PRIMES OF THE FORM $p^2 + nq^2$

BEN GREEN AND MEHTAAB SAWHNEY

Abstract. Suppose that $n \equiv 0$ or $n \equiv 4 \bmod 6$. We show that there are infinitely many primes of the form $p^2 + nq^2$ with both p and q prime, and obtain an asymptotic for their number. In particular, when $n = 4$ we verify the 'Gaussian primes conjecture' of Friedlander and Iwaniec.

We study the problem using the method of Type I/II sums in the number field $\mathbf{Q}(\sqrt{-n})$. The main innovation is in the treatment of the Type II sums, where we make heavy use of two recent developments in the theory of Gowers norms in additive combinatorics: quantitative versions of so-called concatenation theorems, due to Kuca and to Kuca–Kravitz-Leng, and the quasipolynomial inverse theorem of Leng, Sah and the second author.

Contents

1. Introduction

Our main result establishes an asymptotic count for pairs of primes x and y such that $x^2 + ny^2$ is prime.

Theorem 1.1. *Suppose that n is a positive integer with $n \equiv 0$ or $n \equiv 4 \bmod 6$. Let $W \in C_0^\infty(\mathbf{R}^2)$. Let $N > 1$ be a parameter. Then*

$$\sum_{x,y\in\mathbf{Z}} \Lambda(x)\Lambda(y)\Lambda(x^2+ny^2)W\left(\frac{x}{N},\frac{y}{N}\right) = \kappa_n N^2\left(\int_{\mathbf{R}^2} W\right) + O_W\left(\frac{N^2(\log\log N)^2}{\log N}\right),$$

where

$$\kappa_n := \lim_{X\to\infty} \prod_{\substack{p\leqslant X\\(-n|p)=1}} \frac{p(p-3)}{(p-1)^2} \prod_{\substack{p\leqslant X\\(-n|p)\neq 1}} \frac{p}{(p-1)}, \tag{1.1}$$

with $(\cdot \mid p)$ the Legendre symbol. In particular, if $n \equiv 0$ or $n \equiv 4 (\bmod 6)$ then there are infinitely many primes of the form $x^2 + ny^2$ with x, y prime.

Remarks. Here, and throughout the paper, the von Mangoldt function is defined by $\Lambda(n) = \log p$ if $n = p^k$ is a prime power, and by $\Lambda(-n) = \Lambda(n)$ if $n < 0$.

From the statement of Theorem 1.1 one can easily derive asymptotics for the number of x, y in a box such as $[0, N] \times [0, N]$ for which x, y and $x^2 + ny^2$ are all prime, by approximating $1_{[0,1]^2}$ by a smooth function W. We leave the details to the reader. It can be seen from the arguments of Section 4.1 that, under the assumption that W is supported in $B_{10}(0)$, the implied constant in the $O_W(\cdot)$ in Theorem 1.1 can be taken to be $\ll \sup_{0 \leqslant j \leqslant 3} \|\partial^{(j)} W\|_\infty$, by which we mean the maximum over all derivatives of W of order at most 3.

If one wanted an asymptotic over the range $x^2 + ny^2 \leqslant N^2$ then one could use Proposition 4.1 directly, without having to pass to a smooth weight W. This case, when further specialized to $n = 4$, recovers the 'Gaussian primes conjecture' in the form stated by Friedlander and Iwaniec [13, Conjecture 1.1]. We remark that constant κ_n may appear different than as stated in [13]; they are equivalent due to the class number formula which in the case $n = 4$ is the statement that $\frac{\pi}{4} = L(1, \chi_{-4}) = \lim_{X \to \infty} \prod_{p \leqslant X : p \equiv 1 (\mathrm{mod}\, 4)} \left(1 - \frac{1}{p}\right)^{-1} \prod_{p \leqslant X : p \equiv 3 (\mathrm{mod}\, 4)} \left(1 + \frac{1}{p}\right)^{-1}$.

In general the infinite product (1.1) is only conditionally convergent. It has the form predicted by heuristics of Bateman–Horn type, that is to say a product of natural archimedean and p-local factors. One can state an equivalent result with an absolutely convergent product by dividing through by the class number formula for the field $\mathbf{Q}(\sqrt{-n})$.

Finally, the argument could surely be modified in a straightforward manner to handle arbitrary positive definite binary quadratic forms over $\mathbf{Z}$ in place of $x^2 + ny^2$, but this would introduce even more notation to a paper which is already fairly heavy going. It should also not be difficult to modify the argument to prove, for example, that there are infinitely many primes $x^2 + y^2$ with x and $y - 1$ prime.

1.1. **Prior results.** Famously, it was stated by Fermat and proven by Euler that every prime $p \equiv 1 (\mathrm{mod}\, 4)$ can be written as $x^2 + y^2$, and in particular there are infinitely many primes of the form $x^2 + y^2$. Fermat also asked about prime values of $x^2 + ny^2$ for various $n \geqslant 2$, and made analogous conjectures when $n = 2, 3, 5$. The first two of these were also proven by Euler, and the third by Lagrange. The book [2] (which inspired the title and some of the content of this paper) is devoted to primes of the form $x^2 + ny^2$ from an algebraic perspective, linking this to the development of central topics in number theory such as class field theory. We note in particular that it was shown by Weber in 1882 [49] that there are infinitely many primes of the form $x^2 + ny^2$ for any $n \geqslant 1$.

Our main result falls into a long line of work considering refinements of such statements in which some restriction is placed on the coordinates x, y. Most of these have been concerned with the case $n = 1$, that is to say with the form $x^2 + y^2$. Work of Fouvry and Iwaniec [9] handled the case where x is prime. Celebrated work of Friedlander and Iwaniec [10, 11], showed that one may restrict x to be a perfect square (the first polynomial taking values in polynomially thin set shown to capture primes). This was refined in work of Heath–Brown and Li [24], to force x to be a square of a prime. Pratt [46] has proven that x may be forced to lie in a set where one excludes 3 digits from its decimal expansion. We also mention two recent works of Merikoski [38, 39]; in particular, in [38] the variable x is restricted to an arbitrary set S of integers satisfying $|S \cap \{1, \dots N\}| \gg N^{1-\delta}$ for some $\delta > 0$. The work of Fouvry and Iwaniec was generalised to $x^2 + ny^2$ for arbitrary $n \geqslant 1$ in [22].

Short of proving that there are infinitely many primes of the form $x^2 + 4$ (with x prime), for parity reasons one cannot hope to show that there are infinitely many primes of the form $x^2 + y^2$ with *both* x and y prime, and for related reasons (mod 6) one should restrict the search for primes of the form $x^2 + ny^2$ with x, y prime to the case $n \in \{0, 4\} (\mathrm{mod}\, 6)$.

Friedlander and Iwaniec [13] considered the problem in the case $n = 4$ and proved a lower bound for the number of (x, y) with x prime, y having at most 7 prime factors, and $x^2 + 4y^2$ prime.

Theorem 1.1 uses relatively little about the specific form of the functions $\Lambda(x), \Lambda(y)$, and similar techniques may be used for other problems connected with the coordinates of Gaussian primes. Perhaps the most interesting of these, which we will address in [16], is to give asymptotic counts for corners $(x+iy, (x+d)+iy, x+i(y+d))$ in the Gaussian primes.

1.2. **Outline of the argument.** For most of the rest of the introductory discussion we specialise to the case $n = 4$. In this case, Theorem 1.1 is closely related to a statement involving Gaussian primes (that is to say primes in $\mathbf{Z}[i]$) using the fact that $z = x + iy$, $y \neq 0$, is a Gaussian prime if and only $N_{\mathbf{Q}(i)/\mathbf{Q}}(z) = x^2 + y^2$ is a rational prime. In particular, our main theorem implies that there are infinitely many Gaussian primes $x + 2iy$ with x, y prime.

For the sake of this introductory discussion we will consider the problem of estimating

$$\sum_{x^2+4y^2 \leqslant X} f(x)f'(y)1_{x^2+4y^2 \text{ prime}}. \tag{1.2}$$

In the case $f = f' = \Lambda$, this is a special case of Theorem 1.1 with $N = X^{1/2}$, W (a smooth approximation to) the cutoff $1_{u^2+nv^2 \leqslant 1}$, and with the von Mangoldt weight $\Lambda(x^2 + 4y^2)$ (which is essentially always $\log X$) renormalised to 1.

Much of our argument is relevant for fairly general f, f'. Noting that (1.2) is bilinear in f and f', one can hope in any given situation of interest to write f, f' as sums of relatively easily-understood 'main terms' plus error terms for which one hopes that the sum (1.2) is small. We are then left with the task of finding verifiable criteria on f, f' which allow us to assert that (1.2) is small.

The first key idea is to observe that (1.2) is a sum

$$\sum_{\gamma \in \mathbf{Z}[i]^*} w(\gamma), \tag{1.3}$$

where here $\mathbf{Z}[i]^*$ denotes the set of Gaussian primes and $w : \mathbf{Z}[i] \to \mathbf{C}$ is defined by $w(x + 2iy) = f(x)f'(y)$ and $w(x + iy) = 0$ for y odd, and we assume throughout the following discussion that $w(\gamma)$ is supported where $|\gamma|^2 \leqslant X$.

We then examine this sum over Gaussian primes using so-called Type I and Type II sums. For a general introduction to this technique over the rational integers, see for instance [28, Chapter 13] (where the authors do not use the terms Type I/II sum, but the relevant sums are (13.16) and (13.17)) or the comprehensive recent analysis [8]. Adapting the relevant techniques from the rational integers $\mathbf{Z}$ to the Gaussian integers $\mathbf{Z}[i]$ presents only minor technical obstacles in our setting. See [23, Section 3] for a very similar type of argument in the ring $\mathbf{Z}[2^{1/3}]$, and [25, 37] for similar arguments in fields not necessarily enjoying unique factorisation.

For the purposes of this discussion, Type I sums are roughly of the form

$$\sum_{\substack{\ell \in \mathbf{Z}[i] \\ |\ell|^2 \sim L}} \Big| \sum_{\substack{m \in \mathbf{Z}[i] \\ \ell | m}} w(m) \Big| \tag{1.4}$$

for some positive $L \in \mathbf{R}$, which we informally call the *level*.

Type II sums are bilinear sums of the form

$$\sum_{\substack{\ell, m \in \mathbf{Z}[i] \\ |\ell|^2 \sim L}} \alpha_\ell \beta_m w(\ell m), \tag{1.5}$$

where here the coeffients α_ℓ, β_m will satisfy reasonable boundedness conditions but can otherwise be fairly arbitrary.

The estimates we seek for these sums are bounds of the form $O(X(\log X)^{-B})$ for some large B, which represent savings over the trivial bounds by large powers of $\log X$. If we have such bounds,

we informally say that we have Type I (or Type II) information at level L. (Precise statements may be found in Section 3.)

If one has Type I and Type II information at sufficient levels, it is possible to obtain asymptotics for sums over primes (1.3) by sieve methods, details of which may be found in Section 3. Any such argument needs Type I/II information in sufficiently large ranges. The question of exactly what is necessary or sufficient for this purpose is rather complicated (and, over the integers, has been comprehensively addressed in the recent work [8]).

We will use an argument of Duke, Friedlander, and Iwaniec [4, Section 6] which (adapted to the Gaussian primes) shows that, at least for 1-bounded functions f, f', it is enough to have Type I information up to level $X^{1/2-o(1)}$, and Type II information up to $X^{1/3-o(1)}$.

The task, then, is to obtain the relevant Type I and II information. We are able to do this under the assumption that appropriate *Gowers norms* of f or f' are small. The Gowers norms are central objects in modern additive combinatorics; see Section 4.2 for an introduction. Assuming bounds for suitable Gowers norms of f or f', we are able to obtain Type I information up to level $X^{1/2-o(1)}$ and Type II information for levels between $X^{o(1)}$ and $X^{1/2-o(1)}$, which is sufficient for variants of the sieve arguments of Duke, Friedlander and Iwaniec to apply in order to show that (1.3) is small. One point to note here is that, while we have comfortably enough Type II information, the Type I range is just barely enough for the requirements of the Duke–Friedlander–Iwaniec sieve. Worse, the functions f and f' of interest to us for our main theorem are not 1-bounded but come from the von Mangoldt function (with a suitable main term subtracted), and the most naïve application of the Duke–Friedlander–Iwaniec machinery fails to give a usable bound. To get around this issue we require an additional upper bound sieve (of dimension 3). We remark that obtaining Type I information at level $X^{1/2}$ in our setting remains an interesting open question. In fact, even obtaining bounds at this level without the absolute values in (1.4) seems challenging. (A bound of this latter type would be interesting, being a simple case of what is called Type I_2 information. Sufficient information in this direction would allow one to obtain an arbitrary logarithmic savings in Theorem 1.1.)

The most novel part of our work lies in showing that the stated Type I and Type II information can indeed be obtained from suitable Gowers norm control on the functions f and f'. Precise statements of these results are Proposition 4.2 and Proposition 4.3 respectively. The proof of Proposition 4.3 implicitly uses a large number of applications of the Cauchy-Schwarz inequality, organised via what are known as *concatenation* theorems for Gowers-type norms. The first concatenation results were obtained by Tao and Ziegler [48], but those results are rather qualitative. More recently, starting with the work of Peluse and Prendiville [44] and Peluse [43], more quantitative variants have been developed. For our application, we need the very recent work of Kuca [33] and Kravitz–Kuca–Leng [32].

In order to apply these results to prove Theorem 1.1, we need to show that for suitable functions f and f' related to the von Mangoldt function, the relevant Gowers norms are small. We work with $f = f' = \Lambda - \Lambda_{\text{Cramér}}$, where $\Lambda_{\text{Cramér}}$ is a certain low-complexity approximant to the von Mangoldt function defined in (1.6) below. Statements of this type were first shown in a series of works of the first author with Tao [17–19] and with Tao and Ziegler [20]. More recent work of Tao and Teräväinen [47] gave quantitative bounds relying on effective bounds for the inverse theorem for the Gowers norms established by Manners [36]. For our application, however, we need savings of large powers of $\log X$ over the trivial bound. Bounds of this strength were only recently established by Leng [34], whose work relies on the quasipolynomial inverse theorem for the Gowers norms due to Leng, Sah and the second author [35].

The final task of the paper is then to obtain the main term in Theorem 1.1 by evaluating (1.2) with $f = f' = \Lambda_{\text{Cramér}}$. This is a task for classical methods of multiplicative number theory, the

main input being the prime ideal theorem. However, it takes a little work to get all the details in order.

In the above outline we had $n = 4$. To deal with more general values of n, we use the field $K = \mathbf{Q}(\sqrt{-n})$ in place of the Gaussian field. In general, these fields do not enjoy unique factorisation and in the above discussion one must work with ideals rather than with elements of the ring of integers $\mathcal{O}_K$. In fact, for technical reasons related to units it is best to work in this language even in the Gaussian setting (which does have unique factorisation). The fact that not all ideals are principal creates some extra difficulties, but they turn out not to be too serious and are resolved in much the same way that similar difficulties were overcome in [25, 37] (for instance).

1.3. **Organization.** In Section 2 we review some of the basic arithmetic of $K = \mathbf{Q}(\sqrt{-n})$ and recall some simple lemmas about the ring of integers $\mathcal{O}_K$ and the ideals in this ring.

In Section 3, we develop the relevant sieve machinery allowing us to estimate sums over primes such as (1.3) using Type I/II information. Here we follow the arguments of Duke, Friedlander and Iwaniec rather closely, albeit in the number field setting. As explained above, we will additionally require the input of an upper bound sieve of dimension 3 in two variables.

In Section 4 we begin by giving a technical reduction of our main theorem, which is the form we shall actually prove; essentially, this consists in converting from Cartesian to polar coordinates. Then, we recall the definitions and basic properties of the Gowers U^k-norms, which will feature heavily in the paper from this point on. We will also give the outline proof of our main theorem, leaving the proof of the key Type I and II estimates, as well as the main input from analytic number theory, to later sections.

In Section 5, we recall a variant of the Gowers norms due to Peluse [43], which we call the Gowers–Peluse norms. These are the appropriate generalisations of the Gowers norms needed for the concatenation machinery we will require. In this section we also state the main concatenation estimate of Kuca, Kravitz, and Leng [32].

In Section 6, we show that Type I information is controlled by Gowers norms and in Section 7 we do the same for Type II information.

In Section 8, we complete the proof of Theorem 1.1 by evaluating the main term in the asymptotic.

Finally, we provide four appendices. In Appendix A, we prove various properties of Gowers box norms. In Appendix B, we provide self-contained proofs of the main concatenation results. In Appendix C, we briefly provide details of the large sieve in higher dimensions. In Appendix D we collect some standard arithmetic estimates.

1.4. **Notation.** There will be a small number of fixed objects throughout the paper, as follows:

- n will always be the same n as in the main theorem, and we think of it as fixed;
- K will always be $\mathbf{Q}(\sqrt{-n})$, regarded as embedded into $\mathbf{C}$, except in the self-contained Section 3.1 where it could be any number field (but for our application will be $\mathbf{Q}(\sqrt{-n})$);
- X is some positive integer parameter, which we will always assume sufficiently large in terms of n without further comment.

For real $N \geqslant 1$ we write $[N] = \{x \in \mathbf{Z} : 1 \leqslant x \leqslant N\}$ and $[\pm N] = \{x \in \mathbf{Z} : -N \leqslant x \leqslant N\}$.

It is convenient to use a very minor modification of the von Mangoldt function which ignores the prime powers and is symmetric. Thus define $\Lambda'(n) = \log|n|$ if $|n|$ is a prime, and $\Lambda'(n) = 0$ otherwise.

