis can be applied. An alternative (and largely equivalent) approach to the situation is to compute the rate of convergence to the dominant eigenvector (namely $\vec{\rho}_\star$) of the transpose $M_\star^{\mathrm{T}}$. We sketch the ideas here, being somewhat more hands-on than applying a black box.

8.8.1. The desired plan would then be to write $M_\star = T^{-1}DT$ where $D$ is diagonal; thus its entries are the eigenvalues, and indeed $T$ can be taken as the matrix whose columns are the eigenvectors of $M_\star$ (normalized in whatever manner). Then we have $M_\star^u = T^{-1}D^uT$, and if 1 is indeed the dominant eigenvalue we then see that $D^u$ tends rapidly to a matrix with one nonzero entry. Upon suitably bounding the entries of $T$ and $T^{-1}$ this then gives a bound on how far the entries of $M_\star^{\tilde{r}-1-J}\vec{h}_\star^J$ will be from those of $\vec{\rho}_\star$.

It is fairly easy to show that the second largest eigenvalue tends to $1/4$ (and the third to $1/16$, etc.) as the size of the matrix becomes large. However, controlling the size of the entries of $T^{-1}$ (for instance) seems rather nontrivial. Moreover, at some point we need to use some notion of smallness of $J$ (so in particular the entries of $\vec{h}_\star^J$ for $s \ge 2J + O(1)$ are nonzero) with respect to $\tilde{r}$.

Let us take $J = \lfloor (\log\log X)/99 \rfloor$ as a specific value.

8.8.2. We then proceed by truncating $M_\star$ to a finite size (dependent on $X$). First we note that $M_\star^{2J}\vec{h}_\star^J$ will have almost all of its 1-norm in the components up to $U$ for $U = \lfloor\sqrt{\log\log X}/9\rfloor$. This follows since the transition entry for $s \to s-2$ is $\ge 1-3/2^s$, and so multiplying up to $U$ gives roughly $1/2^{U^2/4}$ decay, thus saving a small power of $(\log X)$.

We let $M_\star^{[U]}$ be the resulting truncated transition matrix; this is not quite simply the upper-left $U$-corner of $M_\star$, but also modifies the ultimate diagonal entry to ensure the row sums are still 1. The truncated matrix thus retains 1 as an eigenvalue, though the eigenvector $\vec{\rho}_\star^{[U]}$ is no longer quite $\vec{\rho}_\star$. However, the difference between these is small – indeed, as in Footnote 47, for $s < U$ the ratio of $s$th component to the 0th component (in the even parity case) is still the same, and the final ratio changes negligibly. Rescaling to have unit 1-norm introduces another (negligible) error of size $\ll 1/(\log X)^c$ in each entry of $\vec{\rho}_\star^{[U]}$.

We let $(\vec{h}_\star^{[U]})^{3J}$ be the truncation of $M_\star^{2J}\vec{h}_\star^J$ to the components up to $U$.

The second eigenvalue $\lambda_2$ can be shown to tend to $1/4$ (from below) and indeed the characteristic polynomial tends to $(x-1)(x-1/4)(x-1/16)(x-1/64)(\cdots)$ as $U \to \infty$. One way to handle the error analysis with the size of the entries in the resulting matrix of eigenvectors is to first balance[48] the matrix $M_\star^{[U]}$ by transforming it by a diagonal matrix $B_\star$ so that $S_\star = B_\star^{-1}M_\star^{[U]}B_\star$ is symmetric. Then (as was known to Jacobi) the eigenvector matrix $V$ for $S_\star$ can be taken to have orthonormal columns, so that the inverse is simply the transpose.

For our tridiagonal situation, in the even parity case (the odd case is similar) the balancing matrix will have entries

$$(B_{\mathrm{e}})_{nn} = \prod_{i=1}^{n/2-1}\sqrt{(M_{\mathrm{e}})_{2i,2i-2}/(M_{\mathrm{e}})_{2i-2,2i}}$$

for $n \ge 2$, with $(B_{\mathrm{e}})_{00} = 1$. Thus the size of the $n$th diagonal entry is roughly $2^{n^2}$. We then have that

$$(M_\star^{[U]})^l = (B_\star S_\star B_\star^{-1})^l = (B_\star^{-1}VDV^{-1}B_\star)^l = B_\star^{-1}V^{-1}D^lVB_\star$$

where $D$ is the matrix of eigenvalues. The size bound on the entries of $B_\star$ then implies that $\big((M_\star^{[U]})^l(\vec{h}_\star^{[U]})^{3J} - \vec{\rho}_\star^{[U]}\big)$ has entries bounded as $\ll U^3 2^{U^2}\lambda_2^l$.

Note $(S_\star)_{nn} > (S_\star)_{n,n-2}$, and it follows $S_\star$ is positive definite; so its eigenvalues are nonnegative, and their sum is bounded by the diagonal-sum $\le 4/3$, so $\lambda_2 \le 1/3$.

The above then gives the following Lemma (which is still somewhat of a sketch).

**Lemma 8.8.3.** *With $J = \lfloor(\log\log X)/99\rfloor$ and $\tilde{r} \ge (98/99)\log\log X$, the distribution of 2-Selmer estimations at $(\tilde{r}-1)$ is given by $\vec{\rho}_\star + \vec{\beta}_\star$ for some $\vec{\beta}_\star$ whose 1-norm is $\ll (15/16)^J + 1/(\log X)^{1/500}$ when $E$ has no rational 4-torsion point.*

*Proof.* Lemma 8.7.3 reduces the situation to considering $M_\star^{\tilde{r}-1-J}\vec{h}_\star^J$, and our above truncation process shows that with $U = \lfloor\sqrt{\log\log X}/9\rfloor$ this is sufficiently close to $(M_\star^{[U]})^{\tilde{r}-1-3J}(\vec{h}_\star^{[U]})^{3J}$.

The entries of $\big((M_\star^{[U]})^j(\vec{h}_\star^{[U]})^{3J} - \vec{\rho}_\star^{[U]}\big)$ are then bounded as $\ll U^3 2^{U^2}\lambda_2^j$, and applying this with $j = \tilde{r}-1-3J \ge -1 + (95/99)\log\log X$ and $U \le \sqrt{\log\log X}/9$ gives the result (recalling $\lambda_2 \le 1/3$), since $\vec{\rho}_\star^{[U]}$ differs from $\vec{\rho}_\star$ negligibly. □

[48]This is a case of Parlett-Reinsch balancing [37] from the context of computing eigenvalues.

A more detailed approach to the above Markov error analysis has been given by Koymans and Pagano [28], in the case of 4-ranks of quadratic class groups.

8.9. Finally, we want to interpret this distribution of $\rho_s$-values for $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ in terms of boxes. We thus replicate the sketched computation from the end of §6.4.

**Proposition 8.9.1.** *Suppose $\bar{T}$ is a $(\kappa_0, \eta_1, \eta_{\mathrm{s}})$-pleasant $(X, \eta_0, \mathcal{P})$-box, and assume that $\eta_0/2 > \eta_1 > \eta_{\mathrm{s}} > 0$. Also assume that $E$ has no rational 4-torsion point. Then*

$$\#\{d : d \in \hat{T}^{\pm} \mid s_{\tilde{r}}(E_d) = s + 2\} = \#T \cdot \rho_s + O\Big(\frac{\#T}{(\log\log X)^u}\Big)$$

*for $u = 1/2 - \kappa_0\eta_0 \log\sqrt{2} - \eta_1$.*

*Proof.* For a pleasant box $T$ (defined as in §4.5), from Proposition 6.3.1 we have

$$\#\{d : d \in \hat{T}^{\pm} \mid s_{\tilde{r}}(E_d) = s\} = \sum_{\substack{(\varepsilon,\mathcal{K},\mathcal{L}) \\ s^{\varepsilon}_{\tilde{r}}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) = \sum_{\substack{(\varepsilon,\mathcal{K},\mathcal{L}) \\ s^{\varepsilon}_{\tilde{r}}(\mathcal{K},\mathcal{L})=s}} \frac{\#T}{2^{\binom{\tilde{r}}{2}}\xi^{\tilde{r}}_{\mathcal{P}}} + O\Big(\frac{\#T \cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big),$$

where $k_0 \le \kappa_0\eta_0 \log\log\log X$ and $k_1 \le 3(\log\log X)^{\eta_1}$ (by pleasantness).

By Lemma 8.8.3 and the equal split in parity amongst $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, the proportion of $(\varepsilon, \mathcal{K}, \mathcal{L})$ with $s_{\tilde{r}}$ of $(s+2)$ is $\rho_s/2 + O\big((15/16)^J\big)$ where $J = \lfloor(\log\log X)/99\rfloor$. Noting that there are $2 \cdot 2^{\binom{\tilde{r}}{2}}\xi^{\tilde{r}}_{\mathcal{P}}$ choices of $(\varepsilon, \mathcal{K}, \mathcal{L})$, we thus conclude

$$\#\{d : d \in \hat{T}^{\pm} \mid s_{\tilde{r}}(E_d) = s + 2\} = \#T \cdot \big[\rho_s + O\big((15/16)^J\big)\big] + O\Big(\frac{\#T \cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big),$$

and substituting for $J$ and using the above bounds for $k_0$ and $k_1$ gives the result. □

We should stress that this result, which uses the box machinery of §4 and the box-splitting of §§5-6, is what allows the passage from the "unnatural" ordering to the normal one, as discussed in §1.3.1.

## 9. Recapitulation

Let us put together the various parts of our argument, and optimize the error.

We assume the parameters satisfy $0 < \eta_{\mathrm{s}} < \eta_1 < \eta_0/2 < 1/100$ and also that $E$ has no rational 4-torsion point.

9.1. From Section 4, we know almost all $\tilde{d}$ are represented by pleasant boxes. Indeed, by Lemma 4.5.1 (in our case of $\alpha_{\mathcal{P}} = 1$) the exceptional subset of $\tilde{d} \in S^{\mathcal{P}}(X)$ that are not represented by a $(\kappa_0, \eta_1, \eta_{\mathrm{s}})$-pleasant $(X, \eta_0, \mathcal{P})$-box has size

$$\ll \frac{X}{(\log\log X)^{\eta_0\kappa_0(\log\kappa_0 - 1 - \log(100/99))}} + \frac{X}{(\log\log X)^{99}}.$$

Also, every positive squarefree $\tilde{d}$ that is coprime to $\Omega$ is in at most one such box.

Then in §§5-6 we showed results about the size of boxes when split up by Legendre symbol conditions on the primes involved. In particular, for a pleasant box $\bar{T}$ we found that

$$\#\{d : d \in \hat{T}^{\pm} \mid s_{\tilde{r}}(E_d) = s\} = \sum_{\substack{(\varepsilon,\mathcal{K},\mathcal{L}) \\ s^{\varepsilon}_{\tilde{r}}(\mathcal{K},\mathcal{L})=s}} \#T(\mathcal{K},\mathcal{L}) = \sum_{\substack{(\varepsilon,\mathcal{K},\mathcal{L}) \\ s^{\varepsilon}_{\tilde{r}}(\mathcal{K},\mathcal{L})=s}} \frac{\#T}{2^{\binom{\tilde{r}}{2}}\xi^{\tilde{r}}_{\mathcal{P}}} + O\Big(\frac{\#T \cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big),$$

where $k_0 \le \kappa_0\eta_0 \log\log\log X$ and $k_1 \le 3(\log\log X)^{\eta_1}$ (by pleasantness). This used the bounds on the $\eta$, but not the 4-torsion assumption.

Finally, in §§7-8 we showed (using the no 4-torsion assumption to handle non-genericity) that the proportion of $(\varepsilon, \mathcal{K}, \mathcal{L})$ with $s_{\tilde{r}}$ of $(s+2)$ is $\rho_s/2 + O\big((15/16)^J\big)$ where $J = \lfloor(\log\log X)/99\rfloor$, so that for any pleasant box $\bar{T}$ we have

$$\#\{d : d \in \hat{T}^{\pm} \mid s_{\tilde{r}}(E_d) = s+2\} = \#T \cdot \big[\rho_s + O\big((15/16)^J\big)\big] + O\Big(\frac{\#T \cdot 2^{k_0/2} k_1}{\sqrt{\log\log X}}\Big).$$

Summing over pleasant boxes, we thereby conclude that

$$\frac{\#\{|d| \le X : \mu(d) \ne 0,\ \gcd(d,\Omega) = 1 \mid s(E_d) = s+2\}}{\#\{|d| \le X : \mu(d) \ne 0,\ \gcd(d,\Omega) = 1\}} = \frac{\rho_s}{2} + O\Big(\frac{1}{(\log\log X)^{\min(u,v)}}\Big)$$

where $v = \eta_0\kappa_0(\log\kappa_0 - 1 - \log(100/99))$ and $u = 1/2 - \kappa_0\eta_0\log\sqrt{2} - \eta_1$.

9.1.1. It remains to optimize this exponent in the error. The choice of $\kappa_0 > 3$ is still available to us, and we shall take it so that $\kappa_0\eta_0 = 1/\sqrt{\log(1/\eta_0)}$. In particular, we then have

$$u = 1/2 - \kappa_0\eta_0\log\sqrt{2} - \eta_1 = 1/2 - \frac{\log\sqrt{2}}{\sqrt{\log(1/\eta_0)}} - \eta_1 \to 1/2$$

as $\eta_0 \to 0$ (so that $\eta_1 \to 0$ also), while

$$v = \eta_0\kappa_0\big(\log\kappa_0 - 1 - \log(100/99)\big) = \frac{\log(1/\eta_0\sqrt{\log(1/\eta_0)}) - \log(100e/99)}{\sqrt{\log(1/\eta_0)}}$$
$$= \frac{\log(1/\eta_0) - (1/2)\log\log(1/\eta_0) - \log(100e/99)}{\sqrt{\log(1/\eta_0)}} \to \infty$$

as $\eta_0 \to 0$. Thus $\min(u,v) \to 1/2$, and this then shows the stated Theorem 1.1.2.

## 10. The 4-rank of (narrow) class groups of quadratic fields

A related topic to 2-Selmer ranks of quadratic twists of elliptic curves is the 4-rank of the narrow class group of quadratic fields. One key similarity is that an analysis of the equi-distribution of $(p_i|p_j)$ will play a significant rôle.

10.1. The subject of 4-ranks (and higher 2-power ranks) for narrow class groups of quadratic fields was studied in a series of papers (such as [38, 39]) in the 1930s by Rédei, starting with a joint paper with Reichardt [41].

The most relevant observation for our purposes is that the 4-rank can be written in terms of the dimension of the kernel of the matrix of Legendre symbols (sometimes called the Rédei matrix, or perhaps the Legendre matrix). We write $(a|b)^\star$ for the value of a nonzero Legendre symbol $(a|b)$ when mapped from $\{\pm 1\}$ to $\mathbf{F}_2$.

Let $d = \varepsilon\prod_i p_i$ be a fundamental discriminant with $r$ prime divisors, and write $d = \prod \hat{p}_i$ for its factorization into prime-power discriminants (allowing $-4$ as such, with a slight abuse of notation in that $\hat{2} \in \{-4, -8, 8\}$ is not uniquely defined). Write $R^d$ for the $r$-by-$r$ matrix over $\mathbf{F}_2$ defined by $(\hat{p}_i|p_j)^\star$ for $i \ne j$, and $R^d_{jj} = \sum_{i\ne j} R^d_{ij}$. The 4-rank $e(d)$ of the narrow class group of $\mathbf{Q}(\sqrt{d})$ is then equal to one less than the dimension of the kernel of the Rédei matrix $R^d$, so that $e(d) = \dim(\ker R^d) - 1$.

(This also holds for non-fundamental discriminants, which we do not consider).

The column-sums are zero by construction, and when $d < 0$ the row-sums are all zero by quadratic reciprocity. When $d > 0$ a row-sum is zero when it corresponds to a prime that is 1 mod 4, and is nonzero when it corresponds to a prime that is 3 mod 4; meanwhile, for $\hat{2} = 8$ it is zero, while nonzero for $\hat{2} \in \{-4, -8\}$.

Stevenhagen has written a couple of modern resources [47, 48] on Rédei matrices (and Rédei symbols for 8-ranks).

10.1.1. The special case of "Gaussian" discriminants (often just called "special" discriminants), where $d > 0$ and no prime dividing $d$ is 3 mod 4, yields a symmetric Rédei matrix, and thus can be more readily amenable to further study.[49] Indeed, Rédei [39] claimed to have obtained the distribution of the 4-ranks for Gaussian discriminants when ordering by number of prime factors, though as Gerth and Graham [13] point out, the error analysis is spotty.[50]

Also, the ordinary class group will be the same as the narrow class group except possibly for Gaussian discriminants, and indeed the occurrences of when these differ are then related to the question of the solubility of the negative Pell equation.

10.2. Gerth [12] considered 4-ranks of the class group for general quadratic fields under the ordering by number of prime factors.[51] He was able to determine the 4-rank distribution (in the sense of giving an algorithm to compute it in time polynomial in $r$) for any fixed number $r$ of prime factors, and then gave a somewhat complicated Markov analysis to determine the limiting behavior as $r \to \infty$.

Gerth gets (under his ordering) that the proportion of imaginary quadratic fields with 4-rank equal to $e$ is

$$\gamma_{\mathrm{I}}(e) = \frac{1}{2^{e^2}} \prod_{u=1}^{\infty}(1 - 1/2^u) \prod_{u=1}^{e}(1 - 1/2^u)^{-1} \prod_{u=1}^{e}(1 - 1/2^u)^{-1},$$

while in the case of real quadratic fields the proportion is

$$\gamma_{\mathrm{R}}(e) = \frac{1}{2^{e(e+1)}} \prod_{u=1}^{\infty}(1 - 1/2^u) \prod_{u=1}^{e}(1 - 1/2^u)^{-1} \prod_{u=1}^{e+1}(1 - 1/2^u)^{-1}.$$

Gerth notes that these turn out to be related to the heuristic predictions [4] of Cohen and Lenstra (who only handled odd primes) that one might expect at $p = 2$, namely that the 4-rank considers the 2-rank of the square of the class group, and thus difficulties with genus theory are obviated.

10.2.1. Another interpretation of these proportions is in terms of ranks of random matrices over $\mathbf{F}_2$. The imaginary quadratic case corresponds to the limiting probability (as $r \to \infty$) that a random matrix of size $(r-1)$-by-$(r-1)$ has a kernel of dimension $e$, while the real quadratic case has the matrices of size $r$-by-$(r-1)$. (See Landsberg [32] for counting random matrices with a given kernel dimension).

[49] There are other cases where one can force the Rédei matrix to be symmetric, such as $d < 0$ with exactly one prime factor that is 3 mod 4. I don't know if anyone has tried to describe the 4-rank distribution for such a situation. (A bound for the 8-rank is given by Lu [33]).

[50] Stevenhagen also notes [46, Proposition 2.5] that Rédei's proof does not adequately handle the error term (and it was 50 years until this was remedied), though he erroneously claims the proof was in [40] (and moreover habitually gives the wrong page numbers in his citations of [39]).

[51] Note that this allowed him to ignore even discriminants, as the number of odd parts $d' = d/8$ with $d' \leq X/8$ having $(r-1)$ prime factors is asymptotically negligible compared to the number of $d$ with $r$ prime factors – assuming (of course) that $r$ is fixed while $X \to \infty$, as per the ordering.

Meanwhile, the proportion of Gaussian discriminants with 4-rank equal to $e$ is

$$\gamma_{\mathrm{G}}(e) = \frac{1}{2^{e(e+1)/2}} \prod_{u=1}^{\infty} (1+1/2^u)^{-1} \prod_{u=1}^{e} (1-1/2^u)^{-1},$$

and this corresponds to the limiting probability that a symmetric matrix over $\mathbf{F}_2$ has a kernel of dimension $e$. Note that large $e$ occur much more readily here, as indeed the symmetric nature of the Rédei matrix ensures that the entries are not as independent as in the general case (cf. the discussion of [9, §1.3]).

We can indeed re-interpret the 2-Selmer case for quadratic twists of elliptic curves to have proportions (having fixed the parity) of 2-ranks equal to $s$ as

$$\rho_s = \frac{1}{2^{s(s-1)/2}} \prod_{u=1}^{\infty} (1+1/2^u)^{-1} \prod_{u=1}^{s} (1-1/2^u)^{-1},$$

and this corresponds to the limiting probability that an alternating matrix over $\mathbf{F}_2$ (whose dimension is a given parity) has a kernel of dimension $s$.

In Table 8 we list (approximations to) the above proportions for $e \le 5$.

| | $e=0$ | $e=1$ | $e=2$ | $e=3$ | $e=4$ | $e=5$ |
|---|---|---|---|---|---|---|
| $\gamma_{\mathrm{I}}(e)$ | 0.288788 | 0.577576 | 0.128350 | 0.005239 | 0.000047 | $9.69 \cdot 10^{-8}$ |
| $\gamma_{\mathrm{R}}(e)$ | 0.577576 | 0.385051 | 0.036672 | 0.000699 | 0.000003 | $3.08 \cdot 10^{-9}$ |
| $\gamma_{\mathrm{G}}(e)$ | 0.419422 | 0.419422 | 0.139807 | 0.019972 | 0.001331 | 0.000043 |
| $\rho_s/2$ | 0.209711 | 0.419422 | 0.279615 | 0.079890 | 0.010652 | 0.000687 |

TABLE 8. Asymptotic proportions of kernel dimensions in various cases

10.3. Fouvry and Klüners [8] were then able to adapt Heath-Brown's analysis (from 2-Selmer ranks) of equi-distribution of $(p_i|p_j)$ to the case of general quadratic (fundamental) discriminants. They obtain the main term for the moments[52] by an analysis of unlinked indices; one aspect of their work is that they indeed interpret this main contribution in terms of the heuristic of Cohen and Lenstra.[53]

As with Heath-Brown's work on the 2-Selmer group, the error term was not given too explicitly (for the distribution), and was ineffective (already for the moments).