As mentioned above, we will use a 'low-complexity' model for the primes, whose main task is to encode the fact that the primes are essentially supported only on residue classes coprime to q. Specifically, for $x \in \mathbf{Z}$ and $Q \in \mathbf{Z}^+$, we define the Cramér model at level Q to be

$$\Lambda_{\text{Cramér},Q}(x) = \prod_{p \leqslant Q} \left(1 - \frac{1}{p}\right)^{-1} 1_{(x,p)=1}. \tag{1.6}$$

Throughout the paper we take

$$Q := \exp(\log^{1/10}(X^{1/2})). \tag{1.7}$$

Note that this is the same choice made in [47, Equation (1.1)] and in [34, Section 5] if one takes $N = X^{1/2}$ in those papers. This is the reason for the inclusion of $X^{1/2}$ in (1.7), which otherwise looks peculiar. We will write $\Lambda_{\text{Cramér}} = \Lambda_{\text{Cramér},Q}$ from now on.

As is standard, we let $e(\theta) = e^{2\pi i\theta}$.

We use standard asymptotic notation throughout. We use $A \sim B$ to denote that $B/2 \leqslant A < 2B$ (e.g. A lies within a factor of 2 of B). Given functions $f = f(x)$ and $g = g(x)$, we write $f = O(g)$, $f \ll g$, $g = \Omega(f)$, or $g \gg f$ to mean that there is a constant C such that $|f(x)| \leqslant Cg(x)$ for sufficiently large n. We write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \ll g$ and $g \ll f$, and write $f = o(g)$ to mean $f(x)/g(x) \to 0$ as $x \to \infty$. Subscripts indicate dependence on parameters.

From Section 5 onwards, we will use the language of probability measures on $\mathbf{Z}$. For us, this simply means a finitely-supported function $\mu : \mathbf{Z} \to [0,1]$ with $\sum_x \mu(x) = 1$. In this paper all our probability measures will be symmetric, meaning that $\mu(x) = \mu(-x)$. The convolution $\mu_1 * \mu_2$ of two probability measures is defined by $(\mu_1 * \mu_2)(x) := \sum_y \mu_1(y)\mu(x-y)$, and it is also a probability measure. If $\mu_1, \dots, \mu_\ell$ are probability measures then we write $*_{i=1}^{\ell} \mu_i$ for $\mu_1 * \cdots * \mu_\ell$. If $\mu_1 = \cdots = \mu_\ell = \mu$, we write $\mu^{(\ell)}$ for this measure. Given a (multi)set $S \subseteq \mathbf{Z}$, we let U_S denote the uniform measure on this set. Finally, δ_0 denotes the delta-mass at 0, thus $\delta_0(0) = 1$ and $\delta_0(x) = 0$ for $x \neq 0$. Observe that δ_0 is an identity for convolution in the sense that $\mu * \delta_0 = \mu$ for any probability measure μ. The notation $\mathbb{E}_{x\sim\mu} f(x)$ means the same as $\sum_x f(x)\mu(x)$.

We remark that it is perhaps controversial to use the letter μ for something other than the Möbius function in an analytic number theory paper. However, the Möbius function only appears in Section 3 of our paper, where there are no probability measures, and so there should be no danger of confusion.

When working in number fields, we will use the term *rational prime* to mean a prime number in $\mathbf{Z}$. This is to distinguish from other uses of the word prime (prime ideals or prime elements of the ring of integers).

Finally, if $\ell \in \mathbf{Z}$ then we define the character $\chi_\infty^{(\ell)} : \mathbf{C}^* \to \mathbf{C}^*$ by

$$\chi_\infty^{(\ell)}(z) := (z/|z|)^\ell. \tag{1.8}$$

1.5. **Acknowledgements.** We thank Sarah Peluse and Terry Tao for helpful conversations. We also thank Roger Heath-Brown, Emmanuel Kowalski and James Maynard for comments pertaining to the literature concerning the prime number theorem for ideals, and Noah Kravitz, James Leng, Jori Merikoski, Cédric Pilatte, Joni Teräväinen and Katharine Woo for comments on a draft of the paper.

BG is supported by Simons Investigator Award 376201. This research was conducted during the period MS served as a Clay Research Fellow. Portions of this research were conducted while MS was visiting Oxford and Cambridge both of which provided excellent working conditions. Finally, the authors thank ICMS and the organisers of the workshop on Additive Combinatorics (July 2024) which provided the initial impetus for this collaboration.

2. Number fields and weight functions

In this section we collect some basic facts about number fields, particularly the imaginary quadratic field $K = \mathbf{Q}(\sqrt{-n})$. We also introduce the key notion of a weight function of product form, Definition 2.2.

Given a number field K, denote by $\mathcal{O}_K$ its ring of integers. This is a lattice satisfying $\mathbf{Z}[\sqrt{-n}] \subseteq \mathcal{O}_K \subseteq \frac{1}{2n}\mathbf{Z}[\sqrt{-n}]$. Note in particular that

$$(2n) \subseteq \mathbf{Z}[\sqrt{-n}], \tag{2.1}$$

where $(2n)$ is the ideal generated by $2n$. Write $\mathcal{O}_K^*$ for the units in $\mathcal{O}_K$. This will consist of ± 1 unless the squarefree part of n is 1 (in which case $\mathcal{O}_K^* = \{\pm 1, \pm i\}$) or 3 (in which case $\mathcal{O}_K^*$ consists of the sixth roots of unity).

Ideals. Denote by $\mathrm{Ideals}(\mathcal{O}_K)$ the semi-group of non-zero ideals in $\mathcal{O}_K$ and by $\mathrm{Ideals}^0(\mathcal{O}_K)$ the non-zero principal ideals. We will also encounter fractional ideals. If $\alpha \in K^*$ then it is convenient to write $(\alpha) = \alpha\mathcal{O}_K$ and call this the fractional ideal generated by α. Note that if $\alpha \in \mathcal{O}_K \setminus \{0\}$ then (α) is a principal ideal, and all principal ideals are of this form.

Fraktur letters $\mathfrak{a}, \mathfrak{b}, \mathfrak{d}$ will always denote ideals in $\mathcal{O}_K$, with the letters $\mathfrak{p}, \mathfrak{q}, \mathfrak{r}$ being reserved for prime ideals. We reserve the letter $\mathfrak{c}$ for particular fixed choices of fractional ideals in K, as described in Lemma 2.1 below.

Recall that we may define the norm $N\mathfrak{a}$ of any non-zero fractional ideal $\mathfrak{a}$, and this takes values in $\mathbf{Q}^+$.

Ideal classes. Denote by $C(K)$ the ideal class group of K. We will denote elements of $C(K)$ by g, and will write $[g]$ for the set of ideals in $\mathrm{Ideals}(\mathcal{O}_K)$ in the class represented by g. Sometimes it is necessary to consider fractional ideals and their classes in $C(K)$. Indeed it will be convenient throughout the paper to fix some fractional ideal representatives of each class, as detailed in the following lemma.

Lemma 2.1. *Let $K = \mathbf{Q}(\sqrt{-n})$. We may choose fractional ideals $\mathfrak{c}_g$, $\mathfrak{c}'_g$ for $g \in C(K)$ with the following properties:*

(i) *$\mathfrak{c}_g, \mathfrak{c}'_g$ are in the ideal class corresponding to g;*
(ii) *$\mathfrak{c}_g\mathfrak{c}'_{g^{-1}}$ (which is principal) is generated by $\gamma_g \in \mathbf{Q}$, a positive rational of height $O_n(1)$;*
(iii) *The set $\Pi_g := \{z \in K^* : (z)\mathfrak{c}_g \subseteq \mathcal{O}_K\}$ is a finite index sublattice of $\mathbf{Z}[\sqrt{-n}]$, and similarly for $\Pi'_g := \{z \in K^* : (z)\mathfrak{c}'_g \subseteq \mathcal{O}_K\}$.*

Proof. First of all, for each $g \in C(K)$ pick some $\mathfrak{b}_g \in [g]$ with norm $\leqslant \frac{4}{\pi}\sqrt{n}$; such ideals exist by the Minkowski bound. For each g, $\mathfrak{b}_g\mathfrak{b}_{g^{-1}}$ is a principal ideal. Set $\mathfrak{b}'_{g^{-1}} := \overline{\mathfrak{b}_g\mathfrak{b}_{g^{-1}}}\mathfrak{b}_{g^{-1}}$. Then $N\mathfrak{b}'_{g^{-1}} \ll n^{3/2}$, and $\mathfrak{b}_g\mathfrak{b}'_{g^{-1}} = (\mathfrak{b}_g\mathfrak{b}_{g^{-1}})\overline{\mathfrak{b}_g\mathfrak{b}_{g^{-1}}}$ is generated by a rational integer of magnitude $\ll n^2$. Now, for each g, take $\mathfrak{c}_g$ to be the fractional ideal $\frac{1}{M_g}\mathfrak{b}_g$, where $M_g = O_n(1)$ is an integer such that $M_g\mathfrak{b}_g^{-1} \subseteq (2n)$. Such an M_g exists by clearing denominators in generators of $\mathfrak{b}_g^{-1}$. Define $\mathfrak{c}'_g = \frac{1}{M'_g}\mathfrak{b}'_g$ similarly. Properties (i) and (ii) are then immediate from the above discussion. (We remark that the key point in (ii) is that the generator γ_g lies in $\mathbf{Q}$, rather than merely in K; the second point is in a sense vacuous since there are only $O_n(1)$ ideal classes, however the proof yields an explicit value for this constant if desired.)

We turn to point (iii). It is clear that Π_g is an abelian group under addition. Also, since $\mathfrak{c}_g$ has an integral basis, we see by clearing denominators that there is $M = O_n(1)$ such that $M, M\sqrt{-n} \in \Pi_g$. In the other direction, note that by construction we have $\mathfrak{c}_g^{-1} \subseteq (2n)$. Therefore if $(z)\mathfrak{c}_g \subseteq \mathcal{O}_K$ then $(z) = (z)\mathfrak{c}_g\mathfrak{c}_g^{-1} \subseteq \mathcal{O}_K\mathfrak{c}_g^{-1} \subseteq (2n) \subseteq \mathbf{Z}[\sqrt{-n}]$, the second inclusion being (2.1). Hence $\Pi_g \subseteq \mathbf{Z}[\sqrt{-n}]$, and the claim follows. □

Weight functions. Throughout the paper we will be dealing with weight functions w on $\mathrm{Ideals}(\mathcal{O}_K)$. In Section 3.1 these can be quite general, but for most of the paper we will take $K = \mathbf{Q}(\sqrt{-n})$ and the weight functions will be of a special product form.

Definition 2.2. Suppose that $K = \mathbf{Q}(\sqrt{-n})$. Suppose that $f, f' : \mathbf{Q} \to \mathbf{C}$ are two functions, both supported on $\mathbf{Z}$, and let $\ell \in \mathbf{Z}$. Then we define a function $f \boxtimes_\ell f'$ on ideals as follows:

$$f \boxtimes_\ell f'(\mathfrak{a}) = \sum_{(x+y\sqrt{-n})=\mathfrak{a}} \chi_\infty^{(\ell)}(x+y\sqrt{-n}) f(x) f'(y), \tag{2.2}$$

where $\chi_\infty^{(\ell)}(z) = (z/|z|)^\ell$. Any function $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ having this form is said to be in *product form* with *frequency* ℓ. We say that w is 1-bounded if f, f' are.

Remarks. Note that the sum is empty unless $\mathfrak{a}$ is principal, so $f \boxtimes_\ell f'$ is supported on principal ideals. If $\mathfrak{a}$ is principal the the sum is over the $|\mathcal{O}_K^*|$ associates of any generator for $\mathfrak{a}$.

Further facts about $\mathbf{Q}(\sqrt{-n})$. The following further definitions and facts will be required only in Section 8. Factor

$$n = n_* r^2 \tag{2.3}$$

where n_* is squarefree and $r \in \mathbf{N}$. It is convenient to define a quantity ω by

$$\omega := \begin{cases} 1 & \text{if } n_* \equiv 1, 2 (\mathrm{mod}\ 4) \text{ and} \\ \frac{1}{2} & \text{if } n_* \equiv 3 (\mathrm{mod}\ 4). \end{cases} \tag{2.4}$$

Then it is well-known that $\mathcal{O}_K = \mathbf{Z}[\sqrt{-n_*}]$ if $\omega = 1$, whereas $\mathcal{O}_K = \mathbf{Z}[\frac{1}{2}(1+\sqrt{-n_*})]$ if $\omega = \frac{1}{2}$ (which perhaps explains the notation). Recall that the discriminant Δ is given by the formula

$$\Delta = -4\omega^2 n_*. \tag{2.5}$$

There is a real quadratic character $(\Delta \mid \cdot)$ associated to the field, where the symbol here is the Kronecker symbol, which coincides with the Legendre symbol on odd primes p. This symbol determines the splitting type of a prime p; p splits, is inert or ramifies according to whether $(\Delta \mid p) = 1, -1, 0$ respectively. We also let h_K denote the class number of K. Finally we recall the class number formula, which in our setting and language states that

$$\frac{2\pi h_K}{|\mathcal{O}_K^*||\Delta|^{1/2}} = L(1, (\Delta|\cdot)) = \prod_p \left(1 - \frac{(\Delta|p)}{p}\right)^{-1}. \tag{2.6}$$

3. Sieve setup-reduction to Type I and Type II statements

In this section we detail the main sieve-theoretic parts of the argument which allow us to reduce to proving certain Type I/II estimates.

3.1. The Duke–Friedlander–Iwaniec sieve in a number field.

Here we adapt the sieve arguments of Duke, Friedlander and Iwaniec to an arbitrary number field K. The generalisation is essentially completely routine and follows [4, Section 6] very closely, with only minor technical modifications required to work with ideals rather than integers.

Definition 3.1. Fix a weight function $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$. Given a real number $B \geqslant 1$, w satisfies Type I (with savings $(\log X)^{-B}$)) at L_1 if given $L \leqslant L_1$ then

$$\sum_{\mathfrak{d}: N\mathfrak{d} \sim L} \left| \sum_{\mathfrak{d} | \mathfrak{a}} w(\mathfrak{a}) \right| \ll_B X (\log X)^{-B} \|w\|_\infty. \tag{3.1}$$

We say that w satisfies Type II (with savings $(\log X)^{-B}$) between $[L_1, L_2]$ if given $L \in [L_1, L_2]$, and any 1-bounded sequences $\alpha_\mathfrak{a}, \beta_\mathfrak{b}$ then

$$\sum_{\mathfrak{a},\mathfrak{b}: N\mathfrak{a} \sim L} \alpha_\mathfrak{a} \beta_\mathfrak{b} w(\mathfrak{a}\mathfrak{b}) \ll_B X (\log X)^{-B} \|w\|_\infty. \tag{3.2}$$

Remark. In our applications, the parameter L_1 in Type II estimates will be purely technical and chosen to grow slightly faster than any power of logarithm.

Here is our formulation of Duke, Friedlander and Iwaniec's sieve result in the number field setting.

Lemma 3.2. *Let $A, C > 1$ be parameters, set $B := 2A + 4$, and suppose $C \geqslant B$. Furthermore suppose that $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ satisfies Type I (with savings $(\log X)^{-B}$) at $X^{1/2}(\log X)^{-C}$ and Type II (with savings $(\log X)^{-B}$) between $[(\log X)^C, X^{3/8}]$. Suppose that w is supported on ideals of norm $\leqslant X$. Then*

$$\sum_{\mathfrak{p}} w(\mathfrak{p}) + \sum_{\mathfrak{p}_1,\mathfrak{p}_2} w(\mathfrak{p}_1\mathfrak{p}_2) \ll_{A,C} X(\log X)^{-A}\|w\|_\infty,$$

where the first sum ranges over prime ideals $\mathfrak{p}$, and the second sum ranges over pairs $\mathfrak{p}_1, \mathfrak{p}_2$ of prime ideals with $N\mathfrak{p}_i \geqslant X^{1/2}e^{-(\log\log X)^2}$ for $i = 1, 2$.

Remark. To explain the meaning of this result, we note that the sum $\sum_{\mathfrak{p}} w(\mathfrak{p})$ is the one we are ultimately interested in. The second sum, which is over ideals $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2$ with two similarly-sized prime ideal factors, will later be bounded using an upper bound sieve.

Proof. Since the statement is homogeneous in w, we may assume that w is 1-bounded. By adjusting implicit constants in the statement, we may assume that X is larger than an absolute constant depending on B. Let

$$D = X^{1/2}(\log X)^{-C},\ u = (\log X)^C \text{ and } \ z = X^{1/2}e^{-(\log\log X)^2}. \tag{3.3}$$

We put a total ordering on the prime ideals in $\mathcal{O}_K$ by declaring that $\mathfrak{p}_1 < \mathfrak{p}_2$ if $N\mathfrak{p}_1 < N\mathfrak{p}_2$, and then ordering the prime ideals with the same norm arbitrarily. (The same technical device occurs in [25, Section 4] or [37, Section 6], for example.) Note that there are at most $[K : \mathbf{Q}]$ prime ideals of a given norm. By an *up-set* we mean a set I of prime ideals with the property that if $\mathfrak{p}_1 \in I$, and if $\mathfrak{p}_2 > \mathfrak{p}_1$, then $\mathfrak{p}_2 \in I$. Since the prime ideals are well-ordered by $<$, every up-set is of the form $I(\mathfrak{p}) := \{\mathfrak{p}' : \mathfrak{p}' \geqslant \mathfrak{p}\}$. However, for technical reasons during the proof it is sometimes useful to consider up-sets specified in the form $I(t) := \{\mathfrak{p} : N\mathfrak{p} \geqslant t\}$, where $t \in \mathbf{R}^+$. There should hopefully be no confusion over which definition is being used at any given point.

Let $\mathcal{A} \subseteq \mathrm{Ideals}(\mathcal{O}_K)$ and let I be an up-set. We define

$$S(\mathcal{A}, I) = \sum_{\substack{\mathfrak{a}\in\mathcal{A} \\ \mathfrak{p}|\mathfrak{a} \implies \mathfrak{p}\in I}} w(\mathfrak{a}).$$

Given a set $\mathcal{A}$ of ideals and an ideal $\mathfrak{d}$, we denote by $\mathcal{A}_{\mathfrak{d}}$ the subset of $\mathcal{A}$ consisting of ideals divisible by $\mathfrak{d}$.