10.3.1. Fouvry and Klüners [9, Corollary 2] then considered the narrow 4-rank distribution for Gaussian discriminants, though almost in an *en passant* sense, as it was part of their articulation of a more profound analysis that allowed them to handle both the narrow and the ordinary class group (ultimately giving nontrivial bounds on the frequency of solubility of the negative Pell equation).

The preprint [3] of Chan, Koymans, Milovic, and Pagano then notes that Smith's methods can be used for the case of Gaussian discriminants, in particular replicating the above result of Fouvry and Klüners for the 4-rank (and moreover then utilizing the 8-rank to improve the lower bound on the negative Pell solubility frequency).

[52]It is only in [7] that they pass from the moments to the 4-rank distribution, wherein they note that one needs (in general) a suitable growth bound for this to be unique.

[53]To the best of my knowledge, no one has gone back and tried to interpret Heath-Brown's "linked indices" analysis in terms of Delaunay's elliptic curve adaptation [5] of this heuristic.

10.4. Our work here is largely to re-prove, using Smith's methods, the aforementioned results of Fouvry and Klüners for narrow 4-rank distribution, both for Gaussian discriminants and the general case. We write $\mathcal{F}^{\pm}$ for respectively the sets of positive and negative fundamental discriminants, and $\mathcal{G}$ for the set of Gaussian discriminants (which by our convention are always fundamental).

We shall show the following results, where $e(d)$ is the 4-rank of the narrow class group of $\mathbf{Q}(\sqrt{d})$, and the $\gamma_\star(e)$ are defined as above.

**Theorem 10.4.1.** *For any $\omega < 1/2$ we have*

$$\frac{\#\{d \le X : -d \in \mathcal{F}^- \mid e(-d) = e\}}{\#\{d \le X : -d \in \mathcal{F}^-\}} = \gamma_{\mathrm{I}}(e) + O_\omega\Big(\frac{1}{(\log\log X)^\omega}\Big).$$

**Theorem 10.4.2.** *For any $\omega < 1/2$ we have*

$$\frac{\#\{d \le X : d \in \mathcal{F}^+ \mid e(d) = e\}}{\#\{d \le X : d \in \mathcal{F}^+\}} = \gamma_{\mathrm{R}}(e) + O_\omega\Big(\frac{1}{(\log\log X)^\omega}\Big).$$

**Theorem 10.4.3.** *For any $\omega < 1/2$ we have*

$$\frac{\#\{d \le X : d \in \mathcal{G} \mid e(d) = e\}}{\#\{d \le X : d \in \mathcal{G}\}} = \gamma_{\mathrm{G}}(e) + O_\omega\Big(\frac{1}{(\log\log X)^\omega}\Big).$$

10.4.4. Smith's permutation idea with box-splitting fairly readily reduces the situation to a calculation of what proportion of the $(\mathcal{K}, \mathcal{L})$ have a given dimension of the kernel of the associated Rédei matrix.

The main term in the Gaussian case follows without any unwieldy Markov analysis, as we only have $\mathcal{L}$-conditions, and they exactly require a symmetric matrix over $\mathbf{F}_2$, so the distribution therein falls out almost immediately.

For the general case I do not see any better method to calculate said proportions (in the limit as $r \to \infty$) than Gerth's Markov chain technique; thus we only sketch the main milestones therein.

10.4.5. We will try to conserve relevant notation with §2.1, and will use the basic results from §3 and §4 regarding squarefree integers and boxes, and also Smith's box-splitting from §5 and permutation idea in §6. We will not use anything from the later sections on the 2-Selmer group.

## 11. The general cases of 4-ranks

First we discuss the case of 4-ranks of class groups for general quadratic fields. This naturally splits into two cases for real/imaginary fields, and each of these into three subcases for the 2-valuation of the fundamental discriminant. The latter aspect ultimately has no effect on the 4-rank distribution.

11.1. We take $\mathcal{P}$ to be the set of odd primes, with the modulus $M_{\mathcal{P}}$ as 8. In order to normalize the notation between the cases of even and odd fundamental discriminants, we write $\tilde{d} = |d|/2^f$ where $\tilde{d}$ is odd and $f \in \{0, 2, 3\}$. We then write $\hat{p}_0 \in \{1, -4, -8, 8\}$ and $p_0 \in \{1, 4, 8\}$ so that $d = \prod_i \hat{p}_i$ and $|d| = p_0\tilde{d}$. We again write $\tilde{r}$ for the number of prime divisors of $\tilde{d}$.

We consider each of the possibilities for $f$ and the sign of $d$ separately. As with §2.2, we take parameters $\eta_0/2 > \eta_1 > \eta_{\mathrm{s}} > 0$ and $\kappa_0 > 3$, and will eventually take $\kappa_0\eta_0 = 1/\sqrt{\log(1/\eta_0)}$ and $\eta_0 \to 0$ as in §9 to minimize the error term. From

Lemma 4.5.1, we know that the exceptional set of $\tilde{d}$ that are not represented by a $(\kappa_0, \eta_1, \eta_s)$-pleasant $(X/2^f, \eta_0, \mathcal{P})$-box has size bounded as

$$\ll \frac{X}{(\log\log X)^{\eta_0\kappa_0(\log\kappa_0 - 1 - \log(100/99))}} + \frac{X}{(\log\log X)^{99}}.$$

We let $\mathcal{K}$ specify each divisor $p_i$ of $\tilde{d}$ modulo 8.[54] We let $\mathcal{L}$ be a specification of Legendre symbols so that $\mathcal{L}_{ij} = (p_i|p_j)$ for $1 \le i < j \le \tilde{r}$. As for the 4-rank of the narrow class group of $\mathbf{Q}(\sqrt{d})$ in terms of this, we need be no more specific at this point than to say that it is determined by the tuple $(f, \mathrm{sgn}(d), \mathcal{K}, \mathcal{L})$, and in particular we write it as $e_\varepsilon^f(\mathcal{K}, \mathcal{L})$.

Note also that we may wish to restrict $\mathcal{K}$ to ensure $d$ is a fundamental discriminant; for instance, when $(f, \varepsilon) = (0, -)$ we wish for $\tilde{d}$ to be 3 mod 4, so an odd number of prime divisors (thus $\mathcal{K}$-specifications) should be 3 mod 4. This causes no difficulties in the analysis, as we just sum over half the $\mathcal{K}$ instead.

Let $\bar{T}$ be a pleasant box. As with §6.4 we have

$$\tilde{r}! \sum_{\substack{(\mathcal{K},\mathcal{L}) \\ e_\varepsilon^f(\mathcal{K},\mathcal{L})=e}} \#T(\mathcal{K},\mathcal{L}) = \sum_{\substack{(\mathcal{K},\mathcal{L}) \\ e_\varepsilon^f(\mathcal{K},\mathcal{L})=e}} \sum_{\sigma\in\mathrm{Sym}_r} \#T(\mathcal{K}^\sigma,\mathcal{L}^\sigma) = \sum_{\substack{(\mathcal{K},\mathcal{L}) \\ e_\varepsilon^f(\mathcal{K},\mathcal{L})=e}} \tilde{r}! \frac{\#T}{2^{\binom{r}{2}}\xi_{\mathcal{P}}^r} + O\Big(\frac{\tilde{r}!\#T\cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big),$$

where $k_0 \le \kappa_0\eta_0 \log\log\log X$ and $k_1 \le 3(\log\log X)^{\eta_1}$. By Gerth's analysis (which we briefly outline below) we have an asymptotic (as $\tilde{r} \to \infty$) for the number of $(f, \varepsilon, \mathcal{K}, \mathcal{L})$ with narrow 4-rank $e$. This turns out to be independent of $f$, but not of $\varepsilon$, tending to $\gamma_{\mathrm{I}}(e)$ in the imaginary case and $\gamma_{\mathrm{R}}(e)$ in the real quadratic case. Although Gerth does not explicitly give an error bound, as with the Markov analysis in §8.7 it can be shown to be $O\big(1/(\log X)^c\big)$ – this has recently been carried out explicitly by Koymans and Pagano in [28].

Thus for a pleasant box $\bar{T}$ we have

$$\#\{\tilde{d} \in \hat{T} : e_\varepsilon^f(\tilde{d}) = e\} = \sum_{\substack{(\mathcal{K},\mathcal{L})\in\mathcal{D}(\tilde{r},\mathcal{P}) \\ e_\varepsilon^f(\mathcal{K},\mathcal{L})=e}} \#T(\mathcal{K},\mathcal{L}) = \frac{\gamma_\varepsilon(e)}{\delta(f,\varepsilon)} \cdot \#T + O\Big(\frac{\#T\cdot 2^{k_0/2}k_1}{\sqrt{\log\log X}}\Big)$$

where $\delta(f, \varepsilon)$ is 1 for $(f, \varepsilon) \in \{(3, +), (3, -)\}$ and 2 for $(f, \varepsilon) \in \{(0, \pm), (2, \pm)\}$ (the latter to account for fundamental discriminants). Summing over all such pleasant boxes then leads to same minimization problem for the exponent of $(\log\log X)$ in the error term as in §9, and so we conclude both Theorems 10.4.2 and 10.4.1.

11.2. Now we turn to a brief synopsis of Gerth's method to compute the relevant proportions of $(f, \epsilon, \mathcal{K}, \mathcal{L})$ with a given 4-rank $e$.

[54] We can take the modulus to be 4 when $f = 0$, and also in the imaginary case (where we can exclude the row and column associated to 2 in the Rédei matrix).

11.2.1. Let us first consider the case where $f = 0$ and the sign of $d$ is negative. We recall that the Rédei matrix is defined as

$$\begin{pmatrix} R^d_{11} & (\hat{p}_1|p_2)^\star & \cdots & (\hat{p}_1|p_i)^\star & (\hat{p}_1|p_j)^\star & \cdots & (\hat{p}_1|p_r)^\star \\ (\hat{p}_2|p_1)^\star & R^d_{22} & \cdots & (\hat{p}_2|p_i)^\star & (\hat{p}_2|p_j)^\star & \cdots & (\hat{p}_2|p_r)^\star \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (\hat{p}_i|p_1)^\star & \cdots & \cdots & R^d_{ii} & (\hat{p}_i|p_j)^\star & \cdots & (\hat{p}_i|p_r)^\star \\ (\hat{p}_j|p_1)^\star & \cdots & \cdots & (\hat{p}_j|p_i)^\star & R^d_{jj} & \cdots & (\hat{p}_j|p_r)^\star \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (\hat{p}_r|p_1)^\star & (\hat{p}_r|p_2)^\star & \cdots & (\hat{p}_r|p_i)^\star & (\hat{p}_r|p_j)^\star & \cdots & R^d_{rr} \end{pmatrix}$$

where the diagonal entries are taken to be the sum of the other entries in the column. By quadratic reciprocity the row-sums are also zero. Thus we can remove any one row and any one column from the matrix and its rank will remain the same. Also, we can permute the rows and columns without changing the rank. Finally, the matrix is anti-symmetric for primes that are 3 mod 4 in that $R^d_{ij} \neq R^d_{ji}$ for $i \neq j$ that both correspond to such a prime, and else is symmetric, with $R^d_{ij} = R^d_{ji}$ for all other pairs $(i, j)$.