Now note that if I_1, I_2 are up-sets with $I_1 \subseteq I_2$ then we have the *Buchstab identity*

$$S(\mathcal{A}, I_1) = S(\mathcal{A}, I_2) - \sum_{\mathfrak{p}\in I_2\setminus I_1} S(\mathcal{A}_{\mathfrak{p}}, I(\mathfrak{p})).$$

To see this, observe that if $w(\mathfrak{a})$ appears in the sum $S(\mathcal{A}, I_2)$ but not in $S(\mathcal{A}, I_1)$ then $\mathfrak{a}$ is divisible by a unique smallest prime ideal $\mathfrak{p}$ with $\mathfrak{p} \in I_2 \setminus I_1$. Of course, $\mathfrak{a}$ lies in $\mathcal{A}_{\mathfrak{p}}$, and since $\mathfrak{p}$ was chosen to be smallest, all other prime ideal divisors of $\mathfrak{a}$ are $\geqslant \mathfrak{p}$, or in other words lie in $I(\mathfrak{p})$.

For notational brevity take $\mathcal{C} = \mathrm{Ideals}(\mathcal{O}_K)$. By two applications of the Buchstab identity we have

$$\begin{aligned} S(\mathcal{C}, I(z)) &= S(\mathcal{C}, I(u)) - \sum_{\mathfrak{p}\in I(u)\setminus I(z)} S(\mathcal{C}_{\mathfrak{p}}, I(\mathfrak{p})) \\ &= S(\mathcal{C}, I(u)) - \sum_{\mathfrak{p}\in I(u)\setminus I(z)} \Bigg(S(\mathcal{C}_{\mathfrak{p}}, I(u)) - \sum_{\mathfrak{q}\in I(u)\setminus I(\mathfrak{p})} S(\mathcal{C}_{\mathfrak{p}\mathfrak{q}}, I(\mathfrak{q})) \Bigg) \end{aligned}$$

$$= S(\mathcal{C}, I(u)) - \sum_{u \leqslant N\mathfrak{p} < z} S(\mathcal{C}_{\mathfrak{p}}, I(u)) + \sum_{\substack{N\mathfrak{p} \geqslant u \\ N\mathfrak{q} < z \\ \mathfrak{p} < \mathfrak{q}}} S(\mathcal{C}_{\mathfrak{p}\mathfrak{q}}, I(\mathfrak{p})). \tag{3.4}$$

Note that in the last step here, we switched the roles of $\mathfrak{p}$ and $\mathfrak{q}$, and unpacked the definitions of the $I()$ terms.

On the other hand, note that $z^3 > X$ and recall that if $w(\mathfrak{a})$ is non-zero then $N\mathfrak{a} \ll X$. Therefore the sum $S(\mathcal{C}, I(z))$ contains only two types of term $w(\mathfrak{a})$: those with $\mathfrak{a}$ prime, and those with $\mathfrak{a}$ a product of two prime ideals, both of norm $\geqslant z$. That is,

$$S(\mathcal{C}, I(z)) = \sum_{\mathfrak{p}: N\mathfrak{p} \geqslant z} w(\mathfrak{p}) + \sum_{\mathfrak{p}_1, \mathfrak{p}_2} w(\mathfrak{p}_1\mathfrak{p}_2), \tag{3.5}$$

where the second sum is over $\mathfrak{p}_1, \mathfrak{p}_2$ with $N\mathfrak{p}_i \geqslant z = X^{1/2}e^{-(\log\log X)^2}$, that is to say the second sum here is exactly the one appearing in Lemma 3.2. The constraint $N\mathfrak{p} \geqslant z$ will easily be removed by trivial bounds, so our task is to bound $S(\mathcal{C}, I(z))$, which we shall do via (3.4).

Step 1: Deploying Type I information.

The first step is to bound the first two terms in (3.4), $S(\mathcal{C}, I(u)) - \sum_{u \leqslant N\mathfrak{p} < z} S(\mathcal{C}_{\mathfrak{p}}, I(u))$, via Type I information, or in other words by use of (3.1). Define $P(u) := \prod_{\mathfrak{p}: N\mathfrak{p} < u} \mathfrak{p}$ and let Π_u be the set of all $\mathfrak{a}$ with at most one prime factor of norm $\geqslant u$. Then, unwinding definitions, we have that

$$S(\mathcal{C}, I(u)) - \sum_{u \leqslant N\mathfrak{p} < z} S(\mathcal{C}_{\mathfrak{p}}, I(u)) = \sum_{(\mathfrak{a}, P(u)) = 1} w(\mathfrak{a}) \Big(1 - \sum_{\substack{u \leqslant N\mathfrak{p} < z \\ \mathfrak{p} \mid \mathfrak{a}}} 1 \Big). \tag{3.6}$$

Denote by μ_K the Möbius function on (ideals of) K. Then we have

$$\sum_{\substack{\mathfrak{d} \mid (P(z), \mathfrak{a}) \\ \mathfrak{d} \in \Pi_u}} \mu_K(\mathfrak{d}) = 1_{(\mathfrak{a}, P(u)) = 1} \cdot \Big(1 - \sum_{\substack{u \leqslant N\mathfrak{p} < z \\ \mathfrak{p} \mid \mathfrak{a}}} 1 \Big).$$

Substituting into (3.6) yields

$$S(\mathcal{C}, I(u)) - \sum_{u \leqslant N\mathfrak{p} < z} S(\mathcal{C}_{\mathfrak{p}}, I(u)) = \sum_{\substack{\mathfrak{d} \mid P(z) \\ \mathfrak{d} \in \Pi_u}} \mu_K(\mathfrak{d}) \sum_{\mathfrak{d} \mid \mathfrak{a}} w(\mathfrak{a}). \tag{3.7}$$

We now break the sum into two parts based on where $N\mathfrak{d} \leqslant D$ and else when $N\mathfrak{d} > D$. For the first of these we have that

$$\Big| \sum_{\substack{\mathfrak{d} \mid P(z) \\ N\mathfrak{d} \leqslant D \\ \mathfrak{d} \in \Pi_u}} \mu_K(\mathfrak{d}) \sum_{\mathfrak{d} \mid \mathfrak{a}} w(\mathfrak{a}) \Big| \leqslant \sum_{N\mathfrak{d} \leqslant D} \Big| \sum_{\mathfrak{d} \mid \mathfrak{a}} w(\mathfrak{a}) \Big| \ll X(\log X)^{-B+1} \ll X(\log X)^{-A}; \tag{3.8}$$

here we have applied dyadic summation in $N\mathfrak{d}$ and the Type I estimate (3.1) at each dyadic scale, and recall that w is 1-bounded.

For the sum over $N\mathfrak{d} > D$, we use the trivial bound $|w(\mathfrak{a})| \leqslant 1_{N\mathfrak{a} \leqslant X}$ (that is, w is 1-bounded on supported on ideals of norm $\leqslant X$), obtaining

$$\Big| \sum_{\substack{\mathfrak{d} \mid P(z) \\ N\mathfrak{d} > D \\ \mathfrak{d} \in \Pi_u}} \mu_K(\mathfrak{d}) \sum_{\mathfrak{d} \mid \mathfrak{a}} w(\mathfrak{a}) \Big| \leqslant \sum_{\substack{\mathfrak{d} \mid P(z) \\ D < N\mathfrak{d} \leqslant X \\ \mathfrak{d} \in \Pi_u}} \Big| \sum_{\substack{N\mathfrak{a} \leqslant X \\ \mathfrak{d} \mid \mathfrak{a}}} 1 \Big| \ll \sum_{\substack{\mathfrak{d} \mid P(z) \\ D < N\mathfrak{d} \leqslant X \\ \mathfrak{d} \in \Pi_u}} \frac{X}{N\mathfrak{d}}. \tag{3.9}$$

By dyadic decomposition, this is bounded above by

$$\ll X(\log X)\sup_{L\in[D,2X]}\sum_{\substack{\mathfrak{d}\mid P(z)\\ N\mathfrak{d}\sim L\\ \mathfrak{d}\in\Pi_u}}\frac{1}{L}. \tag{3.10}$$

Note that every $\mathfrak{d}$ appearing in the sum here is squarefree (since it divides $P(z)$) and has at most one prime factor of norm $\geqslant u$. The product of the (distinct) prime divisors of $\mathfrak{d}$ of norm $< u$ must therefore have norm at least D/z, and therefore there must be at least $t := \lfloor \log(D/z)/\log u\rfloor$ of them. Using Lemma D.3 and Lemma D.4, this implies that

$$\sum_{\substack{\mathfrak{d}\mid P(z)\\ N\mathfrak{d}\sim L\\ \mathfrak{d}\in\Pi_u}}\frac{1}{L}\ll\frac{1}{L}\sum_{\substack{\mathfrak{p}_1<\cdots<\mathfrak{p}_t\\ N\mathfrak{p}_i<u}}\frac{L}{N(\mathfrak{p}_1\cdots\mathfrak{p}_t)}\ll\frac{1}{t!}\Big(\sum_{\mathfrak{p}:N\mathfrak{p}<u}\frac{1}{N\mathfrak{p}}\Big)^t\ll\frac{(O_K(\log\log u))^t}{t!}\ll(\log X)^{-A-1}.$$

Substituting into (3.10), we see that the contribution to (3.7) from the sum over $N\mathfrak{d} > D$ is $\ll X(\log X)^{-A}\|w\|_\infty$. Combining with the estimate (3.8) for the sum over $N\mathfrak{d}\leqslant D$, we obtain the bound

$$S(\mathcal{C},I(u))-\sum_{u\leqslant N\mathfrak{p}<z}S(\mathcal{C}_{\mathfrak{p}},I(u))\ll X(\log X)^{-B+1}\ll X(\log X)^{-A}. \tag{3.11}$$

This concludes our analysis of the first two terms in (3.4) by the use of Type I sums.

Step 2: Splitting the $\mathfrak{p}$ variable.

We now proceed to bounding the third term in (3.4), that is to say $\sum_{N\mathfrak{p}\geqslant u,N\mathfrak{q}<z,\mathfrak{p}<\mathfrak{q}}S(\mathcal{C}_{\mathfrak{pq}},I(\mathfrak{p}))$. Here, we will use the Type II estimate (3.2), but first we make some preliminary manoeuvres aimed at decoupling the condition $\mathfrak{p}<\mathfrak{q}$ into conditions on $\mathfrak{p},\mathfrak{q}$ separately. We first estimate the contribution from large $\mathfrak{p}$. To this end set

$$y := X^{3/8}. \tag{3.12}$$

Then we have

$$\sum_{\substack{N\mathfrak{p}\geqslant y\\ N\mathfrak{q}<z\\ \mathfrak{p}<\mathfrak{q}}}S(\mathcal{C}_{\mathfrak{pq}},I(\mathfrak{p}))=\sum_{\substack{N\mathfrak{p}\geqslant y\\ N\mathfrak{q}<z\\ \mathfrak{p}<\mathfrak{q}}}w(\mathfrak{pq})\ll z^2\ll X(\log X)^{-A}. \tag{3.13}$$

In the first step, we observed that the only ideal $\mathfrak{a}$ of norm $\ll X$, divisible by $\mathfrak{p}$ and $\mathfrak{q}$ and with all other prime ideal factors in $I(\mathfrak{p})$, is $\mathfrak{a}=\mathfrak{pq}$ itself. For the last step, we recall the choice (3.3) of z. The estimate (3.13) is clearly acceptable in Lemma 3.2. The remaining task is to estimate the contibution from smaller $\mathfrak{p}$, namely

$$\sum_{\substack{u\leqslant N\mathfrak{p}<y\\ N\mathfrak{q}<z\\ \mathfrak{p}<\mathfrak{q}}}S(\mathcal{C}_{\mathfrak{pq}},I(\mathfrak{p})). \tag{3.14}$$

Now we begin the decoupling manœuver in earnest. Set

$$M := (\log X)^{1+B/2} \tag{3.15}$$

and choose spacing points

$$y=y_0>y_1>y_2>\cdots>y_M=u$$

where $y_m = y(u/y)^{m/M}$. Then splitting according to the size of $N\mathfrak{p}$ gives

$$\sum_{\substack{u\leqslant N\mathfrak{p}<y\\ N\mathfrak{q}<z\\ \mathfrak{p}<\mathfrak{q}}}S(\mathcal{C}_{\mathfrak{pq}},I(\mathfrak{p}))=\sum_{m=0}^{M-1}\sum_{\substack{y_{m+1}\leqslant N\mathfrak{p}<y_m\\ N\mathfrak{q}<z\\ \mathfrak{p}<\mathfrak{q}}}S(\mathcal{C}_{\mathfrak{pq}},I(\mathfrak{p}))$$

$$= \sum_{m=0}^{M-1} \sum_{\substack{y_{m+1} \leqslant N\mathfrak{p}, N\mathfrak{q} < y_m \\ \mathfrak{p}<\mathfrak{q}}} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})) + \sum_{m=0}^{M-1} \sum_{y_{m+1} \leqslant N\mathfrak{p} < y_m \leqslant N\mathfrak{q} < z} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})).$$

Applying (assuming here the conditions $y_{m+1} \leqslant N\mathfrak{p} < y_m \leqslant N\mathfrak{q} < z$, as in the right-hand sum) the Buchstab identity

$$S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})) = S(C_{\mathfrak{pq}}, I(y_{m+1})) - \sum_{\mathfrak{r} \in I(y_{m+1}) \setminus I(\mathfrak{p})} S(\mathcal{C}_{\mathfrak{pqr}}, I(\mathfrak{r})),$$

one confirms that

$$\sum_{\substack{u \leqslant N\mathfrak{p} < y \\ N\mathfrak{q} < z \\ \mathfrak{p}<\mathfrak{q}}} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})) = \sum_{m=0}^{M-1} \sum_{i=1}^{3} E_m^{(i)}, \tag{3.16}$$

where

$$E_m^{(1)} := \sum_{\substack{y_{m+1} \leqslant N\mathfrak{p} \\ N\mathfrak{q} < y_m \\ \mathfrak{p}<\mathfrak{q}}} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})), \qquad E_m^{(2)} := \sum_{\substack{y_{m+1} \leqslant N\mathfrak{p} < y_m \\ y_m \leqslant N\mathfrak{q} < z}} S(\mathcal{C}_{\mathfrak{pq}}, I(y_{m+1}))$$

and

$$E_m^{(3)} := - \sum_{y_{m+1} \leqslant N\mathfrak{r} < N\mathfrak{p} < y_m \leqslant N\mathfrak{q} < z} S(C_{\mathfrak{pqr}}, I(\mathfrak{r})).$$

What has been achieved here is that the first and third types of sum $E_m^{(1)}$ and $E_m^{(3)}$ correspond to atypical factorization patterns (with two prime factors very close in norm) and can ultimately be bounded fairly trivially, whereas in the term $E_m^{(2)}$ the conditions on $\mathfrak{p}, \mathfrak{q}$ are fully decoupled, which will allow us to relate these terms to Type II sums.

We turn now to the actual estimations. Looking first at $E_m^{(1)}$, note that

$$\begin{aligned} |E_m^{(1)}| &\leqslant \sum_{\substack{y_{m+1} \leqslant N\mathfrak{p}, N\mathfrak{q} < y_m \\ \mathfrak{p}<\mathfrak{q}}} \sum_{\mathfrak{pq}|\mathfrak{a}} |w(\mathfrak{a})| \ll X \sum_{\substack{y_{m+1} \leqslant N\mathfrak{p}, N\mathfrak{q} < y_m \\ \mathfrak{p}<\mathfrak{q}}} (N\mathfrak{pq})^{-1} \\ &\leqslant X \Bigg(\sum_{y_{m+1} \leqslant N\mathfrak{p} < y_m} (N\mathfrak{p})^{-1} \Bigg)^2 \ll X \log^2(y_m/y_{m+1}) \ll X(\log X)^2/M^2. \end{aligned} \tag{3.17}$$

The estimation of $E_m^{(3)}$ is almost identical. Since an ideal $\mathfrak{a}$ with $N\mathfrak{a} \leqslant X$ has $O(\log X)$ prime factors, each ideal $\mathfrak{a}$ with $\mathfrak{pr}|\mathfrak{a}$ is counted by $\ll \log X$ choices of $\mathfrak{q}$, and so we have

$$|E_m^{(3)}| \ll (\log X) \sum_{y_{m+1} \leqslant N\mathfrak{r} < N\mathfrak{p} < y_m} \sum_{\mathfrak{pr}|\mathfrak{a}} |w(\mathfrak{a})| \ll X(\log X)^3/M^2, \tag{3.18}$$

finishing here in the same way as the estimation of $E_m^{(1)}$.

Step 3: Deploying Type II information.

We turn now to the heart of the matter, which is the estimation of $E_m^{(2)}$ using Type II information. The key point is observing that $E_m^{(2)}$ *is* a Type II sum. Indeed we have

$$E_m^{(2)} = \sum_{\mathfrak{a},\mathfrak{b}} \alpha_{\mathfrak{a}} \beta_{\mathfrak{b}} w(\mathfrak{ab}) \tag{3.19}$$

with

$$\alpha_{\mathfrak{a}} = 1_{\mathfrak{a}\,\text{prime}} 1_{y_{m+1} \leqslant N\mathfrak{a} < y_m} \qquad \text{and} \qquad \beta_{\mathfrak{b}} = 1_{N\mathfrak{b} \leqslant X} \sum_{\substack{y_m \leqslant N\mathfrak{q} < z \\ \mathfrak{q}|\mathfrak{b}}} 1_{P^-(\mathfrak{b}) \geqslant y_{m+1}},$$

where $P^-(\mathfrak{b})$ denotes the norm of the smallest prime ideal factor of $\mathfrak{b}$. Clearly $|\alpha_{\mathfrak{a}}| \leqslant 1$, and $|\beta_{\mathfrak{b}}|$ is at most the number of prime ideal divisors of $\mathfrak{b}$, which is $\ll \log X$. Note also that $y_m/y_{m+1} = (y/u)^{1/M} < 2$, and so $\alpha_{\mathfrak{a}}$ is already supported on a dyadic range of $\mathfrak{a}$, and so is of the form (3.2) with $L = y_m$. Note (recalling the choice (3.12) of y) that $u \leqslant L \leqslant y = X^{3/8}$, and so for all values of m our assumption on the Type II properties of w gives, from (3.19), that

$$|E_m^{(2)}| \ll X(\log X)^{-B+1}. \tag{3.20}$$

We have now proven all of the necessary estimates; let us conclude the proof of Lemma 3.2 by putting them together.