Gerth then considers the anti-symmetric submatrix of $R^d$ corresponding to the rows/columns that are 3 mod 4. As a technical measure (which he admits at the end of Section 3 is not too relevant) he has the prime associated to the excluded row/column of $R_d$ ensure that the number of anti-symmetric primes remaining is even (in the current case, we know that the number of $p_i$ that are 3 mod 4 is odd, as $|d|$ itself is 3 mod 4).

He then computes in Proposition 3.4 and 3.5 the number of $(n+1)$-by-$(n+1)$ anti-symmetric matrices of rank $u$ that have a given $n$-by-$n$ anti-symmetric upper-left corner of rank $v$. When $n$ is even, the number with $u = v$ is $2^{2v-n}$, the number with $u = v+1$ is $2^{v+2} - 3 \cdot 2^{2v-n}$, and the rest have $u = v+2$. Rewriting this in terms of the kernel dimensions $\tilde{u} = n+1-u$ and $\tilde{v} = n-v$, and dividing out by the total number $2^{n+1}$ of such matrices, we find that the proportions are (see also Markov process D in his Appendix I)

$$\begin{cases} 1/2^{1+2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}+1, \\ 2/2^{\tilde{v}} - 3/2^{1+2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}, \\ 1 - 2/2^{\tilde{v}} + 1/2^{2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}-1. \end{cases}$$

When $n$ is odd there is a minor codicil[55] when the matrix has a given type (see (iv) and (v) in Proposition 3.5), but otherwise the proportions are the same. These "transition probabilities" then give a Markov chain that describes the proportion of anti-symmetric matrices of size $n$ that have a kernel of dimension $\tilde{u}$. Solving this Markov chain gives rapid convergence to the $\gamma_{\mathrm{I}}(\tilde{u})$-proportions given above for the distribution of the dimension of the kernel. However, one must still contend with the symmetric part of $R^d$.

Indeed, almost all $d$ will have nearly $r/2$ (up to an error of size roughly $\sqrt{r}$) prime divisors that are 3 mod 4, so we essentially have an $r$-by-$r$ matrix with an upper-left quadrant (after re-ordering the primes) that is anti-symmetric, with the rest symmetric. Gerth then uses the output of the anti-symmetric Markov chain

[55] I don't think this is quite a "non-generic" case as in Swinnerton-Dyer's analysis, but perhaps could be considered as such; the proportion in cases (iv) and (v) has a $1/2^n$ decay.

as input to another one for the symmetric part, and shows that the latter does not change the distribution of kernel dimensions too much.[56]

The argument for the other imaginary cases is similar (though Gerth does not handle them directly, as per Footnote 51), as there is an extra row/column with $\hat{p}_0$, but otherwise the entries of the matrix have the same pattern,[57] and the limiting analysis is the same. This then gives Theorem 10.4.1.

11.2.2. The situation for real quadratic fields has the distinction that the row-sums (or column-sums in Gerth's version) are 0 only for primes that are 1 mod 4 (and also for $p = 2$ when $\hat{p}_0 = 8$). Thus we can remove any single row from the Rédei matrix without changing the rank, but cannot similarly remove a column. After permuting the anti-symmetric portion of size $l$-by-$(l-1)$ to the upper-left quadrant of the matrix, we will have $R^d_{ij} \neq R^d_{ji}$ for $1 \le i \ne j \le l-1$; with $R^d_{lj} = 1$ for $1 \le j \le l-1$ and $R^d_{lj} = 0$ for $l \le j$; and $R^d_{(i+1),j} = R^d_{ji}$ for $l \le i$ and $1 \le j \le l-1$, with $R^d_{(i+1),j} = R^d_{(j+1),i}$ for $l \le i, j$. Gerth then notes that after column exchanges this can be written as $(v\,M)$ where $v$ is a column of $(l-1)$ ones and the rest ones, and $M$ is a matrix as in the previous case, having an anti-symmetric $(l-1)$-by-$(l-1)$ quadrant with the rest being symmetric. Although Gerth only works with the case that $f = 0$, this ansatz in terms of symmetry and anti-symmetry of $\mathbf{F}_2$-entries also holds when there is a row and column corresponding to a nontrivial $\hat{p}_0$.

Again Gerth gives the proportion of such matrices with a given rank in terms of the rank of a smaller such matrix (one less in each dimension), and again there is a slight codicil, here when $n$ is even (see (iv) and (v) of Proposition 5.6). Here the proportions are

$$\begin{cases} 1/2^{2+2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}+1, \\ 3/2^{\tilde{v}+1} - 3/2^{2+2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}, \\ 1 - 3/2^{\tilde{v}+1} + 2/2^{2+2\tilde{v}} & \text{with } \tilde{u} = \tilde{v}-1, \end{cases}$$

leading to the limiting distribution with $\gamma_{\mathrm{R}}(\tilde{u})$ as above. Upon showing that the appending of the symmetric part does not significantly change the distribution, Theorem 10.4.2 then follows.

(The wider applicability of such Markov chain analysis, particularly to the Selmer case, is considered by Klagsbrun, Mazur, and Rubin [25]).

## 12. The case of Gaussian discriminants

Next we turn to the case of Gaussian discriminants. Recall that these are positive fundamental discriminants $d$ with no prime factor that is 3 mod 4. We first describe the case where $d$ is odd.

12.1. Here we take $\mathcal{P}$ to be the set of primes that are 1 mod 4 (so that $\xi_{\mathcal{P}} = 1$ and $\alpha_{\mathcal{P}} = 1/2$), and the modulus $M_{\mathcal{P}}$ to be 4. Although it is ultimately not

[56]Note that merely flipping one entry of a symmetric matrix (*e. g.*, the $(2,1)$-entry) already has a large effect on the rank distribution (for instance, the proportion of full rank matrices is reduced roughly from 41.9% to 31.5%), though I think one needs to de-symmetrize something on the order of $(\log r)$ entries for the statistics to approach those of random matrices in the limit.

[57]Note that when $f = 2$ it is no longer the row-sum that is zero, but rather the row-sum omitting the $\hat{p}_0$-column; thus one can omit any column but this one, and retain the rank.

important to us, one can give the constant in the asymptotic as

$$\Phi^{\mathcal{P}}(X) \sim \frac{1}{\pi} \prod_{p \equiv 1 \,(4)} \sqrt{1-1/p^2} \cdot \frac{X}{\sqrt{\log X}}$$

for the number of squarefree numbers up to $X$ whose prime factors are all 1 mod 4.

12.1.1. Our setup in Section 4 allows us to still utilize the parametrization of relevant $d$ by pleasant boxes.

By Lemma 4.5.1 we find that the exceptional subset of $\tilde{d} \in S^{\mathcal{P}}(X)$ that are not represented by a $(\kappa_0, \eta_1, \eta_{\mathrm{s}})$-pleasant $(X, \eta_0, \mathcal{P})$-box has size

$$\ll \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{\alpha_{\mathcal{P}}\eta_0\kappa_0(\log\kappa_0 - 1 - \log(100/99))}} + \frac{\Phi^{\mathcal{P}}(X)}{(\log\log X)^{99}}.$$

Here the $\mathcal{K}$-conditions are redundant in that they simply re-specify that the primes are to be taken 1 mod 4. As with §6.4, for a pleasant box $\bar{T}$ we have

$$\tilde{r}! \sum_{\substack{\mathcal{L} \in \mathcal{D}(\tilde{r}, \mathcal{P}) \\ e(\mathcal{L}) = e}} \#T(\mathcal{L}) = \sum_{\substack{\mathcal{L} \\ e(\mathcal{L}) = e}} \sum_{\sigma \in \mathrm{Sym}_{\tilde{r}}} \#T(\mathcal{L}^{\sigma}) = \sum_{\substack{\mathcal{L} \in \mathcal{D}(\tilde{r}, \mathcal{P}) \\ e(\mathcal{L}) = e}} \tilde{r}! \frac{\#T}{2^{\binom{\tilde{r}}{2}}} + O\Big(\frac{\tilde{r}! \#T \cdot 2^{k_0/2} k_1}{\sqrt{\log\log X}}\Big),$$

where $k_0 \leq \alpha_{\mathcal{P}}\kappa_0\eta_0 \log\log\log X$ and $k_1 \leq 3\alpha_{\mathcal{P}}(\log\log X)^{\eta_1}$.

In this case in turns out that we can give a fairly simple exact expression for the number of $\mathcal{L}$ that yield a given rank. Indeed (see below for more details), by excluding one row and column we are left with a symmetric matrix of size $(\tilde{r}-1)$ over $\mathbf{F}_2$, and the ranks of such can be determined by means other than a Markov chain. With $\gamma_{\mathrm{G}}$ the limiting distribution (as above) as $\tilde{r} \to \infty$ we thus find that for a pleasant box $\bar{T}$ we have

$$\#\{\tilde{d} \in \hat{T} : e(\tilde{d}) = e\} = \sum_{\substack{\mathcal{L} \in \mathcal{D}(\tilde{r}, \mathcal{P}) \\ e(\mathcal{L}) = e}} \#T(\mathcal{L}) = \gamma_{\mathrm{G}}(e) \cdot \#T + O\Big(\frac{\#T \cdot 2^{k_0/2} k_1}{\sqrt{\log\log X}}\Big)$$

Summing over all such pleasant boxes leads to a similar minimization problem as in §9 for the exponent of $(\log\log X)$ in the error, and we obtain Theorem 10.4.3.

12.1.2. The final task is then to indeed compute the distribution of 4-ranks from $\mathcal{L}$-specifications. Here we use that the Rédei matrix is symmetric and has row- and column-sums equal to zero; such a matrix is determined by any minor obtained by removing a row and column, and the minor itself will be symmetric when the same row and column is removed.

The problem is thereby reduced to the distribution of kernel dimensions of symmetric matrices of size $(\tilde{r}-1)$ over $\mathbf{F}_2$. As Stevenhagen notes [46, Proposition 2.3], this can be done in a fairly elementary manner, and indeed for each $\tilde{r}$ (not just in the limit as $\tilde{r} \to \infty$).