From (3.16), (3.17), (3.18), (3.20) and the choice (3.15) of M and the fact that $B = 2A + 4$, we have

$$\sum_{\substack{u \leqslant N\mathfrak{p} < y \\ N\mathfrak{q} < z \\ \mathfrak{p} < \mathfrak{q}}} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})) \ll X(\log X)^3 M^{-1} + X(\log X)^{-B+1} M \ll X(\log X)^{-A}.$$

The quantity on the left here arose as (3.14). Together with (3.13), this gives the bound

$$\sum_{\substack{N\mathfrak{p} \geqslant u \\ N\mathfrak{q} < z \\ \mathfrak{p} < \mathfrak{q}}} S(\mathcal{C}_{\mathfrak{pq}}, I(\mathfrak{p})) \ll X(\log X)^{-A}.$$

The quantity on the left here is the third term in (3.4). Combining with the estimate (3.11) for the first two terms in (3.4), we obtain

$$S(\mathcal{C}, I(z)) \ll X(\log X)^{-A}.$$

Finally, from (3.5) and the trivial bound $|\sum_{\mathfrak{p}: N\mathfrak{p} < z} w(\mathfrak{p})| \ll z$, we obtain the bound claimed in Lemma 3.2. □

3.2. **Bounding the sum over semiprimes.** For Lemma 3.2 to be useful for sums over primes, we need a way to estimate the term $\sum_{\mathfrak{p}_1, \mathfrak{p}_2} w(\mathfrak{p}_1\mathfrak{p}_2)$ involving almost equally-sized semiprimes. We do this now under an assumption on w suitable for our intended purpose (of proving Theorem 1.1). We remark that in the work of Ford and Maynard [8, Theorem 4.16], they point to the fact that if one only achieves Type I information below $1/2$ and w is not bounded then one may not deduce any result regarding primes. However the weight function underlying [8, Theorem 4.16] places a substantial portion of weight on semiprimes; the role of the upper bound sieve is precisely to rule out this conspiracy.

At this point we specialise to the case of imaginary quadratic fields $K = \mathbf{Q}(\sqrt{-n})$ and to weight functions of product form (2.2).

Lemma 3.3. *Let $K = \mathbf{Q}(\sqrt{-n})$. Suppose that $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ is in product form (2.2) for some frequency $\ell \in \mathbf{Z}$. Suppose moreover that f, f' satisfy the pointwise bound*

$$|f(x)|, |f'(x)| \leqslant \Lambda_{\text{Cramér}}(x) + \Lambda'(x). \tag{3.21}$$

where $\Lambda_{\text{Cramér}}, \Lambda'$ are defined to be zero on non-integer arguments. Let η, $X^{-1/10} \leqslant \eta < \frac{1}{2}$, be a parameter. Then

$$\sum_{\substack{\mathfrak{p}_1, \mathfrak{p}_2 \\ N\mathfrak{p}_1\mathfrak{p}_2 \leqslant X \\ N\mathfrak{p}_i \geqslant \eta X^{1/2}}} \big|w(\mathfrak{p}_1\mathfrak{p}_2)\big| \ll \frac{X \log(1/\eta)}{(\log X)^2}.$$

Remark. We take the domain of definition of Λ' to be $\mathbf{Q}$, though it is supported on $\mathbf{Z}$. The key point of the lemma is that it represents a nontrivial saving over the trivial bound of $O(X/\log X)$.

Proof. We first dispense with the contribution from either $\mathfrak{p}_1$ or $\mathfrak{p}_2$ being generated by a rational prime. This is straightforward, since the number of such ideals with norm $\leqslant M$ is $\ll M^{1/2}$. Using the trivial bound $|w(\mathfrak{p}_1\mathfrak{p}_2)| \leqslant \log^2 X$, the total contribution from these terms is $\ll (X^{1/2}/\eta)^{3/2}\log^2 X$, which is acceptable by a large margin.

It therefore suffices to establish for each fixed $\mathfrak{p}$ not generated by a rational prime, and with $\eta X^{1/2} \leqslant N\mathfrak{p} \leqslant X^{1/2}$, that

$$\sum_{\mathfrak{q}: N\mathfrak{q}\leqslant X/N\mathfrak{p}} |w(\mathfrak{p}\mathfrak{q})| \ll \frac{X}{(N\mathfrak{p})\log X}, \tag{3.22}$$

where the sum is over prime ideals $\mathfrak{q}$ not generated by a rational prime. Indeed, this implies the result using

$$\sum_{\substack{\mathfrak{p}_1,\mathfrak{p}_2\\ N\mathfrak{p}_1\mathfrak{p}_2\leqslant X\\ N\mathfrak{p}_i\geqslant \eta X^{1/2}}} \big|w(\mathfrak{p}_1\mathfrak{p}_2)\big| \ll \sum_{\eta X^{1/2}\leqslant N\mathfrak{p}\leqslant X^{1/2}} \sum_{\mathfrak{q}: N\mathfrak{q}\leqslant X/N\mathfrak{p}} |w(\mathfrak{p}\mathfrak{q})|$$

and the fact that

$$\sum_{\eta X^{1/2}\leqslant N\mathfrak{p}\leqslant X^{1/2}} (N\mathfrak{p})^{-1} \ll \frac{\log(1/\eta)}{\log X},$$

where in these sums all ideals are prime but not generated by a rational prime.

We turn to the proof of (3.22). An important subtlety here is that, while $w = f\boxtimes_\ell f'$ is supported on principal ideals, the prime ideals $\mathfrak{p}$ and $\mathfrak{q}$ may not be principal. To deal with this, we use the representatives $\mathfrak{c}_g, \mathfrak{c}'_g$ for the ideal classes as described in Lemma 2.1. Suppose that $\mathfrak{p}\in[g]$ for some $g\in \mathrm{Cl}(K)$. Since w is supported on principal ideals, we may restrict attention to $\mathfrak{q}\in[g^{-1}]$. Henceforth, write $\mathfrak{c}=\mathfrak{c}_g$ and $\mathfrak{c}'=\mathfrak{c}'_{g^{-1}}$. Then by Lemma 2.1 we have $\mathfrak{p}=(a+b\sqrt{-n})\mathfrak{c}$ for some $a,b\in\mathbf{Z}$ and $\mathfrak{q}$ is always of the form $(x+y\sqrt{-n})\mathfrak{c}'$ for some $x,y\in\mathbf{Z}$. Also, $\mathfrak{c}\mathfrak{c}'=(\gamma)$ for some rational γ of height $O_n(1)$.

The generators of $\mathfrak{p}\mathfrak{q}$ are the numbers $u\gamma(a+b\sqrt{-n})(x+y\sqrt{-n})$, $u\in\mathcal{O}_K^*$. Writing the units u as $\zeta_j+\zeta'_j\sqrt{-n}$, $j=1,\dots,r$, where $\zeta_j,\zeta'_j\in\frac{1}{2n}\mathbf{Z}$, one may compute that these numbers are $a_{1,j}x+b_{1,j}y+(a_{2,j}x+b_{2,j}y)\sqrt{-n}$, where $a_{1,j}=\gamma(a\zeta_j-nb\zeta'_j)$, $b_{1,j}=-\gamma n(a\zeta'_j+b\zeta_j)$, $a_{2,j}=\gamma(a\zeta'_j+b\zeta_j)$, and $b_{2,j}=\gamma(a\zeta_j-nb\zeta'_j)$. Note here that we have $a_{i,j},b_{i,j}\in\frac{\gamma}{2n}\mathbf{Z}$ for all i,j. Thus from the assumption that w has the product form (2.2) and the bound (3.21) we have

$$|w(\mathfrak{p}\mathfrak{q})| \leqslant \sum_{j=1}^{r} (\Lambda_{\text{Cramér}}+\Lambda')(a_{1,j}x+b_{1,j}y)(\Lambda_{\text{Cramér}}+\Lambda')(a_{2,j}x+b_{2,j}y). \tag{3.23}$$

Here, $\Lambda_{\text{Cramér}}$ and Λ' take the value zero if the arguments are not integers. Observe also that

$$\frac{1}{n}(a_{1,j}^2n+b_{1,j}^2) = a_{2,j}^2n+b_{2,j}^2 = a_{1,j}b_{2,j}-a_{2,j}b_{1,j} = \gamma^2(a^2+nb^2)\neq 0. \tag{3.24}$$

Let $W:=X^{1/100}$. Then, recalling the definition (1.6), we have the pointwise bound $\Lambda'(x)\ll\Lambda_{\text{Cramér},W}(x)$ for $W<|x|\ll X^2$, and so from (3.23) we have

$$|w(\mathfrak{p}\mathfrak{q})| \ll \sum_{j=1}^{r} \max_{Q_1,Q_2\in\{Q,W\}} \Lambda_{\text{Cramér},Q_1}(a_{1,j}x+b_{1,j}y)\Lambda_{\text{Cramér},Q_2}(a_{2,j}x+b_{2,j}y). \tag{3.25}$$

unless $|a_{i,j}x+b_{i,j}y|\leqslant W$ for some i,j.

We turn to the task of proving (3.22). The constraint $N\mathfrak{q}\leqslant X/N\mathfrak{p}$ is contained in the constraint $|x|,|y|\ll(X/N\mathfrak{p})^{1/2}$ (note here that $\mathfrak{c}'$ is simply a fixed fractional ideal with norm $\asymp_n 1$). The contribution from $|a_{i,j}x+b_{i,j}y|\leqslant W$ will be negligible. We sum (3.25) over all other x,y with $|x|,|y|\ll(X/N\mathfrak{p})^{1/2}$, ignoring any congruence conditions necessary for $\mathfrak{q}$ to actually be an ideal in

$\mathcal{O}_K$, as we may by the positivity of all the terms involved. In order for $\mathfrak{q}$ to be a prime ideal, we must have $N\mathfrak{q} = (x^2+ny^2)N\mathfrak{c}'$ equal to a rational prime or the square of a rational prime.

Thus, writing $\gamma' := N\mathfrak{c}'$ and $a_1 = a_{1,j}$ etc, it is enough to prove that

$$\sum_{|x|,|y|\ll (X/N\mathfrak{p})^{1/2}} 1_{\text{prime}}((x^2+ny^2)\gamma')\Lambda_{\text{Cramér},Q_1}(a_1x+b_1y)\Lambda_{\text{Cramér},Q_2}(a_2x+b_2y) \ll \frac{X}{(N\mathfrak{p})\log X},$$

for any $Q_1, Q_2 \in \{Q, W\}$. (We omit the contribution from $(x^2+ny^2)\gamma'$ the square of a prime, which is easily shown to be considerably smaller.) Since $1_{\text{prime}}(t) \ll \frac{1}{\log X}\Lambda_{\text{Cramér},W}(t)$ for $|t| > W$, it is enough to prove that

$$\sum_{|x|,|y|\ll N} \Lambda_{\text{Cramér},W}((x^2+ny^2)\gamma')\Lambda_{\text{Cramér},Q_1}(a_1x+b_1y)\Lambda_{\text{Cramér},Q_2}(a_2x+b_2y) \ll N^2 \tag{3.26}$$

for any $Q_1, Q_2 \in \{Q, W\}$, where here we have omitted the negligible contribution from $|(x^2+ny^2)\gamma'| \leqslant W$, and now have written

$$N := (X/N\mathfrak{p})^{1/2}. \tag{3.27}$$

Note that since $\eta X^{1/2} \leqslant N\mathfrak{p} \leqslant X^{1/2}$ we have $X^{1/4} \ll N \ll X^{1/4}\eta^{-1/2}$. Recalling that $\eta^{-1} \ll X^{1/10}$, it follows that $N \leqslant X^{1/3}$, so certainly $W < N^{1/30}$.

The desired estimate (3.26) is a standard type of sieve bound, but it is not easy to find a suitable reference because (a) the sieving ranges W, Q_1, Q_2 can be different, so the underlying sieve does not have a well-defined dimension; (b) we need to track uniformity in the parameters a_i, b_i; and (c) there are two variables x, y. We have found an approach based on an application of the large sieve to be the most convenient. The result we need is Lemma C.3, which is a 2-dimensional variant of [28, Equation (7.38)]. Whilst generalising this kind of result from 1 to 2 or more dimensions presents no difficulty, we do not know of a reference for exactly what we need, and so we provide some details in Appendix C.

Returning to the proof of (3.26), suppose that $Q_1 \leqslant Q_2 \leqslant W$ (the other case $Q_2 \leqslant Q_1 \leqslant W$ is essentially identical). For rational primes p, we define sets Ω_p as follows. Denote by $\mathcal{P}_*$ the set of primes dividing $2n$, a^2+nb^2 or the numerator or denominator of γ or γ'. Note that $\mathcal{P}_*$ is a set of $O_n(1)$ primes, since $(a^2+nb^2)\gamma^2/\gamma' = N\mathfrak{p}$ and $\mathfrak{p}$ is a prime ideal.

We will be using Lemma C.3; let us now define the sets of residues Ω_p with which we will apply that result. If $p \in \mathcal{P}_*$, set

$$\Omega_p = \emptyset.$$

If $p \notin \mathcal{P}_*$, define Ω_p as follows. For $p \leqslant Q_1$, define

$$\Omega_p := \{(u,v) \in (\mathbf{Z}/p\mathbf{Z})^2 : (u^2+nv^2)(a_1u+b_1v)(a_2u+b_2v) \equiv 0 (\text{mod } p)\};$$

for $Q_1 < p \leqslant Q_2$ (which may be an empty range) define

$$\Omega_p := \{(u,v) \in (\mathbf{Z}/p\mathbf{Z})^2 : (u^2+nv^2)(a_2u+b_2v) \equiv 0 (\text{mod } p)\};$$

for $Q_2 < p \leqslant W$ (which may also be an empty range) define

$$\Omega_p := \{(u,v) \in \mathbf{Z}/p\mathbf{Z} : u^2+nv^2 \equiv 0 (\text{mod } p)\},$$

and define $\Omega_p = \emptyset$ for $p > W$. The purpose of making these definitions is that we have

$$\Lambda_{\text{Cramér},W}((x^2+ny^2)\gamma')\Lambda_{\text{Cramér},Q_1}(a_1x+b_1y)\Lambda_{\text{Cramér},Q_2}(a_2x+b_2y) \ll (\log Q_1)(\log Q_2)(\log W)\prod_p 1_{\Omega_p^c}(x,y) \tag{3.28}$$

(recall here that $a_1, a_2, b_1, b_2 \in \frac{\gamma}{2n}\mathbf{Z}$).

In order that we may apply Lemma C.3, we compute the densities $\alpha_p = |\Omega_p|/p^2$ in the different ranges. Here, note that if $p \notin \mathcal{P}_*$ then, from (3.24), we have

$$p \nmid (a_1b_2 - a_2b_1)(na_1^2 + b_1^2)(na_2^2 + b_2^2). \tag{3.29}$$

From this, we may compute that

$$\alpha_p = \frac{1}{p}\big(d(p) + (-n \mid p)\big) + O\bigg(\frac{1}{p^2}\bigg), \tag{3.30}$$

where $d(p) = 3$ if $p \leqslant Q_1$, $d(p) = 2$ if $Q_1 < p \leqslant Q_2$, and $d(p) = 1$ if $Q_2 < p \leqslant W$. (Here d is essentially the dimension of the sieve on the corresponding ranges.) In particular, for all p we have $\alpha_p \leqslant 4/p$.

Now we bound the sum $\sum_{q \leqslant N^{1/2}} \mu^2(q)h(q)$ appearing in Lemma C.3 from below. Recall that h is the multiplicative function satisfying $h(p) = \alpha_p/(1-\alpha_p)$. To do this we use essentially Rankin's trick, but it is instructive to phrase this in a probabilistic language. Define a random variable $Z := \prod_{p \leqslant W} Z_p$, where the Z_p are independent random variables with $Z_p = p$ with probability α_p, and $Z_p = 1$ otherwise. Thus for squarefree q we have

$$\mathbb{P}(Z = q) = h(q) \prod_{p \leqslant W} (1 - \alpha_p).$$

We give an upper bound for the quantity in (C.2). We have

$$\sum_{q \leqslant N^{1/2}} \mu^2(q)h(q) = \prod_{p \leqslant W} (1-\alpha_p)^{-1} \mathbb{P}\big(Z \leqslant N^{1/2}\big). \tag{3.31}$$

On the other hand,

$$\mathbb{E} \log Z = \sum_{p \leqslant W} \mathbb{E} \log Z_p = \sum_{p \leqslant W} \alpha_p \log p \leqslant 4 \sum_{p \leqslant W} \frac{\log p}{p} < \frac{\log N}{4}.$$

By Markov's inequality it follows that $\mathbb{P}(Z > N^{1/2}) < \frac{1}{2}$, and so from (3.31)

$$\sum_{q \leqslant N^{1/2}} \mu^2(q)h(q) \geqslant \frac{1}{2} \prod_{p \leqslant W} (1-\alpha_p)^{-1}. \tag{3.32}$$

Now from (3.30) we have (since the product of $1 + O(p^{-2})$ terms converges)

$$\prod_{p \leqslant W} (1 - \alpha_p) \asymp \prod_{p \leqslant W} \bigg(1 - \frac{d(p)}{p}\bigg) \prod_{p \leqslant W} \bigg(1 - \frac{(-n \mid p)}{p}\bigg). \tag{3.33}$$

However we have

$$\prod_{p \leqslant W} \bigg(1 - \frac{(-n \mid p)}{p}\bigg) \asymp 1$$

since $L(1, (-n \mid \cdot)) \neq 0$. (See, e.g, [42, Theorem 4.11] for a full discussion of the convergence of the infinite product here, which is only conditionally convergent.) Substituting into (3.33) and recalling the definition of the dimension function $d(\cdot)$ gives

$$\prod_{p \leqslant W} (1 - \alpha_p) \asymp \prod_{p \leqslant W} \bigg(1 - \frac{d(p)}{p}\bigg) \asymp \frac{1}{(\log Q_1)(\log Q_2)(\log W)}.$$

From this, (3.28) and (3.32) and Lemma C.3, we obtain the stated bound (3.26). This concludes the proof of Lemma 3.3. □

Combining Lemmas 3.2 and 3.3 and taking $A = 4$ immediately leads to the following, which is the only result from this section that we will need in what follows.

Proposition 3.4. *Fix $C \geqslant 28$. Let $K = \mathbf{Q}(\sqrt{-n})$, and suppose that $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ is in product form (2.2) with some frequency $\ell \in \mathbf{Z}$, and suppose moreover that f, f' in that definition satisfy the pointwise bound $|f(x)|, |f'(x)| \leqslant (\Lambda_{\mathrm{Cramér}} + \Lambda')(x/2n)$. Define $\tilde{w}(\mathfrak{a}) := w(\mathfrak{a})1_{N\mathfrak{a}\leqslant X}$, and suppose that $\tilde{w}$ satisfies Type I (with savings $(\log X)^{-12}$) at $X^{1/2}(\log X)^{-C}$ and Type II (with savings $(\log X)^{-12}$) between $[(\log X)^C, X^{3/8}]$. Then*

$$\sum_{\mathfrak{p}: N\mathfrak{p}\leqslant X} w(\mathfrak{p}) \ll_C \frac{X(\log\log X)^2}{(\log X)^2}.$$

4. Gowers norms and main proof framework

We are now in a position to give the overview of the proof of Theorem 1.1 in more detail, reducing it to the proof of three main results: Propositions 4.2, 4.3, and 4.5.

4.1. Reduction to polar coordinates.

We begin with an exercise in partial summation, reducing Theorem 1.1 to the following slightly more technical-looking statement, which removes the von Mangoldt weight on $x^2 + ny^2$ and replaces the rectangular cutoffs of Theorem 1.1 with a norm cutoff $x^2 + ny^2 \leqslant X$ and an angular frequency; this may be thought of as conversion to a kind of polar coordinates. This is essential in order to bring in the prime ideal theorem (Proposition 8.3).

Recall from Section 1.4 the definitions of Λ' and of $\chi_\infty^{(\ell)}$.

Proposition 4.1. *Suppose that $n \equiv 0$ or $4 \bmod 6$. Let $\ell \in \mathbf{Z}$, $|\ell| \leqslant (\log X)^{10}$ be a frequency. Then we have that*

$$\sum_{x,y\in\mathbf{Z}: x^2+ny^2\leqslant X} \chi_\infty^{(\ell)}(x + y\sqrt{-n})\Lambda'(x)\Lambda'(y)1_{x^2+ny^2 \text{ prime}} = \frac{\pi\kappa_n 1_{\ell=0} X}{\sqrt{n}\log X} + O\left(\frac{X(\log\log X)^2}{(\log X)^2}\right)$$

where κ_n is the constant in (1.1).

Remark. The exponent 10 is somewhat arbitrary, being chosen so that Proposition 4.1 suffices for the deduction of Theorem 1.1. If desired it could be replaced by any fixed constant A, at the expense of requiring the $O()$ term to depend on A.

Proof of Theorem 1.1, assuming Proposition 4.1. Let $W \in C_0^\infty(\mathbf{R}^2)$ be as in the statement of Theorem 1.1. Rescaling N by a factor of $O_W(1)$ if necessary, we may assume that W is supported on $B_{10}(0)$. It is convenient to write $\|W\|$ for the maximum L^∞-norm of any partial derivative of W of order at most 3. Set $\eta := (\log N)^{-10}$ (say), and suppose that $\eta \leqslant R \leqslant 10$. Setting $X = (RN)^2$ in the conclusion of Proposition 4.1 gives

$$(2\log N)\sum_{x,y\in\mathbf{Z}} \Lambda'(x)\Lambda'(y)1_{x^2+ny^2 \text{ prime}}\psi_{R,\ell}\left(\frac{x}{N}, \frac{y}{N}\right) = \kappa_n N^2\left(\int_{\mathbf{R}^2}\psi_{R,\ell}\right) + O\left(\frac{N^2(\log\log N)^2}{\log N}\right), \tag{4.1}$$

where

$$\psi_{R,\ell}(r\cos 2\pi\theta, n^{1/2}r\sin 2\pi\theta) := 1_{r\leqslant R}\cdot e(\ell\theta),$$

and this is valid for $|\ell| \leqslant (\log N)^{10}$. Note that here we have replaced the factor $\log RN$ which arises in Proposition 4.1 by $\log N$; the error in doing this can be absorbed into the $O(\cdot)$ term on the RHS of (4.1).

For $x^2 + ny^2 \geqslant \eta N^2$, we have

$$W\left(\frac{x}{N}, \frac{y}{N}\right) = -\int_\eta^{10}\sum_{\ell\in\mathbf{Z}} c(R,\ell)\psi_{R,\ell}\left(\frac{x}{N}, \frac{y}{N}\right)dR, \tag{4.2}$$

where the $c(\cdot,\cdot)$ are defined in terms of the Fourier expansion

$$\frac{\partial}{\partial R}W(R\cos 2\pi\theta, n^{1/2}R\sin 2\pi\theta) = \sum_{\ell\in\mathbf{Z}} c(R,\ell)e(\ell\theta).$$

(To verify this, substitute $\frac{x}{N} = r\cos 2\pi\theta$, $\frac{y}{N} = n^{1/2}r\sin 2\pi\theta$.) We claim that

$$\int_\eta^{10} \sum_{|\ell|>(\log N)^{10}} |c(R,\ell)|dR \ll (\log N)^{-10}\|W\| \tag{4.3}$$

and that

$$\int_\eta^{10} \sum_{\ell\in\mathbf{Z}} |c(R,\ell)|dR \ll \|W\|. \tag{4.4}$$

To prove these statements, observe that

$$|c(R,\ell)| \ll |\ell|^{-2}\int_0^1 \Big|\frac{\partial^2}{\partial\theta^2}\frac{\partial}{\partial R}W(R\cos 2\pi\theta, n^{1/2}R\sin 2\pi\theta)\Big|d\theta \ll |\ell|^{-2}\|W\| \tag{4.5}$$

for $\ell\neq 0$. Here, the first bound follows by integrating the definition of Fourier transform twice by parts, and the second from the chain rule. We also have

$$|c(R,0)| \ll \|W\|. \tag{4.6}$$

Using (4.5) and (4.6) immediately leads to (4.3) and (4.4).

From (4.2) and (4.3) it follows that

$$W\Big(\frac{x}{N},\frac{y}{N}\Big) = -\int_\eta^{10} \sum_{|\ell|\leqslant(\log N)^{10}} c(R,\ell)\psi_{R,\ell}\Big(\frac{x}{N},\frac{y}{N}\Big)dR + O((\log N)^{-10}\|W\|). \tag{4.7}$$

when $x^2+ny^2\geqslant\eta N^2$. Now consider

$$S := (2\log N)\sum_{x,y\in\mathbf{Z}} \Lambda'(x)\Lambda'(y)1_{x^2+ny^2\text{ prime}}W\Big(\frac{x}{N},\frac{y}{N}\Big). \tag{4.8}$$

The contribution to S from $x^2+ny^2<\eta N^2$ is $\ll N^2(\log N)^{-9}\|W\|_\infty$. By (4.7), the sum over the remaining range is

$$-\int_\eta^{10} dR \sum_{|\ell|\leqslant(\log N)^{10}} c(R,\ell) \sum_{\substack{x,y\in\mathbf{Z}\\ x^2+ny^2\geqslant\eta N^2}} \Lambda'(x)\Lambda'(y)1_{x^2+ny^2\text{ prime}}\psi_{R,\ell}\Big(\frac{x}{N},\frac{y}{N}\Big) + O(N^2(\log N)^{-9}\|W\|).$$

We now remove the condition $x^2+ny^2\geqslant\eta N^2$ from the sums; the error in doing so is (crudely)

$$\ll \eta N^2(\log N)^2\int_\eta^{10} dR\sum_{\ell\in\mathbf{Z}}|c(R,\ell)| \ll N^2(\log N)^{-8}\|W\|,$$

by (4.4). Therefore

$$S = -(2\log N)\int_\eta^{10} dR \sum_{|\ell|\leqslant(\log N)^{10}} c(R,\ell)\sum_{x,y\in\mathbf{Z}} \Lambda'(x)\Lambda'(y)1_{x^2+ny^2\text{ prime}}\psi_{R,\ell}\Big(\frac{x}{N},\frac{y}{N}\Big) + O(N^2(\log N)^{-8}\|W\|).$$

Applying (4.1) and then (4.4) gives

$$S = -\int_\eta^{10} dR \sum_{|\ell| \leqslant (\log N)^{10}} c(R,\ell)\Big(\kappa_n N^2 \int \psi_{R,\ell} + O\Big(\frac{N^2(\log\log N)^2}{\log N}\Big)\Big) + O(N^2(\log N)^{-8}\|W\|)$$
$$= -\int_\eta^{10} dR \sum_{|\ell| \leqslant (\log N)^{10}} c(R,\ell)\Big(\kappa_n N^2 \int_{\mathbf{R}^2} \psi_{R,\ell}\Big) + O\Big(\frac{N^2(\log\log N)^2}{\log N}\|W\|\Big).$$

Finally, comparing this with the integral of (4.7) we see that

$$S = \kappa_n\Big(\int_{\mathbf{R}^2} W\Big) + O\Big(\frac{N^2(\log\log N)^2}{\log N}\|W\|\Big). \tag{4.9}$$

Recall that S is given by (4.8). To deduce Theorem 1.1, we add back in the contribution of prime powers to the two Λ' terms and replace $(2\log N)1_{x^2+ny^2 \text{ prime}}$ with $\Lambda(x^2+ny^2)$. The former replacement is trivial. The error in making the latter replacement is bounded by

$$\ll (\log N)^2 \sum_{\substack{x^2+ny^2 \leqslant N^2 \\ x,y,x^2+ny^2 \text{ prime}}} W\Big(\frac{x}{N},\frac{y}{N}\Big)\log\Big(\frac{x^2+ny^2}{N^2}\Big) = -(\log N)^2 \int_1^{N^2} \frac{dt}{t} \sum_{\substack{x^2+ny^2 \leqslant t \\ x,y,x^2+ny^2 \text{ prime}}} W\Big(\frac{x}{N},\frac{y}{N}\Big).$$

The inner sum on the right is $\ll t(\log N)^{-3}\|W\|$ by a standard upper bound sieve. (For instance, one could use Lemma C.3 with $k=2$ and analyse as in (3.32), noting that in this case one has $\alpha_p = \frac{4}{p} + O(\frac{1}{p^2})$ for p with $(-n \mid p) = 1$ and $\alpha_p = \frac{2}{p} + O(\frac{1}{p^2})$ for all p with $(-n \mid p) = -1$.) Thus we see that the error in passing from (4.9) to Theorem 1.1 is $O(N^2\|W\|/\log N)$, which may be absorbed into the error term in Theorem 1.1. □

Remark. One of the main ingredients of our argument is the prime ideal theorem Proposition 8.3. For ease of reference we have stated that result with a sharp cutoff $N\mathfrak{a} \leqslant X$. If one used a smooth cutoff instead then one could probably improve the dependence on the derivatives of W in the above analysis.

4.2. **Gowers norms.** The Gowers norms, introduced by Gowers [14], are a fundamental notion in additive combinatorics. A discussion of their basic properties may be found in many places; see for instance the introductions to [35, 47]. We briefly recall the definitions. Let $f : \mathbf{Z} \to \mathbf{C}$ be a function with finite support. We first define, for $h \in \mathbf{Q}$, the difference operators

$$\Delta_h f(x) = f(x)\overline{f(x+h)}. \tag{4.10}$$

(It is convenient for technical reasons later on to allow $h \in \mathbf{Q}$, but of course $\Delta_h f$ will be zero unless $h \in \mathbf{Z}$). Now let $k \geqslant 2$ be a parameter. Then we first define

$$\|f\|_{U^k(\mathbf{Z})}^{2^k} := \sum_{x,h_1,\dots,h_k \in \mathbf{Z}} \Delta_{h_1}\cdots\Delta_{h_k} f(x). \tag{4.11}$$

The Gowers $U^k(\mathbf{Z})$-norm $\|f\|_{U^k(\mathbf{Z})}$ is then defined to be the unique non-negative 2^k-th root of this quantity. Given a real number $N \geqslant 1$, we then define

$$\|f\|_{U^k[N]} := \|f\|_{U^k(\mathbf{Z})}/\|1_{[N]}\|_{U^k(\mathbf{Z})}. \tag{4.12}$$

This definition will be relevant when f is supported on $[\pm O(N)]$. It is easy to see that this is completely equivalent to the definition given in, for instance, [35, Definition 1.1], which is specifically geared to functions f which are supported on $[N]$. The normalising factor $\|1_{[N]}\|_{U^k(\mathbf{Z})}$ has the effect

that the maximum value of $\|f\|_{U^k[N]}$ is exactly 1 if f is supported on $[N]$. For us, precise constants will never be relevant and it will suffice to know that

$$\|1_{[N]}\|_{U^k(\mathbf{Z})}^{2^k} \asymp_k N^{k+1}, \tag{4.13}$$

which is clear from the definition. That the Gowers norms are well-defined and are norms are well-known; proofs of this statement may be found in many places, for instance [17, Appendix B.5]. We have the nesting property

$$\|f\|_{U^2[N]} \ll \|f\|_{U^3[N]} \ll \cdots \ll \|f\|_{U^k[N]} \ll \cdots \tag{4.14}$$

Further properties of the Gowers norms relevant to our paper may be found in Appendix A. As a final remark, we note that all Gowers norms in the main part of our paper are at scale $X^{1/2}$, where X is our main global parameter.

4.3. **Type I and II sums with product weights and Gowers norms.** In this section we state our two main results linking Gowers norms and Type I/II estimates for weight functions in product form. We begin with the Type I statement.

Proposition 4.2. *There is an absolute constant $C_{4.2}$ with the following property. Let $\delta \in (0,1)$ be a parameter. Let $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ be a weight function in product form (2.2) with frequency ℓ. Suppose that f, f' are 1-bounded and supported on $[\pm 2nX^{1/2}]$. Suppose $L \leqslant \delta^{C_{4.2}} X^{1/2}$ and that*

$$\sum_{\mathfrak{d}: N\mathfrak{d} \sim L} \Big| \sum_{\substack{\mathfrak{d} | \mathfrak{a} \\ N\mathfrak{a} \leqslant X}} w(\mathfrak{a}) \Big| \geqslant \delta X. \tag{4.15}$$

Then

$$\|f\|_{U^3[X^{1/2}]}, \|f'\|_{U^3[X^{1/2}]} \gg \Big(\frac{\delta}{(|\ell|+1)\log X} \Big)^{O(1)}.$$

Remark. Jori Merikoski has indicated to us a rather different argument using the multiplicative large sieve, which allows one to show that in fact f, f' correlate with linear phases $e(\theta x)$ with θ close to a rational with denominator $\ll (\log X/\delta)^{O(1)}$. We have opted to retain our original argument so as to give a unified treatment of the Type I and II sums. With some additional argument it may also be used to give the aforementioned stronger conclusion, as we will discuss in [16].

Now we give the Type II statement.

Proposition 4.3. *There exists an absolute constant $C_{4.3}$ and a (large) positive integer k such that the following holds. Let $w : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ be a weight function in product form (2.2) with frequency ℓ, and suppose that f, f' are 1-bounded and supported on $[\pm 2nX^{1/2}]$. Suppose that $\delta^{-C_{4.3}} \ll L \leqslant X^{3/8}$, and suppose that for some 1-bounded weights $\alpha, \beta : \mathrm{Ideals}(\mathcal{O}_K) \to \mathbf{C}$ we have*

$$\Big| \sum_{\substack{\mathfrak{a},\mathfrak{b}: N\mathfrak{a} \sim L \\ N\mathfrak{a}\mathfrak{b} \leqslant X}} \alpha_{\mathfrak{a}} \beta_{\mathfrak{b}} w(\mathfrak{a}\mathfrak{b}) \Big| \geqslant \delta X. \tag{4.16}$$

Then we have

$$\|f\|_{U^k[X^{1/2}]}, \|f'\|_{U^k[X^{1/2}]} \geqslant \Big(\frac{\delta}{(|\ell|+1)\log X} \Big)^{O(1)}.$$

Remarks. A trivial modification to the proof gives a similar statement in the range $L \leqslant X^{1/2-\kappa}$, with $k = O_\kappa(1)$. The value of k is extremely large; for further comments see Appendix B.

4.4. **Approximations to Λ' in Gowers norm.** Define $\Lambda_{\text{Cramér}}$ as in (1.6). Extend the domain of definition to $\mathbf{Q}$, with the convention that $\Lambda_{\text{Cramér}}(t) = 0$ if $t \notin \mathbf{Z}$. A crucial ingredient in our arguments is the fact that these functions are good approximants to Λ' in the Gowers norms.