One can initially work over an arbitrary finite field with $q$ elements, where the number of nonsingular symmetric matrices of size $n$ is

$$A_n(q) = q^{\binom{n+1}{2}} \prod_{\substack{1 \leq k \leq n \\ k \text{ odd}}} (1 - 1/q^k),$$

and the number of symmetric matrices with rank $w$ is $A_w(q) \cdot \begin{bmatrix} n \\ w \end{bmatrix}_q$ where

$$\begin{bmatrix} n \\ w \end{bmatrix}_q = \prod_{i=1}^{n} (q^i - 1) \Big/ \prod_{i=1}^{w} (q^i - 1) \prod_{j=1}^{n-w} (q^j - 1)$$

is the number of $w$-dimensional subspaces of a vector space of dimension $n$. Upon taking $q = 2$ this then gives an exact expression for the proportion of Rédei matrices of size $\tilde{r}$ and rank $\tilde{r} - 1 - e$ (thus kernel dimension $(e+1)$ and hence 4-rank of $e$) as

$$\gamma_{\mathrm{G}}^{\tilde{r}}(e) = \frac{1}{2^{\binom{e+1}{2}}} \prod_{j=e+1}^{\tilde{r}-1} (1 - 1/2^j) \Big/ \prod_{j=1}^{(\tilde{r}-e-1)/2} (1 - 1/2^{2j}),$$

and taking the limit as $\tilde{r} \to \infty$ then gives the stated formula.

(One might hope that the first part of Gerth's analysis, that an anti-symmetric matrix has a distribution of kernel dimensions matching that of a random matrix, might have a similar interpretation as here – I must admit to being fairly ignorant of the subject of such anti-symmetric matrices).

12.1.3. The case of $d$ even is mostly the same, with $d$ replaced by $\tilde{d} = d/8$ in the application of the results from Section 4, and then the prime 2 can be taken to be the one excluded from the Rédei matrix. The latter choice ensures that the $\mathcal{K}$-specifications will only need to be modulo 4, rather than 8 as might seem at first glance from the appearance of $(2|p_i)$. Thus they are again irrelevant and the rank only depends on $\mathcal{L}$, and we conclude as before.

12.2. Finally, let us say something about the extension of Fouvry and Klüners [9] to consider the 4-rank of the ordinary class group. Their main result for this (Theorem 2) is that the proportion of Gaussian discriminants with narrow 4-rank $e$ and ordinary 4-rank $e$ is $\gamma_{\mathrm{G}}(e)/2^e$, and thus the proportion for those with narrow 4-rank $e$ and ordinary 4-rank $(e-1)$ is $\gamma_{\mathrm{G}}(e)(1 - 1/2^e)$.

Following their argument, the condition that the 4-ranks are equal can be detected by Rédei symbols involving the infinite prime, and these can be given as quartic residue symbols in this case (see [9, §3.3, §4], or Stevenhagen's version given in [47, §4] or [48, §10]). Thus it appears one should replace the Legendre specifications by $\mathcal{Q}$-specifications (taking values in $\{\pm 1, \pm\zeta_4\}$) for quartic symbols between the prime divisors, and then the ordinary 4-rank is determined by $(\mathcal{K}, \mathcal{Q})$ (also fixing whether $d$ is even or odd). Writing $\pi_j$ for a primary factor of $p_j$, these symbols are then detected by $\left[1 + \bar{\mathcal{Q}}_{ij}(p_i|\pi_j) + \mathcal{Q}_{ij}^2(p_i|\pi_j)^2 + \mathcal{Q}_{ij}(p_i|\pi_j)^3\right]$, with the relevant boxes then being reduced roughly by a factor of 4 for each pair $(i, j)$. As Fourvy and Klüners note (§6.3), the methods for bilinear bounds are quite robust and can rather easily be adapted give the desired cancellation. Meanwhile, one can introduce Hecke Grössencharacters over $\mathbf{Q}(\sqrt{-1})$ to handle the congruential sums (see [9, §5, Proposition 7]).

Smith's application of permutations should also readily extend to this case.

Finally, one would need to determine the 4-rank distribution by an analysis of what proportion of $(\mathcal{K}, \mathcal{Q})$ yield given 4-ranks of the class groups.

However, the introduction of all the necessary machinery would take us too far afield, and so we leave this problem for the interested reader.

12.3. Let us briefly mention the case of 8-ranks, where instead of Legendre symbols one can determine it in terms of Rédei symbols (see [48, Definition 4.4]). These are certainly not so nice from an analytic standpoint, though can still be analyzed via Artin representations (in the guise of Frobenius equi-distribution and the Chebotarev density theorem), where at least some results about equi-distribution exist.

In any event, Fouvry and Klüners [10] are able to show various distribution results regarding the 8-rank. For instance (Theorem 2) they show that the proportion of Gaussian discriminants with both 4-ranks equal to 1 is equally split between narrow 8-rank 0 and 1 (both proportions being $\gamma_{\mathrm{G}}(1)/4$). My understanding is that their $\lambda_D$ (see Theorem 3, Definition 2, and §4.3) detects a weaker condition than the Rédei symbol equalities that would actually determine the 8-rank; but on the other hand this $\lambda_D$ can be usefully shoe-horned into the analysis over $\mathbf{Q}(\sqrt{-1})$ in a manner similar to [9]. (Thus they avoid any specifics with Artin representations).

12.3.1. The recent preprint of Chan, Koymans, Milovic, and Pagano [3] considers the 8-rank distribution (by an analogue of Smith's later methods that we didn't discuss here), and in particular they re-obtain the result for the proportion of Gaussian discriminants with equal narrow and ordinary 4-ranks. Contrary to the above, their arguments do not use any analytic number theory over $\mathbf{Q}(\sqrt{-1})$, as they instead work via the Chebotarev density theorem.

Our sequel [54] to the current work gives an exposition of this; from the standpoint of the $2^\infty$-Selmer rank distribution as considered by Smith, this corresponds to the 4-Selmer case. Here there are various simplifications that occur so as to avoid the need to get too heavily into co-homological machinery (for instance, when defining the higher $2^k$-pairing matrices), and the methods (due mostly to Smith) turn out to be largely combinatorial in nature, involving an arrangement to show equi-distribution of a Frobenius element.

Indeed, for generic $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$, Smith isolates three of the primes dividing $d$ (one of them restricted to be large, say $\log\log p_k \sim (2/3)\log\log X$, and the other two small, say $\log\log p_j \sim (1/3)\log\log X$), and then describes situations where the 4-Selmer pairing matrix (or more properly, the characters on its ambient space) depends on a Frobenius element for the large prime in a field defined by the small primes. However, this leads to much too large of a field degree when all the small primes $p_j \in T_j$ are used at once, so Smith proceeds to cover the $T_j$ by suitable "grids" (subsets with a few extra properties, such as Legendre conditions) of much smaller size, thereby reducing the field degree and allowing the Chebotarev theorem to be gainfully applied. (In the general $2^k$-Selmer case there are $k$ small primes involved; but more crucially, for $k > 2$ setting up the beneficial situations where the Selmer pairing matrix depends on a Frobenius element involves significantly deeper co-homological considerations).

## 13. Exercises

13.1. Let us do the exercise mentioned in Footnote 1. Our proof of this is somewhat tedisome, though essentially elementary.

**Lemma 13.1.1.** *Suppose that an elliptic curve $E$ over $\mathbf{Q}$ has full 2-torsion. Then there is some isogenous curve (possibly $E$ itself) that has full 2-torsion and no rational 4-torsion point.*

*Proof.* The generic form of a curve $E$ with full 2-torsion is $y^2 = x(x+1)(x+\lambda)$, and this has a 4-torsion point $(\mu, \mu^2+\mu)$ when $\lambda = \mu^2$. Thus we can assume $\lambda$ is square (and neither 0 nor 1, as these are singular), or else we are already done.

We can write $E$ as $A(1+\lambda, \lambda)$ for $A(a,b) : y^2 = x(x^2+ax+b)$, and recall there is a 2-isogeny map from $A(a,b)$ to $A(-2a, a^2-4b)$. In particular, our given curve $E$ is 2-isogenous to the curve $E'$ given by $A\big(-2(1+\lambda), (\lambda-1)^2\big)$. This $E'$ has a model $y^2 = x(x-(\mu-1)^2)(x-(\mu+1)^2)$, which can then be transformed to $y^2 = x(x+1)(x+\beta)$ where $\beta = 1-(\mu-1)^2/(\mu+1)^2 = 4\mu/(\mu+1)^2$. Now $E'$ has full 2-torsion, and so unless $\beta$ (ergo $\mu$) is square we are done.

Otherwise we write $\theta^2 = 4\mu/(\mu+1)^2$, so that $E'$ is isomorphic to $A(1+\theta^2, \theta^2)$, and is thus 2-isogenous to the curve $E''$ given by $A\big(-2(1+\theta^2), (\theta^2-1)^2\big)$. Again this $E''$ has a model $y^2 = x(x-(\theta-1)^2)(x-(\theta+1)^2)$ which can be transformed to $y^2 = x(x+1)(x+\gamma)$ where $\gamma = 1-(\theta-1)^2/(\theta+1)^2 = 4\theta/(\theta+1)^2$. We see that $E''$ has full 2-torsion, and the condition that it have a 4-torsion point is that $\theta$ be square, say $\theta = \rho^2$. We would then have $\rho^4 = 4\mu/(\mu+1)^2$, which we will find has no rational solutions other than from $\mu \in \{0,1\}$.

Indeed, upon writing $z = \rho(\mu+1)$ and $\mu = r/s$ with $\gcd(r,s) = 1$ we see that $z^4 = 4\mu(\mu+1)^2 = 4(r/s)(r/s+1)^2$. This yields $z^4 s^3 = 4r(r+s)^2$, where exactly one of $r, s, (r+s)^2$ is even, and thus has 2-valuation congruent to 2 mod 4, while the other two 2-valuations are multiples of 4. Breaking this up into the three cases, first when $r$ is even we write $(r, s, (r+s)^2) = (4a^4, b^4, z^4b^{12}/16a^4)$, and upon writing $c^4 = z^4b^{12}/16a^4 = (r+s)^2$, by equating the sum of $r$ and $s$ with $(r+s)$ we are left with the equation $4a^4 + b^4 = \pm c^2$. Similarly, when $s$ is even we have $a^4 + 4b^4 = \pm c^2$, while when $(r+s)$ is even we get $a^4 + b^4 = \pm 2c^2$ from $(r, s, (r+s)^2) = (a^4, b^4, 4z^4b^{12}/16a^4)$. Clearly none of these is solvable with the minus sign; with the plus sign, by Fermat descent the only coprime solution with no co-ordinate of 0 is $(\pm 1, \pm 1, \pm 1)$ to the third, which corresponds to $\mu = r/s = 1$.

Thus $E''$ has no rational 4-torsion point and we are done. □

13.2. Our second exercise appeared in §7.2.1. We want to show that exactly half of the choices of $(\mathcal{K}_\varepsilon, \mathcal{L})$ yield 2-Selmer rank of each parity. As Yu [58] points out, Monsky [36] shows that a stronger statement (namely the root number changes by $(d|-N_E)$ when twisting by a fundamental discriminant $d$ coprime to $N_E$) follows from work culminated by Kolyvagin [26]; this follows directly by [26] when the analytic rank is $\le 1$, and one can always reduce to such a case by a suitable quadratic twist, with calculations of Kramer [29] then giving the applicable parity under twisting. What we show here is that we can avoid Kolyvagin's work if we only aim to show that a 50-50 split in parity (essentially, we can ignore the "base case" of the twisting calculation).

Let us recap our situation. We fix an elliptic curve $E$ with full 2-torsion that is twist-minimal in the sense that no prime has the same nonzero valuation at all the $\delta_{ij}$. We then wish to show that the 2-Selmer ranks for specifications $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ have an equal 50-50 split in parity. As noted in Swinnerton-Dyer's analysis, the 2-Selmer rank has the same parity as the 2-Selmer estimation for the 0th restriction $(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})[0]$, which contains no information about $\mathcal{L}$, while the $\tilde{\mathcal{K}}_\varepsilon[0]$-conditions specify the sign of $d$ and (by convention) the image of $|d|$ in $\prod_{l\in\Omega} \mathbf{Q}_l^\star/(\mathbf{Q}_l^\star)^2$. We denote the $\tilde{\mathcal{K}}_\varepsilon^{\mathrm{op}}[0]$-conditions to mean those that flip the local condition at each $q \in \tilde{\Omega}$ that is not 1 mod 4, corresponding to the image of $d$ itself (when negative).

13.2.1. Let us review what appears in the literature for the 2-Selmer rank when twisting. The main paper is that of Kramer [29]. We let $E/\mathbf{Q}$ be an elliptic curve and will consider twisting it by odd squarefree $d$ coprime to the conductor $N_E$.

As Mazur and Rubin catalogue [34, Theorem 2.7], one consequence of Kramer's results is that the 2-Selmer ranks $s(E)$ of $E$ and $s(E_d)$ of $E_d$ are related as

$$s(E_d) \equiv s(E) + \sum_p \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big),$$

where the congruence is mod 2 (as are all future ones), while $\delta_p$ is defined as the codimension of the image $E_{\mathbf{N}}(\mathbf{Q}_p)$ of the local norm map $E(\mathbf{Q}(\sqrt{d})_{\mathfrak{p}}) \to E(\mathbf{Q}_p)$, that is, $\delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) = \dim_{\mathbf{F}_2}\big(E(\mathbf{Q}_p)/E_{\mathbf{N}}(\mathbf{Q}_p)\big)$. For our result, we need only discuss the parity of the $\delta_p$. We write $\Delta_E$ for the discriminant of $E$, which we assume is given as a minimal model.

| $p$ | $\mathbf{Q}_p(\sqrt{d})/\mathbf{Q}_p$ | $E$ at $p$ | $\delta_p \equiv$ | where |
|---|---|---|---|---|
| finite | trivial | any | $0 = (\Delta_E, d)_p = (p\Delta_E, d)_p$ | obvious |
| finite | unramified | good | $0 = (\Delta_E, d)_p$ | Mazur |
| finite | unramified | split | $(p\Delta_E, d)_p = (\Delta_E, d)_p + 1$ | Prop 1 |
| finite | ramified | split | $(\Delta_E, d)_p + 1$ | Prop 1 |
| finite | unramified | nonsplit | $(p\Delta_E, d)_p = (\Delta_E, d)_p + 1$ | Prop 2a |
| finite | ramified | nonsplit | $(\Delta_E, d)_p$ | Prop 2b |
| odd | ramified | good | $(\Delta_E, d)_p$ | Prop 3 |
| 2 | ramified | supersingular | $(\Delta_E, d)_p$ | Prop 4 |
| 2 | ramified | ordinary | $(\Delta_E, d)_p$ | Prop 5 |
| $\infty$ | | | $(-\Delta_E, d)_p$ | Prop 6 |

TABLE 9. Computation of the parity of $\delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big)$

We sum up Kramer's work in Table 9. From it, one can read off (from the fourth column) the parity of $\delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big)$ for a semistable elliptic curve $E/\mathbf{Q}$ at any place $p$ (including 2 and $\infty$, and those that divide $d$ or $N_E$). The first line simply records the case when $d$ is a square in $\mathbf{Q}_p$, with the convention then that $\delta_p = 0$. The results from Propositions 1 and 2 refer to the cases of multiplicative reduction.

Adding up the results, we can use the reciprocity law $\sum_p (\Delta_E, d)_p = 0$ to simplify. First we might note that when $\gcd(d, 2\Delta_E) = 1$ the lines with ramified primes for multiplicative reduction do not occur, so Kramer's work implies that when twisting a semistable curve $E$ by an odd squarefree $d$ coprime to $N_E$ we have

$$\sum_p \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) = (-1, d)_\infty + \sum_{\substack{p \mid \Delta_E \\ \chi_d(p) = -1}} 1 = (-1, d)_\infty + \sum_{p \mid N_E} (p, d)_p.$$

When $d$ is a fundamental discriminant this says $(-1)^{\sum_p \delta_p} = (d|-N_E)$, so indeed matches with the previous formulation for root numbers.

13.2.2. Kramer's hope of resolving the case of additive reduction in the future does not seem to have come to fruition. However, we can still show enough of the necessary results for our purposes, in particular, that we have equi-distribution of the parity of $s(\tilde{\mathcal{K}}_\varepsilon, \mathcal{L})$ over the possibilities for $\tilde{\mathcal{K}}_\varepsilon[0]$.

We recall that we assume we are twisting by odd squarefree $d$ coprime to $N_E$.

From the above reformulation of Mazur and Rubin we have

$$s(E_d) - s(E) \equiv \sum_p \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big).$$

Splitting the primes/places into three types: infinite, bad for $E$, and good for $E$, this becomes

$$s(E_d) - s(E) \equiv (-\Delta_E, d)_\infty + \sum_{p|N_E} \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) + \sum_{p\nmid\infty N_E} \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big).$$

Then we use that $\delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) \equiv (\Delta_E, d)_p$ for good primes, so that

$$s(E_d) - s(E) \equiv (-\Delta_E, d)_\infty + \sum_{p|N_E} \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) + \sum_{p\nmid\infty N_E} (\Delta_E, d)_p.$$

Applying the reciprocity law of $\sum_p (\Delta_E, d)_p = 0$ then gives

$$s(E_d) - s(E) \equiv (-1, d)_\infty + \sum_{p|N_E} \delta_p\big(E, \mathbf{Q}(\sqrt{d})/\mathbf{Q}\big) + \sum_{p|N_E} (\Delta_E, d)_p.$$

Since $d$ is odd and coprime to $N_E$, the summands in both the second and third sums are determined by whether $d$ is a square mod $p$ (or its class in $\mathbf{Q}_2^\star/(\mathbf{Q}_2^\star)^2$ for $p = 2$), and so are determined by $\tilde{\mathcal{K}}[0]$. Meanwhile, the $(-1, d)_\infty$-term is independent of the $p$-adic conditions from the $p|N_E$. Since $(-1, d)_\infty$ itself is equi-distributed with respect to $d$, the entire right-hand expression is also. In other words, for any choice of $\tilde{\mathcal{K}}[0]$, the 2-Selmer rank parities of $\tilde{\mathcal{K}}_+[0]$ and $\tilde{\mathcal{K}}_-^{\mathrm{op}}[0]$ will differ.[58] (One reason to include this exercise is to emphasize how the prime at infinity enforces the parity split; compare [24, Corollary 7.10] of Klagsbrun, Mazur, and Rubin, who compute the general disparity of 2-Selmer ranks for quadratic twists).

This then shows the desired parity equi-distribution. (Our argument did not use the presence of full 2-torsion).

13.3. Next we show a variant of Heilbronn's bilinear estimate (Lemma 3.4.3). We want to show that if $\{\alpha_m\}$ and $\{\beta_n\}$ are sequences of complex numbers bounded by 1 and supported on odd squarefree integers $m$ and $n$ with $M \le m \le 2M$ and $N \le n \le 2N$, then $\sum_m \sum_n \alpha_m \beta_n (m|n) \ll MN/\min(M, N)^{1/9}$.

By dividing $m$ and $n$ into congruence classes mod 8 we can replace $(m|n)$ by $(n|m)$ with a fixed sign, and thus by symmetry we can assume that $N \le M$.

13.3.1. Almost every author seems to have their own version of this type of result. The history is a bit eclectic, as Heilbronn's original argument is one paragraph long, using Cauchy's inequality and partial character sum estimates (of Pólya and Vinogradov); then, various authors gave more weighty ratiocination to the methods (sometimes to debatable avail in the end result) such as the partial character sum estimate of Burgess or the large sieve; later, it was realized that one can actually dispense with partial character sum estimates (using periodicity instead) and still obtain an adequate result (via Hölder's inequality rather than Cauchy's).

Heilbronn [20] originally considered $\sum_p \sum_q (p|q)$ with both variables prime and of size $X$, and saved $X^{1/4}$ over the trivial estimate via using Cauchy's inequality twice and the estimate of Pólya and Vinogradov for partial character sums. It is

[58] Note that this is not the same as twisting by $-1$ (which need not flip the parity). For instance, for the congruent number curve, we have that the $d \equiv 1, 3\,(8)$ classes retain the parity for $d > 0$, while these same $d \equiv 1, 3\,(8)$ classes flip the parity for $d < 0$.

fairly routine to modify this to include weights $\alpha_p$ and $\beta_q$, and allow the variables to run over dyadic intervals of different sizes.

Gerth and Graham [13, Theorems 4 & 3] essentially give the "obvious" generalization of Heilbronn's result to sequences supported on the squarefree[59] integers, using Cauchy's inequality twice and the estimate of Pólya and Vinogradov.

Heath-Brown's version in [17, Lemma 4] desires $\vec{\alpha}, \vec{\beta}$ to be supported on odd squarefree integers rather than primes, and thus the second application of Cauchy's inequality would induce a divisor function – he instead applies the estimate of Burgess after only one application, so saves only $(1/16 - \epsilon)$ instead of $1/4$ in the exponent.

In a later paper [19, Corollary 4], whose main topic is such estimates, Heath-Brown gives a version (for odd[60] integers) that saves a larger power ($1/2$ instead of $1/4$) of the parameter $N$, but also has an extra $(MN)^\epsilon$ included. I don't think this is a "large sieve" *per se*, and indeed he terms it a mean-value estimate.[61] The interposition of $(MN)^\epsilon$ here can cause difficulties when $M$ and $N$ are on much different scales, and unlike other occurrences of this, I don't think this is just a shorthand for a divisor function (which can be bounded by a log-power on average) in his estimate.