Proposition 4.4. *Fix $A \geqslant 1$ and $k \geqslant 2$. Then we have that*

$$\|(\Lambda' - \Lambda_{\text{Cramér}})1_{[X^{1/2}]}\|_{U^k[X^{1/2}]}^{2^k} \ll_A (\log X)^{-A}.$$

Proof. We will apply the results of [47] and [34, Section 5], taking $N = X^{1/2}$ in those papers. Note (cf. the comments following (1.7)) that our value of Q is precisely the same as the one in those papers. Let Λ_{Siegel} be the Siegel model for the primes, introduced in [47, Definition 2.1]; the definition in [34, Section 5] is the same. Let q_{Siegel} be the modulus of the character for which there is a Q-Siegel zero β, if one exists (again, see [47, Definition 2.1] for the definition).

Then by [34, Theorem 7] (and ignoring the contribution of the prime powers to Λ, which is negligible) we have

$$\|(\Lambda' - \Lambda_{\text{Siegel}})1_{[X^{1/2}]}\|_{U^k[X^{1/2}]} \ll e^{-(\log X)^{c_k}}. \tag{4.17}$$

On the other hand, by [47, Theorem 2.5] we have

$$\|(\Lambda_{\text{Siegel}} - \Lambda_{\text{Cramér},Q})1_{[X^{1/2}]}\|_{U^k[X^{1/2}]} \ll q_{\text{Siegel}}^{-c_k}, \tag{4.18}$$

Note here that the exponent c in [47, Theorem 2.5] does depend on k, as per the convention laid out on the first page of that paper. If there is no Siegel zero, the LHS of (4.18) is zero. By Siegel's theorem we have the (ineffective) bound

$$q_{\text{Siegel}} \gg_A (\log X)^A. \tag{4.19}$$

(In more detail, [28, Theorem 5.28 (2)] with $\varepsilon = 1/10A$ implies that $1 - \beta \gg_A q_{\text{Siegel}}^{-1/10A}$; on the other hand, by the definition [47, Definition 2.1] we have $1 - \beta \ll \frac{1}{\log Q} \sim (\log X)^{-1/10}$.)

The result follows from this and (4.17) and (4.18) by the triangle inequality for the Gowers norms. □

Remark. This is the point in the paper where the quasi-polynomial bounds for the inverse theorem for the Gowers norms [35] has been applied, this being the major ingredient in the proof of (4.17).

4.5. **Evaluating the main term.** The following will be the main result of Section 8.

Proposition 4.5. *Let $\ell \in \mathbf{Z}$ and suppose that $|\ell| \leqslant e^{\sqrt{\log X}}$. Then we have*

$$(\log X) \sum_{x,y \in \mathbf{Z}: x^2+ny^2 \leqslant X} \chi_\infty^{(\ell)}(x + y\sqrt{-n})\Lambda_{\text{Cramér}}(x)\Lambda_{\text{Cramér}}(y)1_{x^2+ny^2 \text{ prime}}$$
$$= \frac{\pi\kappa_n 1_{\ell=0}}{\sqrt{n}} X + O(X(\log X)^{-1}),$$

where κ_n is the constant in (1.1) and $\chi_\infty^{(\ell)}(z) = (z/|z|)^\ell$.

4.6. **Proof of the main theorem.** Recall our introductory discussion in Section 1.2, and in particular the assertion (1.3) that our main theorem in the case $n = 4$ could be considered in terms of a sum over Gaussian primes. We begin by formulating such a statement for general n.

Lemma 4.6. *Let $\ell \in \mathbf{Z}$, and let $K = \mathbf{Q}(\sqrt{-n})$. Then we have*

$$\sum_{x^2+ny^2 \leqslant X} \chi_\infty^{(\ell)}(x + y\sqrt{-n})\Lambda'(x)\Lambda'(y)1_{x^2+ny^2 \text{ prime}} = \sum_{\mathfrak{p}: N\mathfrak{p} \leqslant X} w(\mathfrak{p}) + O(X^{1/2+o(1)}), \tag{4.20}$$

where $\mathfrak{p}$ runs over prime ideals in $\mathcal{O}_K$ and

$$w = \Lambda' \boxtimes_\ell \Lambda'. \tag{4.21}$$

Proof. Recalling the definition of $\boxtimes_\ell$ (see (2.2)) we see that the sum on the right of (4.20) is the sum of $\chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda'(x)\Lambda'(y)$ over $(x,y)\in\mathbf{Z}^2\setminus\{0\}$ for which $(x+y\sqrt{-n})$ is a prime ideal.

Now the prime ideals $\mathfrak{p}$ in $\mathcal{O}_K$ are of two types: those which sit above a rational prime p which splits (or is ramified), in which case $N\mathfrak{p}=p$, or those which are equal to (p) for some rational prime p which does not split, in which case $N\mathfrak{p}=p^2$. The contribution of the non-split terms to the LHS of (4.20) is $\ll(\log X)^2\sum_{p\leqslant X^{1/2}}\sum_{x,y}1_{x^2+ny^2=p^2}\ll X^{1/2+o(1)}$, where the last bound follows from Lemma D.2 and the divisor bound. The contribution of the non-split terms to the RHS of (4.20) is at most $(\log X)^2$ times the number of ideals with norm $p^2\leqslant X$ for some p, which is again $\ll X^{1/2+o(1)}$. □

We may now prove the main theorem itself, assuming Propositions 4.2, 4.3, and 4.5.

Proof of Proposition 4.1. We start with the LHS of the expression in Proposition 4.1, which we expressed as a sum over $w(\mathfrak{p})$ in (4.20). Note that $w(\mathfrak{a})$ as given in (4.21) is in product form, which brings to mind Proposition 3.4. However, it would not be a good idea to try and apply Proposition 3.4 immediately, since $\tilde{w}(\mathfrak{a})=w(\mathfrak{a})1_{N\mathfrak{a}\leqslant X}$ does not satisfy Type I estimates (it is irregularly distributed mod 3, for instance). To get around this latter issue, we introduce a decomposition using the approximant $\Lambda_{\mathrm{Cramér}}$, defined in (1.6). Thus we write

$$w=\sum_{j=1}^{3}w_j,\qquad w_j=f_j\boxtimes_\ell f_j' \tag{4.22}$$

where

$$f_1=\Lambda',\quad f_1'=f_2=\Lambda'-\Lambda_{\mathrm{Cramér}},\quad\text{and}\quad f_2'=f_3=f_3'=\Lambda_{\mathrm{Cramér}}. \tag{4.23}$$

A trivial change to the proof of Lemma 4.6 (replacing Λ' by $\Lambda_{\mathrm{Cramér}}$ throughout) shows that

$$\sum_{\mathfrak{p}:N\mathfrak{p}\leqslant X}w_3(\mathfrak{p})=\sum_{x^2+ny^2\leqslant X}\chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda_{\mathrm{Cramér}}(x)\Lambda_{\mathrm{Cramér}}(y)1_{x^2+ny^2\text{ prime}}+O(X^{1/2+O(1)}).$$

In view of Proposition 4.5, it therefore suffices to show that

$$\sum_{\mathfrak{p}:N\mathfrak{p}\leqslant X}w_1(\mathfrak{p}),\sum_{\mathfrak{p}:N\mathfrak{p}\leqslant X}w_2(\mathfrak{p})\ll X\frac{(\log\log X)^2}{(\log X)^2}. \tag{4.24}$$

For this, we use Proposition 3.4. Note that w_1,w_2 do satisfy the boundedness condition (3.21) of that result. Since the concern in that result is with $\tilde{w}_j(\mathfrak{a})=w_j(\mathfrak{a})1_{N\mathfrak{a}\leqslant X}$, we lose nothing by assuming that f_j,f_j' are supported on $[\pm 2nX^{1/2}]$.

The Type I/II information that we need for the $\tilde{w}_j$ comes from (the contrapositive of) Propositions 4.2 and 4.3 respectively, together with Proposition 4.4, noting that for $j=1,2$, at least one of the functions f_j,f_j' is $\Lambda'-\Lambda_{\mathrm{Cramér},Q}$. In applying Propositions 4.2 and 4.3, one should take $\delta=(\log X)^{-12}$; the Gowers norm conclusions of those propositions are then contradicted by Proposition 4.4 for sufficiently large A. (In particular, we have that we may apply Proposition 3.4 with $C=12\max(C_{4.2},C_{4.3})+1$.) Note that, although Propositions 4.2 and 4.3 were stated only for 1-bounded functions, they apply equally well to functions bounded by $\log X$, simply by applying the results with $(\log X)^{-1}f$ and $(\log X)^{-1}f'$ and absorbing the extra factors of $\log X$ in the $(\log X)^{O(1)}$ part of those bounds.

This concludes the proof of Proposition 4.1, assuming Propositions 4.2, 4.3, and 4.5. □

5. Preliminaries on concatenation and Gowers–Peluse norms

In this section we develop some preliminary material connected with concatenation theorems for Gowers norms. When dealing with concatenation, it is useful to introduce generalisations of Gowers norms first considered by Peluse [43, Definition 2.1]. Peluse calls these norms the *Gowers box norms*, but we prefer to see the term 'box norm' reserved for purely combinatorial graph- and hypergraph-norms and so will instead refer to these norms as Gowers–Peluse norms.

We first state the definition of the Gowers–Peluse norm.

Definition 5.1. Given a function $f : \mathbf{Z} \to \mathbf{C}$ with finite support and $h, h' \in \mathbf{Q}$, we define

$$\Delta_{(h,h')}f(x) = f(x+h)\overline{f(x+h')}. \tag{5.1}$$

Given a multiset $\Omega = \{\mu_1, \dots, \mu_k\}$ of probability measures on $\mathbf{Q}$ and a positive integer scale $N \geqslant 1$, we define

$$\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\Omega]} = \|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]} := \frac{1}{N}\sum_{x\in\mathbf{Z}} \mathbb{E}_{h_i,h_i'\sim\mu_i}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_k,h_k')}f(x). \tag{5.2}$$

(As shown in Lemma A.1, the RHS is a non-negative real number, and so $\|f\|_{U_{\mathrm{GP}}[N;\Omega]}$ is uniquely defined as a nonnegative real number.)

Remark. The only difference between our definition and that of Peluse is the fact that we work with general probability measures rather than uniform measures on sets as Peluse did. (The two settings are equivalent, at least if one passes from sets to multisets as in [32], with the multiset situation being a special case of the measure one, and the measure one a limiting case of the multiset one.)

The definition of the difference operators $\Delta_{(h,h')}$ should be compared with (4.10), of which it is a mild variant. Once again, for minor technical reasons it is convenient to allow $h, h' \in \mathbf{Q}$, but $\Delta_{(h,h')}f$ will be identically zero unless $h' - h \in \mathbf{Z}$. Note that $\Delta_{(h,h')}$, $\Delta_{(k,k')}$ commute. Later on it will be convenient to adopt the shorthand $\Delta_{M(h,h')}f := \Delta_{(Mh,Mh')}f$.

If f is supported on $[\pm N]$ and if the μ_i are all uniform on $[\pm N]$ then $\|f\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}$ is a kind of 'smoothed' Gowers norm which is comparable to, but not exactly the same as, $\|f\|_{U^k[N]}$. We will not actually need this fact in the paper: for equivalent results see [45, Lemma C.3] or [31, Lemma A.3].

We next record a certain general–purpose lemma and corollary which will be used in order to convert Gowers-Peluse norms whose underlying measures are somewhat close to uniform on $[\pm N]$ into genuine Gowers norms, albeit of a higher order.

Lemma 5.2. *Let $\delta \in (0,1)$, $N, T \geqslant 1$ and consider a 1-bounded function $f : \mathbf{Z} \to \mathbf{C}$, supported on $[\pm N]$. Furthermore let μ be a probability measure supported on $[\pm N]$ which is somewhat uniform in the sense that $\|\mu\|_2^2 \leqslant T/N$. Suppose that $\|f\|^2_{U_{\mathrm{GP}}[N;\mu]} \geqslant \delta$. Then $\|f\|_{U^2[N]} \gg (\delta^2/T)^{1/4}$.*

Proof. The condition $\|f\|^2_{U_{\mathrm{GP}}[N;\mu]} \geqslant \delta$ is, written out,

$$\sum_{x\in\mathbf{Z}} \mathbb{E}_{h_1,h_1'\sim\mu} f(x+h_1)\overline{f(x+h_1')} = \sum_{x\in\mathbf{Z}} \mathbb{E}_{h_1,h_1'\sim\mu} f(x)\overline{f(x+h_1-h_1')} \geqslant \delta N.$$

This implies that

$$\sum_{t\in\mathbf{Z}} \mathbb{E}_{h_1,h_1'\sim\mu} 1_{h_1-h_1'=t}\Big|\sum_{x\in\mathbf{Z}} f(x)\overline{f(x+t)}\Big| \geqslant \delta N.$$

Removing the averaging over h_1', we see that there is some h_1' such that

$$\sum_{t\in\mathbf{Z}} \mathbb{E}_{h_1\sim\mu} 1_{h_1-h_1'=t}\Big|\sum_{x\in\mathbf{Z}} f(x)\overline{f(x+t)}\Big| \geqslant \delta N,$$

or in other words

$$\sum_{t\in\mathbf{Z}}\mu(h_1'+t)\Big|\sum_{x\in\mathbf{Z}}f(x)\overline{f(x+t)}\Big|\geqslant\delta N.$$

Cauchy-Schwarz and the assumption of the lemma give

$$\sum_{t\in\mathbf{Z}}\Big|\sum_{x\in\mathbf{Z}}f(x)\overline{f(x+t)}\Big|^2\geqslant\frac{(\delta N)^2}{T/N}=\delta^2N^3/T.$$

Expanding the left-hand side gives

$$\sum_{x,x',t\in\mathbf{Z}}f(x)\overline{f(x+t)f(x')}f(x'+t)=\sum_{x,t,t'\in\mathbf{Z}}\Delta_t\Delta_{t'}f(x)\geqslant\delta^2N^3/T.$$

However, the left-hand side is $\gg N^3$ times $\|f\|_{U^2[N]}^4$. □

The following corollary details how the above may be applied iteratively in order to replace a Gowers–Peluse norm with nearly uniform measures by a geunine Gowers norm (of twice the order).

Corollary 5.3. *Let $\delta\in(0,1)$, let $N,T\geqslant1$ be parameters with N an integer, and let $f:\mathbf{Z}\to\mathbf{C}$ be a 1-bounded function supposed on $[\pm N]$. Let $\mu_1,\dots,\mu_k$ be probability measures supported on $[\pm N]$, all of which are somewhat uniform in the sense that $\|\mu_i\|_2^2\leqslant T/N$. Suppose that $\|f\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}\geqslant\delta$. Then $\|f\|_{U^{2k}[N]}\gg(T^{-k}\delta^{2^k})^{1/2^{2k}}$.*

Proof. Let μ be the uniform measure on $[\pm N]$. We will prove by downward induction on $j=k,k-1,\dots,0$ that, for any j,

$$\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant j}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu\\j+1\leqslant i\leqslant k}}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_j,h_j')}\Delta_{h_{j+1}}\Delta_{h_{j+1}'}\cdots\Delta_{h_k}\Delta_{h_k'}f(x)\geqslant T^{j-k}\delta^{2^k}.\tag{5.3}$$

The assumption is the case $j=k$. First note that this may be rewritten as

$$\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant j-1}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu\\j+1\leqslant i\leqslant k}}\|\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{j-1},h_{j-1}')}\Delta_{h_{j+1}}\Delta_{h_{j+1}'}\cdots\Delta_{h_k}\Delta_{h_k'}f\|_{U_{\mathrm{GP}}[N;\mu_j]}^2\geqslant T^{j-k}\delta^{2^k}.$$

Write $\delta(h)$ (where $h=(h_1,h_1',\dots,h_{j-1},h_{j-1}',h_{j+1},h_{j+1}',\dots,h_k,h_k')$) for the size of the inner $U_{\mathrm{GP}}[N;\mu_j]$-norm squared. Then, applying Lemma 5.2, we obtain using Cauchy-Schwarz

$$\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant j-1}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu\\j+1\leqslant i\leqslant k}}\|\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{j-1},h_{j-1}')}\Delta_{h_{j+1}}\Delta_{h_{j+1}'}\cdots\Delta_{h_k}\Delta_{h_k'}f\|_{U_2[N]}^4$$
$$\geqslant T^{-1}\mathbb{E}\delta(h)^2\geqslant T^{j-1-k}\delta^{2^k}.$$

Writing in the definition of the U^2-norm using dummy variables h_j,h_j', this is

$$\frac{1}{N}\sum_{x\in\mathbf{Z}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant j-1}}\mathbb{E}_{\substack{h_i,h_i'\sim\mu\\j\leqslant i\leqslant k}}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{j-1},h_{j-1}')}\Delta_{h_j}\Delta_{h_j'}\cdots\Delta_{h_k}\Delta_{h_k'}f(x)\geqslant T^{-1}\mathbb{E}\delta(h)^2\geqslant T^{j-1-k}\delta^{2^k}.$$

This in turn is exactly the statement (5.3) in the case $j-1$, so the induction goes through.

The case $j=0$ of (5.3) is the statement we want. □

Another key corollary of Lemma 5.2 is that if a and b are coprime then an average over $ax+by$ being large can be converted to U^2-control.

Lemma 5.4. *Let $\eta\in(0,1)$ be sufficiently small and let $M=O(1)$ be a positive integer. Let $a,b\in\frac{1}{M}\mathbf{Z}\setminus\{0\}$ satisfy $|a|\geqslant\eta|b|$, $|b|\geqslant\eta|a|$ and $d:=\gcd(Ma,Mb)\leqslant1/\eta$. Set $Q:=\max(|a|,|b|)$,*

and suppose that $N \geqslant Q^2/\eta^3$. Let $I_1, I_2 \subseteq [\pm N/\eta Q]$ be subintervals of $\mathbf{Z}$. Suppose that $f : \mathbf{Q} \to \mathbf{C}$ is 1-bounded and supported on $\mathbf{Z} \cap [\pm N]$ and that

$$\Big| \sum_{\substack{x,y \in \mathbf{Z} \\ x \in I_1, y \in I_2}} f(ax+by) \Big| \geqslant \eta N^2/Q^2.$$

Then $\|f\|_{U^2[N]}^4 \gg \eta^{O(1)}$.