Smith (Proposition 6.6) has the support be on primes, and adapts a result of Jutila [22, Lemma 3]; he terms this a form of the large sieve, though I don't think this is the best description.[62]

Kane's Lemma 15 also has the support be on primes, though he gives his own derivation which relies on a multiplicative large sieve estimate (in Lemma 14) instead of either the bound of Pólya and Vinogradov or of Burgess.[63]

Fouvry and Klüners [8, Lemmata 14 & 15] has the support on odd squarefree integers; they give two slightly different versions, the first of which comes directly from Heath-Brown's later version [19], and the second of which (to save $(MN)^\epsilon$) is given as a consequence of the large sieve (again this is perhaps overly high-powered).

Their version in [9, §6] is over $\mathbf{Z}[\sqrt{-1}]$, but they note in general that one needs nothing more than reciprocity (for symmetry in the variables), bi-multiplicativity, and some sort of cancellation in character sums (indeed, they merely use the trivial bound from periodicity). They save $(1/8 - \epsilon)$ in the exponent, with a clean $1/8$ when the support is on the primes. This is essentially the argument that I've chosen to present; as they note, it appears in a different guise in various other works (most

[59] They don't state this requirement, but in the proof of Theorem 3 they use that $n \neq r$ (rather than that $nr$ is non-square) to be able to apply the estimate of Pólya and Vinogradov.

[60] Without such a requirement, at least on the lower entry in $(m|n)$, one must contend with non-periodicity when $m$ is an odd nonfundamental discriminant; for instance, $(3|n)$ is not periodic since $(3|2^k) = -(3|7 \cdot 2^k) = -(3|2^k + 3 \cdot 2^{k+1})$ for all $k$.

[61] If I were pressed to make a distinction, a sum over all characters for moduli up to $Q$ would be a large sieve, whilst a sum over only the real characters would not – though below both Kane [23] and Fouvry and Klüners [8] bound the latter by the former (trivially by inclusion), which seems to me to be a rather artificial usage of the large sieve in this context.

[62] Jutila's Lemma 3 is deduced from the prior Lemma 2, which he mis-cites as being from his [2] instead of his [3], but in any event it seems to be a mean value theorem for real character sums rather than a large sieve (see his Introduction).

[63] With his Lemma 14 he mentions the similarity to Lemma 4 of [18] – I suspect he means Lemma 4 of [17] (which is Lemma 3 in [18]), and in any case his comment would seem more apropos adjoining his Lemma 15.

notably, Friedlander and Iwaniec [11, §21] gave such a version over $\mathbf{Z}[\sqrt{-1}]$; also, Koymans and Milovic [27, §3.4, §5] have a result for general number fields).

13.3.2. We use the $m \sim M$ notation to indicate a dyadic interval and thus write

$$S(\vec{\alpha}, \vec{\beta}) = \sum_{m\sim M} \sum_{n\sim N} \alpha_m \beta_n (m|n)$$

In many instances it is natural to have the coefficients bounded by a divisor function (rather than by 1), and we write $\tau_l$ for the $l$-fold divisor function, with $l \ge 1$.

First we show a result that is useful when one variable significantly exceeds the other, for instance $N \ll_\epsilon \sqrt{M}/M^\epsilon$.

**Lemma 13.3.3.** *Suppose that $|\alpha_m| \le \tau_a(m)$ and $|\beta_n| \le \tau_b(n)$ are sequences supported on odd integers in $[M, 2M]$ and $[N, 2N]$. Then*

$$\left|S(\vec{\alpha}, \vec{\beta})\right|^2 \underset{a,b}{\ll} (MN)^2 \left(\frac{1}{N} + \frac{N^2}{M}\right) (\log MN)^{a^2+2b^2+2b}.$$

*Proof.* By Cauchy's inequality and expanding the square we have

$$\begin{aligned}\left|S(\vec{\alpha}, \vec{\beta})\right|^2 &= \left|\sum_{m\sim M} \alpha_m \sum_{n\sim N} \beta_n (m|n)\right|^2 \le \sum_{m\sim M} |\alpha_m|^2 \cdot \sum_{m\sim M} \left|\sum_{n\sim N} \beta_n (m|n)\right|^2 \\ &\le \sum_{m\ll M} \tau_a(m)^2 \cdot \sum_{n_1\sim N} \tau_b(n_1) \sum_{n_2\sim N} \tau_b(n_2) \left|\sum_{m\sim M} (m|n_1 n_2)\right|.\end{aligned}$$

When $n_1 n_2$ is square the inner $m$-sum is $\ll M$, and else it is bounded as $\ll n_1 n_2$. Writing $u^2 = n_1 n_2$ and recalling $\sum_{u\le X} \tau_l(u)^j \ll_{j,l} X(\log X)^{l^j-1}$, this gives

$$\begin{aligned}\left|S(\vec{\alpha}, \vec{\beta})\right|^2 &\underset{a,b}{\ll} M(\log M)^{a^2-1} \cdot \left(M \sum_{u\ll N} \tau_{2b}(u^2) + N^2 \cdot N^2 (\log N)^{2(b-1)}\right) \\ &\underset{a,b}{\ll} M^2 N^2 \cdot \left(\frac{(\log N)^{2b^2+b-1}}{N} + \frac{N^2}{M} (\log N)^{2(b-1)}\right) \cdot (\log M)^{a^2-1}.\end{aligned}$$

where we used $\tau_{2b}(u^2) \le \tau_{\binom{2b+1}{2}}(u) = \tau_{b(2b+1)}(u)$. □

The above is not quite symmetrical in the variables. In many circumstances we can induce symmetry by fixing the sign of $(m|n)(n|m)$ by restricting $m$ and $n$ to suitable arithmetic progressions. However, this again is somewhat unneeded.

**Lemma 13.3.4.** *Suppose that $|\alpha_m| \le \tau_a(m)$ and $|\beta_n| \le \tau_b(n)$ are sequences supported on odd integers in $[M, 2M]$ and $[N, 2N]$. Then*

$$\left|S(\vec{\alpha}, \vec{\beta})\right|^2 \underset{a,b}{\ll} (MN)^2 \left(\frac{1}{M} + \frac{M^2}{N}\right) (\log MN)^{b^2+2a^2+2a}.$$

*Proof.* This is the same proof, except we expand out the $m$-variable by Cauchy's inequality, and use the periodicity of $(m_1 m_2|n)$ when $m_1 m_2$ is non-square. □

Now we show the main result.

**Proposition 13.3.5.** *Suppose that $|\alpha_m| \le \tau_c(m)$ and $|\beta_n| \le \tau_c(n)$ are sequences supported on odd integers in $[M, 2M]$ and $[N, 2N]$. Then*

$$\left|S(\vec{\alpha}, \vec{\beta})\right| \underset{c}{\ll} \frac{MN}{\min(M, N)^{1/9}}.$$

*Proof.* First we assume $N \le M$. We apply Hölder's inequality to $S(\vec{\alpha}, \vec{\beta})$ and get

$$\left|S(\vec{\alpha}, \vec{\beta})\right|^4 \le \Big(\sum_{n\sim N} |\beta_n|^4\Big)^3 \Big(\sum_{n\sim N}\Big|\sum_{m\sim M} \alpha_m (m|n)\Big|^4\Big) \ll \Big(N(\log N)^{c^4-1}\Big)^3 \cdot \left|S(\vec{\alpha}', \vec{\beta}')\right|$$

where $\alpha'_u$ is the sum of $\alpha_{m_1}\alpha_{m_2}\bar{\alpha}_{m_3}\bar{\alpha}_{m_4}$ over representations $u = m_1m_2m_3m_4$, and thus $|\alpha'_u| \le \tau_{4c}(u)$, while $\beta'_n$ is 1 on $[N, 2N]$. Then we split up the interval $[M^4, 16M^4]$ into 4 dyadic intervals and apply Lemma 13.3.3 to get

$$\left|S(\vec{\alpha}', \vec{\beta}')\right|^2 \underset{c}{\ll} (M^4N)^2\Big(\frac{1}{N} + \frac{N^2}{M^4}\Big)(\log MN)^{20c^2},$$

so that

$$\left|S(\vec{\alpha}, \vec{\beta})\right| \underset{c}{\ll} MN\Big(\frac{1}{N^{1/8}} + \frac{N^{1/4}}{\sqrt{M}}\Big)(\log MN)^{4c^4},$$

and then using $N \le M$ gives the stated result.

The argument when $M \le N$ is similar, except we apply Hölder's inequality to $n$ and use Lemma 13.3.4. □

## References


[1] A. O. L. Atkin, J. Lehner, *Hecke operators on* $\Gamma_0(m)$. Math. Ann. **185** (1970), no. 2, 134–160. http://eudml.org/doc/161948

[2] M. Bhargava, D. M. Kane, H. W. Lenstra, B. Poonen, and E. Rains, *Modeling the distribution of ranks, Selmer groups, and Shafarevich-Tate groups of elliptic curves*. Camb. J. Math. **3** (2015), no. 3, 275–321. http://doi.org/10.4310/CJM.2015.v3.n3.a1

[3] S. Chan, P. Koymans, D. Milovic, C. Pagano, *On the negative Pell equation*. Preprint, 2019.

[4] H. Cohen, H. W. Lenstra, *Heuristics on class groups of number fields*. In *Number Theory (Noordwijkerhout, 1983)*, Springer LNM **1068** (1984), 33–62. http://doi.org/10.1007/BFb0099440

[5] C. Delaunay, *Heuristics on Tate-Shafarevitch groups of elliptic curves defined over* **Q**. Experiment. Math. **10** (2001), no. 2, 191–196. http://doi.org/10.1080/10586458.2001.10504442

[6] P. Erdös, M. Kac. *The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions*. Amer. J. Math. **62**, no. 1, 738–742. http://doi.org/10.2307/2371483

[7] É. Fouvry, J. Klüners, *Cohen-Lenstra heuristics of quadratic number fields*. In *Algorithmic Number Theory*, edited by F. Hess, S. Pauli, and M. Pohst. Springer LNCS **4076** (2006), 40–55. http://doi.org/10.1007/11792086_4

[8] É. Fouvry, J. Klüners, *On the 4-rank of class groups of quadratic number fields*. Invent. Math. **167** (2007), 455–513. http://doi.org/10.1007/s00222-006-0021-2

[9] É. Fouvry, J. Klüners, *On the negative Pell equation*. Ann. Math. **172** (2010), no. 3, 2035–2104. http://doi.org/10.4007/annals.2010.172-3

[10] É. Fouvry, J. Klüners, *The parity of the period of the continued fraction of* $\sqrt{d}$. Proc. London Math. Soc. **101** (2010), 337–391. http://doi.org/10.1112/plms/pdp057

[11] J. Friedlander, H. Iwaniec, *The polynomial* $X^2 + Y^4$ *captures its primes*. Ann. Math. **148** (1998), 945–1040. http://doi.org/10.2307/121034