Proof. We generally do not indicate dependence on M in implied constants. Write $I_1 := \ell_1 + [L_1]$ and $I_2 := \ell_2 + [L_2]$, where $L_i := |I_i|$ (and $[L] = \{1, \dots, L\}$). Set $t = a\ell_1 + b\ell_2$ and write $a' = Ma/d$ and $b' = Mb/d$. Then the assumption may be written as

$$\Big| \sum_{\ell \in \frac{d}{M}(a'[L_1] + b'[L_2])} f(t+\ell) \Big| \geqslant \eta N^2/Q^2. \tag{5.4}$$

(Note here that the sum is over the *multi*set $\frac{d}{M}(a'[L_1] + b'[L_2])$.) Let $H := \lfloor \eta^3 N/Q \rfloor$. Note that $H > 1$ by the assumption on N. Suppose that $\ell' \in \frac{d}{M}(a'[H] + b'[H])$. Then

$$\Big| \sum_{\ell \in \frac{d}{M}(a'[L_1]+b'[L_2])} f(t+\ell+\ell') - \sum_{\ell \in \frac{d}{M}(a'[L_1]+b'[L_2])} f(t+\ell) \Big| \ll \max(L_1, L_2) H,$$

since the two ranges of summation differ in only $O(\max(L_1, L_2)H)$ terms and f is 1-bounded. Averaging over ℓ', it follows from this and (5.4) that

$$\frac{1}{H^2} \Big| \sum_{\ell \in \frac{d}{M}(a'[L_1]+b'[L_2])} \sum_{\ell' \in \frac{d}{M}(a'[H]+b'[H])} f(t+\ell+\ell') \Big| \geqslant \eta N^2/Q^2 - O(\max(L_1, L_2)H) \gg \eta N^2/Q^2$$

assuming η small enough, using here that $\max(L_1, L_2) \ll N/\eta Q$. By the triangle inequality, and multiplying through by H^2,

$$\sum_{\ell \in \frac{d}{M}(a'[L_1]+b'[L_2])} \Big| \sum_{\ell' \in \frac{d}{M}(a'[H]+b'[H])} f(t+\ell+\ell') \Big| \gg \eta^7 N^4/Q^4. \tag{5.5}$$

At this point we note for future reference that the assumption of the lemma and the fact that $I_1, I_2 \subseteq [\pm N/\eta Q]$ imply that $L_1, L_2 \gg \eta^2 N/Q$. Therefore, since we are assuming $N \geqslant Q^2/\eta^3$, we have $L_1, L_2 \geqslant Q$.

Now the maximum number of ways to write any $u \in \mathbf{Z}$ as $a'x_1 + b'x_2$ with $x_1 \in [L_1]$, $x_2 \in [L_2]$ is $\ll 1 + L_1/Q \ll L_1/Q$, since if there is one solution then any other one is given by $x_1' = x_1 + \lambda b'$, $x_2' = x_2 - \lambda a'$, using here that $\gcd(a', b') = 1$. It follows from this observation and (5.5) that

$$\sum_{x \in t + \mathbf{Z}} \Big| \sum_{\ell' \in \frac{d}{M}(a'[H]+b'[H])} f(x+\ell') \Big| \gg \frac{\eta^7 N^4}{Q^4} \cdot \frac{Q}{L_1} \gg \frac{\eta^8 N^3}{Q^2}.$$

Note here that since $|da'H| = |aH| \leqslant QH < \eta^3 N$, and similarly for $db'H$, and since $|t| < 2N/\eta$, and since $\mathrm{supp}(f) \subseteq [\pm N]$, we may restrict x to the range $\{x \in t + \mathbf{Z} : |x| \ll N/\eta\}$. Then by Cauchy–Schwarz we have

$$\sum_{x \in t+\mathbf{Z}} \sum_{\ell_1', \ell_2' \in \frac{d}{M}(a'[H]+b'[H])} f(x+\ell_1') \overline{f(x+\ell_2')} \gg \eta^{17} N^5/Q^4,$$

or equivalently

$$\|f\|_{U_{\mathrm{GP}}[N;\mu]}^2 \gg \eta^{17/4}, \tag{5.6}$$

where μ is the uniform measure on the multiset $\big(\frac{d}{M}(a'[H]+b'[H])-t\big)\cap\mathbf{Z}$. Since $t=\frac{d}{M}(a'\ell_1+b'\ell_2)$, this multiset has size at least $\gg H^2/M$.

To apply Lemma 5.2 we need an upper bound on $\|\mu\|_2^2$. Similarly to the remark above, for any $u\in\mathbf{Z}$ the number of ways to write $u=a'x_1+b'x_2$ with $x_1,x_2\in[H]$ is $\ll 1+H/Q\ll H/Q$, and so

$$\|\mu\|_2^2=\sum_u\mu(u)^2\leqslant\sup_{u\in\mathbf{Z}}\mu(u)\ll\frac{H}{Q}\cdot\frac{M}{H^2}\ll 1/\eta^3N.$$

Applying Lemma 5.2 (and recalling (5.6)) then gives $\|f\|_{U^2[N]}\gg\eta^3$, as required. □

The key ingredient in our Type II estimate is the following result of Kravitz, Kuca, and Leng [32, Corollary 6.4]. We state it now, and for the sake completeness (and because some of our notation is a little different) we reproduce a proof of Lemma 5.5 in the appendix.

Lemma 5.5. *Fix $\delta\in(0,1/2)$, $s\geqslant 1$, I an indexing set and m a power of 2. Suppose that for $i\in I$ and $j\in[s]$, we have symmetric probability measures μ_{ji} supported on $[\pm N]$. Let $f:\mathbf{Z}\to\mathbf{C}$ be 1-bounded and such that*

$$\mathbb{E}_{i\in I}\|f\|^{2^s}_{U_{\mathrm{GP}}[N;\mu_{1i},\dots,\mu_{si}]}\geqslant\delta.$$

Then there exists $t=t_0(m,s)$ such that

$$\mathbb{E}_{i_1,\dots,i_t\in I}\|f\|^{2^d}_{U_{\mathrm{GP}}[N;\Omega]}\geqslant\delta^{O_{m,s}(1)},$$

where Ω is the collection of probability measures

$$\Omega=(\mu_{ji_{k_1}}*\cdots*\mu_{ji_{k_m}})_{j\in[s],1\leqslant k_1<k_2<\cdots<k_m\leqslant t}$$

where $d=s\cdot\binom{t}{m}$.

6. Type I up to $X^{1/2}/(\log X)^A$

In this section we prove Proposition 4.2. We remark that the crucial manipulation, which occurs at the end of the proof, is essentially the base case for the concatenation proven by Peluse [43] (see [43, Lemma 5.4]).

Proof. Replacing δ by $\min(\delta,|\ell|^{-1})$, we may assume throughout that $|\ell|\leqslant 1/\delta$.

We start with the assumption (4.15), that is to say

$$\sum_{\mathfrak{d}:N\mathfrak{d}\sim L}\Big|\sum_{\substack{\mathfrak{d}|\mathfrak{a}\\ N\mathfrak{a}\leqslant X}}w(\mathfrak{a})\Big|\geqslant\delta X. \tag{6.1}$$

We are working under the assumption that $w=f\boxtimes_\ell f'$ (see Definition 2.2) and that f,f' are supported on $\mathbf{Z}\cap[\pm 2nX^{1/2}]$. The aim is to prove that both f and f' have large Gowers $U^3[X^{1/2}]$-norm. We prove the bound for f; the proof of f' is symmetric.

The initial stages of the argument consist in manipulating things so that we can work with the fact that w is in product form. First, split the sum over $\mathfrak{d}$ according to which ideal class $\mathfrak{d}$ belongs to. For some such class, the sum is $\gg\delta X$. Let $g\in C(K)$ be the inverse of this class, and let $\mathfrak{c}_g,\mathfrak{c}'_{g^{-1}}$ be as in Lemma 2.1. Since w is supported on principal ideals, the inner sum over $\mathfrak{a}$ is supported on ideals of form $\mathfrak{a}=\mathfrak{d}\mathfrak{b}$, where $\mathfrak{b}\in[g]$. By Lemma 2.1, we may parametrise these ideals by $\mathfrak{b}=(z_2)\mathfrak{c}_g$ where $z_2\in\Pi_g\setminus\{0\}$ and $\mathfrak{d}=(z_1)\mathfrak{c}'_{g^{-1}}$, where $z_1\in\Pi'_{g^{-1}}\setminus\{0\}$, where $\Pi_g,\Pi'_{g^{-1}}$ are the sublattices of $\mathbf{Z}[\sqrt{-n}]$ described in Lemma 2.1.

This parametrisation covers each $\mathfrak{b}$ and $\mathfrak{d}$ a total of r^2 times, since associate values of z_1, z_2 (i.e. differing up to a unit) give the same $\mathfrak{b}$ and $\mathfrak{d}$. Thus (6.1) gives

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}}\Big|\sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |z_1z_2|\leqslant X^{1/2}/\gamma}} w((\gamma z_1z_2))\Big| \gg \delta X. \tag{6.2}$$

Here, $\gamma_1 := (N\mathfrak{c}'_{g^{-1}})^{-1}$ and $\gamma = \gamma_g$ is the generator of $\mathfrak{c}_g\mathfrak{c}'_{g^{-1}}$, a positive rational of height $O_n(1)$ (see Lemma 2.1 (ii)). Using the fact that $w = f\boxtimes_\ell f'$ gives

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}}\Big|\sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |z_1z_2|\leqslant X^{1/2}/\gamma}} \Big(\frac{z_1z_2}{|z_1z_2|}\Big)^\ell f\big(\gamma\operatorname{Re}(z_1z_2)\big)f'\big(n^{-1/2}\gamma\operatorname{Im}(z_1z_2)\big)\Big| \gg \delta X. \tag{6.3}$$

(Note here that all associates of γz_1z_2 are automatically included in the sum (6.2), since the set Π_g is invariant under multiplication by units.) This may be written as

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}}\Big|\sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |\operatorname{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\operatorname{Im}(z_2)|\leqslant C(X/L)^{1/2}}} \Psi_\ell\Big(\frac{\gamma z_1z_2}{X^{1/2}}\Big)f\big(\gamma\operatorname{Re}(z_1z_2)\big)f'\big(n^{-1/2}\gamma\operatorname{Im}(z_1z_2)\big)\Big| \gg \delta X, \tag{6.4}$$

where

$$\Psi_\ell(z) := \Big(\frac{z}{|z|}\Big)^\ell 1_{|z|\leqslant 1}. \tag{6.5}$$

Note here that (for an appropriately large constant C) we can insert the two conditions $|\operatorname{Re}(z_2)|\leqslant C(X/L)^{1/2}$ and $n^{-1/2}|\operatorname{Im}(z_2)|\leqslant C(X/L)^{1/2}$ without harm, since f, f' are supported on $[\pm O(X^{1/2})]$ and so $|z_2|\ll (X/L)^{1/2}$ on the support of the sum. It will be useful to carry a condition of this type through the proof, and we have written it in a Cartesian form with later manipulations in mind.

The next step is to perform 'cosmetic surgery' in the sense of Harman [21] to remove the cutoff involving Ψ_ℓ, using Fourier analysis to replace this by products of phases in $\operatorname{Re}(z_1z_2)$ and $\operatorname{Im}(z_1z_2)$. This allows us to fully decouple into functions of $\operatorname{Re}(z_1z_2)$ and $\operatorname{Im}(z_1z_2)$.

Lemma 6.1. *Let $\ell\in\mathbf{Z}$ and let $\varepsilon>0$. Then there is a smooth approximation $\tilde\Psi_\ell:\mathbf{C}\to\mathbf{C}$ with $\|\tilde\Psi_\ell\|_\infty\leqslant 1$, such that $\tilde\Psi_\ell$ and Ψ_ℓ agree outside of the domain*

$$\mathcal{D}_\varepsilon := \{z\in\mathbf{C}:|z|<\varepsilon\}\cup\{z\in\mathbf{C}:1-\varepsilon<|z|<1+\varepsilon\} \tag{6.6}$$

and such that for $z\in\mathbf{C}$ we have the Fourier expansion

$$\tilde\Psi_\ell(z) = \int_{\mathbf{R}^2} W(\xi_1,\xi_2)e(\xi_1\operatorname{Re}z+\xi_2\operatorname{Im}z)d\xi_1d\xi_2 \tag{6.7}$$

with

$$\|W\|_{L^1(\mathbf{R}^2)}\ll(1+|\ell|)^4\varepsilon^{-4}. \tag{6.8}$$

Proof. Identify $\mathbf{C}$ with $\mathbf{R}^2$ in the usual way. Let $\psi\in C_0^\infty(\mathbf{R}^2)$ be supported on the unit ball and have $\int\psi=1$, and set $\tilde\Psi_0 := (\Psi_0*\psi(\frac{\cdot}{\varepsilon}))(1-\psi(\frac{\cdot}{\varepsilon}))$. Since $\psi(\cdot/\varepsilon)$ is supported on the ball of radius ε, it follows that Ψ_0 and $\tilde\Psi_0$ agree outside of $\mathcal{D}_\varepsilon$. By applications of the Leibniz rule, any kth derivative of $\tilde\Psi_0$ is bounded by $O_k(\varepsilon^{-k})$ in $L^\infty(\mathbf{R}^2)$.

One can check that the kth derivatives of $z\mapsto(z/|z|)^\ell$ are (considered as a function from $\mathbf{R}^2$ to $\mathbf{C}$) are bounded by $\ll_k(1+|\ell|)^k\varepsilon^{-k}$ in the domain $\mathcal{D}_\varepsilon$. To see this note that $(z/|z|)^\ell=(z/\bar z)^{\ell/2}$, and thus

$$\Big|\frac{\partial^j}{\partial z^j}\frac{\partial^{j'}}{\partial\bar z^{j'}}(z/|z|)^\ell\Big| = \Big|z^{\ell/2-j}\bar z^{-\ell/2-j'}\prod_{i=0}^{\ell-1}\Big(\frac{\ell}{2}-i\Big)\prod_{i'=0}^{\ell'-1}\Big(\frac{\ell}{2}-i'\Big)\Big| \ll_{j,j''}(|\ell|+1)^{j+j'}|z|^{-j-j'}.$$

As the standard coordinates on $\mathbf{R}^2$ are $(z+\overline{z})/2$ and $(z-\overline{z})/2$ the result follows easily.

Now define $\tilde{\Psi}_\ell(z) := (z/|z|)^\ell \tilde{\Psi}_0(z)$. This has the desired support properties and, by the above derivative bounds and more applications of the Leibniz rule, has kth derivatives bounded by $\ll_k (1+|\ell|)^k \varepsilon^{-k}$. Since $\tilde{\Psi}_\ell$ is smooth it has a Fourier expansion (6.7) with $W(\xi_1,\xi_2) = \widehat{\tilde{\Psi}_\ell}(\xi_1,\xi_2)$. Integrating the definition of Fourier transform twice by parts in each variable gives

$$|W(\xi_1,\xi_2)| \ll |\xi_1|^{-2}|\xi_2|^{-2}\|\partial_x^2\partial_y^2\tilde{\Psi}_\ell\|_\infty \ll (1+|\ell|)^4\varepsilon^{-4}|\xi_1|^{-2}|\xi_2|^{-2}.$$

Combining this with the trivial bound $|W(\xi_1,\xi_2)| \ll 1$ gives that $\|W\|_{L^1(\mathbf{R}^2)} \ll (1+|\ell|)^4\varepsilon^{-4}$, which is the desired bound (6.8). □

Returning to the main line of argument, we replace Ψ_ℓ in (6.4) with the smoothed approximant $\tilde{\Psi}_\ell$ from Lemma 6.1 for an appropriate value of ε. Since f, f' are 1-bounded, the error in doing this is bounded by

$$\#\{(z_1,z_2) \in \Pi'_{g^{-1}} \times \Pi_g : |z_1| \asymp L^{1/2}, \gamma z_1 z_2/X^{1/2} \in \mathcal{D}_\varepsilon\},$$

where $\mathcal{D}_\varepsilon$ is the domain (6.6). Since the lattices $\Pi_g, \Pi'_{g^{-1}}$ are both contained in $\mathbf{Z}[\sqrt{-n}]$, one may see that this error is $\ll \varepsilon X$. Taking $\varepsilon = c\delta$ for an appropriately small constant $c \asymp_n 1$, this error may be absorbed by the RHS of (6.4), and so we indeed have

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}} \Big| \sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |\mathrm{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\mathrm{Im}(z_2)|\leqslant C(X/L)^{1/2}}} \tilde{\Psi}_\ell\Big(\frac{\gamma z_1 z_2}{X^{1/2}}\Big) f\big(\gamma\,\mathrm{Re}(z_1z_2)\big) f'\big(n^{-1/2}\gamma\,\mathrm{Im}(z_1z_2)\big)\Big| \gg \delta X. \tag{6.9}$$

Applying the Fourier expansion (6.7), we now have

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}} \Big| \int_{\mathbf{R}^2} W(\xi_1,\xi_2) \sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |\mathrm{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\mathrm{Im}(z_2)|\leqslant C(X/L)^{1/2}}} f_{\xi_1}(\gamma\,\mathrm{Re}(z_1z_2)) f'_{\xi_2}(n^{-1/2}\gamma\,\mathrm{Im}(z_1z_2)) d\xi_1 d\xi_2 \Big| \gg \delta X.$$

where

$$f_{\xi_1}(t) := f(t)e\Big(\frac{\xi_1 t}{X^{1/2}}\Big), \qquad f'_{\xi_2}(t) := f'(t)e\Big(\frac{n^{1/2}\xi_2 t}{X^{1/2}}\Big).$$