[12] F. Gerth, *The 4-class ranks of quadratic fields*. Invent. Math. **77** (1984), 489–515. http://doi.org/10.1007/BF01388835

[13] F. Gerth, S. Graham, *Application of a character sum estimate to a 2-class number density*, J. Num. Theory **19** (1984), 239–247. http://doi.org/10.1016/0022-314X(84)90108-2

[14] D. M. Goldfeld, *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **3** (1976), no. 4, 624–663. http://eudml.org/doc/83732

[15] D. M. Goldfeld, A. Schinzel, *On Siegel's zero*. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **2** (1976), no. 4, 571–583. http://eudml.org/doc/83704

[16] B. H. Gross, D. B. Zagier, *Heegner points and derivatives of L-series*. Invent. Math. **84** (1986), no. 2, 225–320. http://eudml.org/doc/143341

[17] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem*. Invent. Math **111** (1993), 171–195. http://doi.org/10.1007/BF01231285
[18] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem, II*. Invent. Math **118** (1994), 331–370. http://doi.org/10.1007/BF01231536
[19] D. R. Heath-Brown, *A mean value estimate for real character sums*. Acta Arith. **72** (1995), 235–275. http://eudml.org/doc/206794
[20] H. Heilbronn, *On the averages of some arithmetical functions of two variables*. Mathematika **5** (1958), 1–7. http://doi.org/10.1112/S0025579300001273
[21] H.-K. Hwang, *Sur la répartition des valeurs des fonctions arithmétiques. Le nombre de facteurs premiers d'un entier*. (French) [On the distribution of values of arithmetic functions. The number of prime factors of an integer]. J. Num. Theory **69** (1998), 135–152. http://doi.org/10.1006/jnth.1997.2216
[22] M. Jutila, *On mean values of Dirichlet polynomials with real characters*. Acta Arith. **XXVII** (1975), 191–198. http://eudml.org/doc/205338
[23] D. Kane, *On the ranks of the* 2*-Selmer groups of twists of a given elliptic curve*. Algebra & Number Theory **7** (2013), no. 5, 1253–1279. http://doi.org/10.2140/ant.2013.7.1253
[24] Z. Klagsbrun, B. Mazur, K. Rubin, *Disparity in Selmer ranks of quadratic twists of elliptic curves*. Ann. Math. **178** (2013), 1–34. http://doi.org/10.4007/annals.2013.178.1.5
[25] Z. Klagsbrun, B. Mazur, K. Rubin, *A Markov model for Selmer ranks in families of twists*. Compositio Math. **150** (2014), 1077–1106. http://doi.org/10.1112/S0010437X13007896
[26] V. A. Kolyvagin, *Finiteness of* $E(\mathbf{Q})$ *and* $Ш(E,\mathbf{Q})$ *for a subclass of Weil curves*. Izv. Akad. Nauk SSR Ser. Mat. **52** (1988), no. 3, 522–540 (Russian), Math. USSR Izv. **32** (1989), 523–541 (English translation). http://doi.org/10.1070/IM1989v032n03ABEH000779
———, *Euler systems*. In *The Grothendieck Festschrift* (Vol. II), ed. P. Cartier et al., Prog. in Math. **87** Birkhäuser Boston (1990), 435–483. http://doi.org/10.1007/978-0-8176-4575-5
[27] P. Koymans, D. Milovic, *On the 16-Rank of Class Groups of* $\mathbf{Q}(\sqrt{-2p})$ *for Primes* $p \equiv 1$ *mod 4*. IMRN **2019**, 7406–7427. http://doi.org/10.1093/imrn/rny010
[28] P. Koymans, C. Pagano, *Effective convergence of coranks of random Rédei matrices*. Preprint, 2020.
[29] K. Kramer, *Arithmetic of elliptic curves upon quadratic extension*. Trans. AMS **264** (1981), 121–135. http://doi.org/10.2307/1998414
[30] E. Landau, *Uber die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate*. (German) [On the division of the positive integers into four classes according to the minimum number of required squares in their additive decomposition]. Arch. der Math. und Phys. **13** (1908), 305–312. http://archive.org/details/archivdermathem37unkngoog/page/n324
[31] E. Landau, *Über die Klassenzahl imaginär-quadratischer Zahlkörper*. (German) [On the class number of imaginary quadratic fields]. Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. (1918), 285–295. http://www.digizeitschriften.de/dms/resolveppn/?PID=GDZPPN002505142
[32] G. Landsberg, *Ueber eine Anzahlbestimmung und eine damit zusammenhängende Reihe*. (German) [On a numerical determination and a related series]. J. reine angew. Math. **111** (1893), 87–88. http://eudml.org/doc/148874
[33] Q. Lu, *8-rank of the class group and isotropy index*. Sci. China Math. **58** (2015), 1433–1444. http://doi.org/10.1007/s11425-014-4898-8
[34] B. Mazur, K. Rubin, *Ranks of twists of elliptic curves and Hilbert's tenth problem*. Invent. Math. **181** (2010), 541–575. http://doi.org/10.1007/s00222-010-0252-0
[35] F. Mertens, *Ein Beitrag zur analytischen Zahlentheorie*. (German) [A contribution to analytic number theory]. J. reine angew. Math. **78** (1874), 46–62. http://resolver.sub.uni-goettingen.de/purl?GDZPPN002155656
[36] P. Monsky, *Generalizing the Birch-Stephens theorem I. Modular curves*. Math. Z. **221** (1996), 415–420. http://doi.org/10.1007/BF02622123
[37] B. N. Parlett, C. Reinsch, *Balancing a Matrix for Calculation of Eigenvalues and Eigenvectors*. Numerische Mathematik **13**, no. 4 (1969), 293–304. Also published as Contribution II/11 in *Linear Algebra. Handbook for Automatic Computation, vol. 2*, edited by F. L. Bauer (1971), 315–326. http://doi.org/10.1007/978-3-662-39778-7_22
[38] L. Rédei, *Arithmetischer Beweis des Satzes über die Anzahl der durch vier teilbaren Invarianten der absoluten Klassengruppe im quadratischen Zahlkörper*. (German) [Arithmetic

proof of the theorem about the number of 4-divisible invariants of the absolute class group of a quadratic field]. J. reine angew. Math. **171** (1934), 55–60. http://eudml.org/doc/149883

[39] L. Rédei, *Über einige Mittelwertfragen im quadratischen Zahlkörper*. J. reine angew. Math. **174** (1936), 15–55. (German) [On some questions of the mean in quadratic fields]. http://eudml.org/doc/149937

[40] L. Rédei, *Ein neues zahlentheoretisches Symbol mit Anwendungen auf die Theorie der quadratischen Zahlkörper*. (German) [A new number-theoretical symbol with applications to the theory of quadratic fields]. J. reine angew. Math. **180** (1939), 1–43. http://eudml.org/doc/150050

[41] L. Rédei, H. Reichardt, *Die Anzahl der durch 4 teilbaren Invarianten der Klassengruppe eines beliebigen quadratischen Zahlkörpers*. (German) [The number of 4-divisible invariants in the class group in any quadratic field]. J. reine angew. Math. **170** (1934), 69–74. http://eudml.org/doc/149862

[42] L. G. Sathe, *On a Problem of Hardy on The Distribution of Integers having a Given Number of Prime Factors, I, II, III, IV*. J. Indian Math. Soc. **17** (1953), 63–82, 83–141, **18** (1954), 27–42, 43–81. http://doi.org/10.18311/jims/1953/17038 http://doi.org/10.18311/jims/1953/17033 http://doi.org/10.18311/jims/1954/17016 http://doi.org/10.18311/jims/1954/17017

[43] A. Selberg, *Note on a paper by L. G. Sathe*. J. Indian Math. Soc. **18** (1954), 83–87. http://doi.org/10.18311/jims/1954/17018

[44] C. L. Siegel, *Über die Classenzahl quadratischer Zahlkörper*. (German) [On the class number of quadratic fields]. Acta Arithmetica **1** (1935), 83–86. https://eudml.org/doc/205054

[45] A. Smith, $2^\infty$*-Selmer groups,* $2^\infty$*-class groups, and Goldfeld's conjecture*. Preprint, 2017.

[46] P. Stevenhagen, *The Number of Real Quadratic Fields Having Units of Negative Norm*. Experiment. Math. **2** (1993), no. 2, 121–136. http://doi.org/10.1080/10586458.1993.10504272

[47] P. Stevenhagen, *Rédei-matrices and applications*. In *Number theory (Paris, 1992–1993)*, edited by S. David. LMS Lecture Note Ser. **215**, 245–259. http://doi.org/10.1017/CBO9780511661990.015

[48] P. Stevenhagen, *Redei reciprocity, governing fields, and negative Pell*. To appear in Math. Proc. Camb. Phil. Soc. http://doi.org/10.1017/S0305004121000335

[49] P. Swinnerton-Dyer, *2-descent through the ages*. In *Ranks of Elliptic Curves and Random Matrix Theory*, edited by J. B. Conrey, D. W. Farmer, F. Mezzadri, and N. C. Snaith, LMS Lecture Notes Ser. **341**, 345–356. http://doi.org/10.1017/CBO9780511735158.023

[50] P. Swinnerton-Dyer, *The effect of twisting on the* 2*-Selmer group*. Math. Proc. Camb. Phil. Soc. **145** (2008), 513–526. http://doi.org/10.1017/S0305004108001588

[51] C. Tudesq, *Majoration de la loi locale de certaines fonctions additives*. (French) [Majorization of the local law of certain additive functions]. Arch. Math. (Basel) **67** (1996), no. 6, 465–472. http://doi.org/10.1007/BF01270610

[52] P. Turán, *On a Theorem of Hardy and Ramanujan*. J. London Math. Soc. **9** (1934), 274–276. http://doi.org/10.1112/jlms/s1-9.4.274

[53] A. Walfisz, *Zur additiven Zahlentheorie. II.* (German) [On additive number theory. II.]. Math. Z. **40**, 592-–607. http://doi.org/10.1007/BF01218882

[54] M. Watkins, *Notes on 4-Selmer and 8-class ranks*. Preprint, 2020.

[55] M. Xiong, *On Selmer groups of quadratic twists of elliptic curves with a two-torsion over* $\mathbf{Q}$. Mathematika **59** (2013), 303–319. http://doi.org/10.1112/S0025579312001143

[56] M. Xiong, A. Zaharescu, *Distribution of Selmer groups of quadratic twists of a family of elliptic curves*. Adv. Math. **219** (2008), 523–553. http://doi.org/10.1016/j.aim.2008.05.005

[57] G. Yu, *Rank 0 Quadratic Twists of a Family of Elliptic Curves*. Compositio Math. **135** (2003), 331–356. http://doi.org/10.1023/A:1022258905572

[58] G. Yu, *On the quadratic twists of a family of elliptic curves*. Mathematika **52** (2005), 139–154. http://doi.org/10.1112/S0025579300000413