By (6.8) (and recalling that $\max(1,|\ell|) \leqslant 1/\delta$), there is some choice of ξ_1, ξ_2 such that

$$\sum_{\substack{z_1\in\Pi'_{g^{-1}}\\ |z_1|^2\sim\gamma_1 L}} \Big| \sum_{\substack{z_2\in\Pi_g\setminus\{0\}\\ |\mathrm{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\mathrm{Im}(z_2)|\leqslant C(X/L)^{1/2}}} f_{\xi_1}(\gamma\,\mathrm{Re}(z_1z_2)) f'_{\xi_2}(n^{-1/2}\gamma\,\mathrm{Im}(z_1z_2))\Big| \gg \delta^9 X. \tag{6.10}$$

Now we pass to real and imaginary parts, writing $z_1 = a + b\sqrt{-n}$ and $z_2 = x + y\sqrt{-n}$ where $a,b,x,y \in \mathbf{Z}$ (since, by Lemma 2.1, $\Pi_g, \Pi'_{g^{-1}}$ are sublattices of $\mathbf{Z}[\sqrt{-n}]$). Set

$$\Gamma := \{(u,v) : u + v\sqrt{-n} \in \Pi'_{g^{-1}}\}, \qquad \Gamma' := \{(u,v) : u + v\sqrt{-n} \in \Pi_g\}; \tag{6.11}$$

thus Γ, Γ' are sublattices of $\mathbf{Z}^2$, of index $O_n(1)$. Then (6.10) implies that

$$\sum_{\substack{(a,b)\in\Gamma\\ |a|,|b|\ll L^{1/2}}} \Big| \sum_{\substack{(x,y)\in\Gamma'\\ |x|,|y|\leqslant C(X/L)^{1/2}}} f_{\xi_1}(\gamma(ax-nby)) f'_{\xi_2}(\gamma(bx+ay))\Big| \gg \delta^9 X. \tag{6.12}$$

Recall that our task is to show that f and f' have large Gowers U^3-norms. This norm is invariant under multiplication by linear phases, as can immediately be seen from the definition (4.11) since the second and higher derivatives of any linear phase equal 1. Therefore, to achieve our task we

may replace f, f' by $fe(\lambda\cdot)$ and $f'e(\lambda'\cdot)$ for any λ, λ'. In other words, we may assume henceforth that $\xi_1 = \xi_2 = 0$. Changing variables to $a_1 := \gamma a$, $b_1 := -n\gamma b$, we have

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}}} \Big| \sum_{\substack{(x,y)\in\Gamma'\\ |x|,|y|\leqslant C(X/L)^{1/2}}} f(a_1x+b_1y)f'\Big(-\frac{b_1}{n}x+a_1y\Big)\Big| \gg \delta^9 X. \tag{6.13}$$

Here, $M = O_n(1)$ is the denominator of γ; note that, since $\Gamma \subseteq \mathbf{Z}^2$, we do indeed have $a_1, b_1 \in \frac{1}{M}\mathbf{Z}$ for all terms in the sum (6.12), and there is no problem with any extra terms being included since the summands are non-negative.

Now we restrict to the situation where Ma_1, Mb_1 are essentially coprime. Note that as there are $\ll \sum_{d\geqslant D} L/d^2 \ll L/D$ pairs with $\gcd(\ell,\ell') \geqslant D$, we can further restrict the sum over a_1, b_1 to those terms with $\gcd(Ma_1, Mb_1) \ll \delta^{-9}$. By a further pigeonholing we may restrict to some fixed value $\gcd(Ma_1, Mb_1) = d$, $1 \leqslant d \ll \delta^{-9}$, such that we have

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}} \Big| \sum_{\substack{(x,y)\in\Gamma'\\ |x|,|y|\leqslant C(X/L)^{1/2}}} f(a_1x+b_1y)f'\Big(-\frac{b_1}{n}x+a_1y\Big)\Big| \gg \delta^{18} X. \tag{6.14}$$

Let $t = -\frac{b_1}{n}x + a_1y$; via triangle inequality and the 1-boundedness of f' and the fact that f' is supported on $[\pm O(X^{1/2})]$ we have that

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}} \sum_{\substack{t\in\frac{1}{Mn}\mathbf{Z}\\ |t|\ll X^{1/2}}} \Big| \sum_{\substack{(x,y)\in\Gamma'\\ |x|,|y|\leqslant C(X/L)^{1/2}\\ -\frac{b_1}{n}x+a_1y=t}} f(a_1x+b_1y)\Big| \gg \delta^{18} X. \tag{6.15}$$

Now it is time to deal with the constraint $(x,y) \in \Gamma'$. For this we use the orthogonality relation

$$1_{\Gamma'}(x,y) = \mathbb{E}_{(\zeta,\zeta')\in(\Gamma')^\perp/\mathbf{Z}^2} e(\zeta x + \zeta' y), \tag{6.16}$$

where $(\Gamma')^\perp$ denotes the dual lattice of Γ'. Substituting into (6.15), it follows that there is some choice of ζ, ζ' such that

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}} \sum_{\substack{t\in\frac{1}{Mn}\mathbf{Z}\\ |t|\ll X^{1/2}}} \Big| \sum_{\substack{x,y\in\mathbf{Z}\\ |x|,|y|\leqslant C(X/L)^{1/2}\\ -\frac{b_1}{n}x+a_1y=t}} f(a_1x+b_1y)e(\zeta x+\zeta' y)\Big| \gg \delta^{18} X.$$

Now we apply Cauchy–Schwarz, which gives

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}} \sum_{\substack{x,y,x',y'\in\mathbf{Z}\\ |x|,|x'|,|y|,|y'|\leqslant C(X/L)^{1/2}\\ b_1x-a_1yn=b_1x'-a_1y'n}} f(a_1x+b_1y)\overline{f(a_1x'+b_1y')}e(\zeta(x-x')+\zeta'(y-y')) \gg \delta^{36}X^{3/2}/L.$$

Note that $b_1x - a_1yn = b_1x' - a_1y'n$ and the coprimality of $Ma_1/d, Mb_1/d$ imply that $x' - x = (Ma_1/d)h$ and $y' - y = (Mb_1/dn)h$ for some $h \in \mathbf{Z}$; therefore the above gives

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}} \sum_{\substack{x,y,h\in\mathbf{Z}\\ |x|,|y|\leqslant C(X/L)^{1/2}\\ |h|\ll X^{1/2}/L}} e\Big(\Big(a_1\zeta+\frac{b_1\zeta'}{n}\Big)\frac{Mh}{d}\Big)\Delta_{M(na_1^2+b_1^2)h/dn}f(a_1x+b_1y) \gg \delta^{36}X^{3/2}/L.$$

By the triangle inequality we may eliminate the phase, obtaining

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ |a_1|,|b_1|\ll L^{1/2}\\ \gcd(Ma_1,Mb_1)=d}}\sum_{\substack{h\in\mathbf{Z}\\ |h|\ll X^{1/2}/L}}\Big|\sum_{\substack{x,y\in\mathbf{Z}\\ |x|,|y|\leqslant C(X/L)^{1/2}}}\Delta_{M(na_1^2+b_1^2)h/dn}f(a_1x+b_1y)\Big|\gg\delta^{36}X^{3/2}/L.$$

It follows that there are $\gg\delta^{36}X^{1/2}$ triples (a_1,b_1,h) with $|a_1|,|b_1|\ll L^{1/2}$ and $|h|\ll X^{1/2}/L$ and with $\gcd(Ma_1,Mb_1)=d\ll\delta^{-9}$ such that

$$\Big|\sum_{\substack{x,y\in\mathbf{Z}\\ |x|,|y|\leqslant C(X/L)^{1/2}}}\Delta_{M(na_1^2+b_1^2)h/dn}f(a_1x+b_1y)\Big|\gg\delta^{36}X/L.$$

By discarding exceptions, the same is true with the additional condition that $|a_1|,|b_1|\gg\delta^{36}L^{1/2}$.

Our aim now is to apply Lemma 5.4 with $\eta=\delta^{C_1}$ and $Q=C_2L^{1/2}$ for appropriate constants C_1,C_2, and with $N=2nX^{1/2}$. If $C_1>12$ and if C_2 is large enough then all the conditions of that lemma are satisfied. (The condition $N\geqslant Q^2/\eta^3$ is satisfied due to the assumption $L\leqslant\delta^{C_{4.2}}X^{1/2}$ in the statement of Proposition 4.2, assuming $C_{4.2}$ large enough.)

Applying Lemma 5.4, we have for each such a_1,b_1,h the lower bound

$$\|\Delta_{M(a_1^2n+b_1^2)h/dn}f\|_{U^2[X^{1/2}]}\gg\delta^{O(1)}.$$

Squaring and summing over a_1,b_1,h (dropping the condition $\gcd(Ma_1,Mb_1)=d$, which has now served its purpose, which we may by positivity) yields

$$\sum_{\substack{a_1,b_1\in\frac{1}{M}\mathbf{Z}\\ \delta^{12}L^{1/2}\ll|a_1|,|b_1|\ll L^{1/2}}}\sum_{|h|\ll X^{1/2}/L}\|\Delta_{M(na_1^2+b_1^2)h/dn}f\|_{U^2[X^{1/2}]}^2\gg\delta^{O(1)}X^{1/2}.\tag{6.17}$$

The idea now that that, as a_1,b_1,h range over $a_1,b_1\in\frac{1}{M}\mathbf{Z}$ with $\delta^{12}L^{1/2}\ll|a_1|,|b_1|\ll L^{1/2}$ and $h\in\mathbf{Z}$ with $|h|\ll X^{1/2}/L$, then $M(a_1^2n+b_1^2)h/dn$ ranges somewhat uniformly over its range. If it was *exactly* uniform then the LHS would be essentially $X^{1/2}$ times the $U^3[O(X^{1/2})]$-norm of f. Making this heuristic rigorous is the first instance of "concatenation" in the paper.

The contribution from $h=0$ is $\ll L$ and can be ignored by the assumption that $L\leqslant\delta^{C_{4.2}}X^{1/2}$. Writing, for $t\in\mathbf{Z}$,

$$\sigma(t):=\#\Big\{(a_1,b_1,h)\in\frac{1}{M}\mathbf{Z}\times\frac{1}{M}\mathbf{Z}\times\mathbf{Z}:\delta^{12}L^{1/2}\ll|a_1|,|b_1|\ll L^{1/2},\\ 0<|h|\ll(X/L)^{1/2},M(a_1^2n+b_1^2)h/dn=t\Big\},$$

we can rewrite (6.17) (with the $h=0$ term removed) as

$$\sum_{\substack{t\in\mathbf{Z}\\ 0<|t|\ll\delta^{-O(1)}X^{1/2}}}\sigma(t)\|\Delta_tf\|_{U^2[X^{1/2}]}^2\gg\delta^{O(1)}X^{1/2}.\tag{6.18}$$

(We can restrict to $t\in\mathbf{Z}$, since $\Delta_tf\equiv0$ otherwise on account of f being supported on $\mathbf{Z}$, and from the definition we see that $\sigma(0)=0$.)

Now if (a_1,b_1,h) is one of the triples being counted in $\sigma(t)$, we have $h|tdn$, and hence $\sigma(t)\leqslant\sum_{h|dnt}r(Mdnt/h)$, where $r(m)$ is the number of representations of m as u^2n+v^2 for integer u,v (which, for us, are Ma_1 and Mb_1). Since $r\ll\tau$ (see Lemma D.2) it follows that $\sigma(t)\ll\sum_{h|dnt}\tau(Mdnt/h)\leqslant\tau(Mdnt)^2\ll\tau(t)^2$. By Lemma D.1 with $m=4$ we have

$$\sum_{0<|t|\ll\delta^{-O(1)}X^{1/2}}\sigma(t)^2\ll\delta^{-O(1)}X^{1/2}(\log X)^{15}.$$

Applying Cauchy-Schwarz to (6.18) then gives

$$\sum_{\substack{t\in\mathbf{Z}\\ 0<|t|\ll\delta^{-O(1)}X^{1/2}}} \|\Delta_t f\|_{U^2[X^{1/2}]}^4 \gg \delta^{O(1)}(\log X)^{-15}X^{1/2}.$$

Expanding the left-hand side using the definitions (4.11) and (4.12), we obtain

$$\|f\|_{U^3(\mathbf{Z})}^8 = \sum_{t,x,h_1,h_2\in\mathbf{Z}} \Delta_{h_1}\Delta_{h_2}\Delta_t f(x) \gg \delta^{O(1)}(\log X)^{-15}X^2.$$

Finally, Proposition 4.2 follows from this and (4.12). □

7. Type II estimates up to $X^{1/2-o(1)}$

In this section we prove Proposition 4.3. As in the proof of Proposition 4.2, we may assume that $|\ell| \leqslant 1/\delta$ throughout. We handle the bound involving f', the one with f being essentially identical. We begin, of course, with the assumption (4.16), that is to say

$$\Big|\sum_{\substack{\mathfrak{a},\mathfrak{b}:N\mathfrak{a}\sim L\\ N\mathfrak{a}\mathfrak{b}\leqslant X}} \alpha_{\mathfrak{a}}\beta_{\mathfrak{b}}w(\mathfrak{a}\mathfrak{b})\Big| \geqslant \delta X,$$

and recall once again that $w = f\boxtimes_\ell f'$ is in product form (Definition 2.2) and that f, f' are supported on $[\pm 2nX^{1/2}]$.

By pigeonhole and the fact that w is supported on principal ideals, there is some ideal class $g \in C(K)$ such that the contribution from $\mathfrak{a} \in [g^{-1}]$ is $\gg \delta X$. By Cauchy, triangle and the 1-boundedness of α, β it follows that

$$\sum_{\substack{\mathfrak{a},\mathfrak{a}'\in[g^{-1}]\\ N\mathfrak{a},N\mathfrak{a}'\sim L}} \Big|\sum_{\substack{\mathfrak{b}\in[g]\\ N\mathfrak{a}\mathfrak{b},N\mathfrak{a}'\mathfrak{b}\leqslant X}} w(\mathfrak{a}\mathfrak{b})\overline{w(\mathfrak{a}'\mathfrak{b})}\Big| \gg \delta^2 XL. \tag{7.1}$$

Now we parametrise ideals as in Section 6. Let $\mathfrak{c}_g, \mathfrak{c}'_{g^{-1}}$ be as in Lemma 2.1, and parametrise $\mathfrak{b}$ by $(z_2)\mathfrak{c}_g$ with $z_2 \in \Pi_g \setminus \{0\}$, and $\mathfrak{a}, \mathfrak{a}'$ by $(z_1)\mathfrak{c}'_{g^{-1}}$, $(z_1')\mathfrak{c}'_{g^{-1}}$ respectively, where $z_1 \in \Pi'_g \setminus \{0\}$. Here $\Pi_g, \Pi'_{g^{-1}}$ are the sublattices of $\mathbf{Z}[\sqrt{-n}]$ described in Lemma 2.1. The parametrisation covers each triple $\mathfrak{a}, \mathfrak{a}', \mathfrak{b}$ a total of r^3 times due to associate values of z_1, z_1', z_2 giving the same ideals. With these parametrisations, (7.1) becomes

$$\sum_{\substack{z_1,z_1'\in\Pi'_{g^{-1}}\\ |z_1|,|z_1'|\sim\gamma_1 L}} \Big|\sum_{\substack{z_2\in\Pi_g\\ |z_1z_2|,|z_1'z_2|\leqslant X^{1/2}/\gamma}} w((\gamma z_1z_2))\overline{w((\gamma z_1'z_2))}\Big| \gg \delta^2 XL.$$

Here, $\gamma_1 := (N\mathfrak{c}'_{g^{-1}})^{-1}$ and $\gamma = \gamma_g$ is the generator of $\mathfrak{c}_g\mathfrak{c}'_{g^{-1}}$, a positive rational of height $O_n(1)$ (see Lemma 2.1 (ii)). Using the fact that $w = f\boxtimes_\ell f'$ gives

$$\sum_{\substack{z_1,z_1'\in\Pi'_{g^{-1}}\\ |z_1|^2,|z_1'|^2\sim\gamma_1 L}} \Big|\sum_{\substack{z_2\in\Pi_g\\ |\mathrm{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\mathrm{Im}(z_2)|\leqslant C(X/L)^{1/2}}} \Psi_\ell\Big(\frac{\gamma z_1z_2}{X^{1/2}}\Big)\overline{\Psi}_\ell\Big(\frac{\gamma z_1'z_2}{X^{1/2}}\Big) f\big(\gamma\,\mathrm{Re}(z_1z_2)\big) f'\big(n^{-1/2}\gamma\,\mathrm{Im}(z_1'z_2)\big)\times$$
$$\times\,\overline{f\big(\gamma\,\mathrm{Re}(z_1'z_2)\big) f'\big(n^{-1/2}\gamma\,\mathrm{Im}(z_1'z_2)\big)}\Big| \gg \delta^2 XL.$$

with Ψ_ℓ as in (6.5). The two rough cutoffs Ψ_ℓ may be replaced by the smooth cutoffs $\tilde{\Psi}_\ell$ constructed in Lemma 6.1, now with $\varepsilon = c\delta^2$. Analogously to (6.10), we obtain

$$\sum_{\substack{z_1,z_1'\in\Pi'_{g^{-1}}\\ |z_1|^2,|z_1'|^2\sim\gamma_1 L}}\Bigg|\sum_{\substack{z_2\in\Pi_g\\ |\mathrm{Re}(z_2)|\leqslant C(X/L)^{1/2}\\ n^{-1/2}|\mathrm{Im}(z_2)|\leqslant C(X/L)^{1/2}}} f_{\xi_1}\big(\gamma\,\mathrm{Re}(z_1z_2)\big)f'_{\xi_2}\big(n^{-1/2}\gamma\,\mathrm{Im}(z_1'z_2)\big)\times$$
$$\times\,\overline{f_{\xi_1'}\big(\gamma\,\mathrm{Re}(z_1'z_2)\big)f'_{\xi_2'}\big(n^{-1/2}\gamma\,\mathrm{Im}(z_1'z_2)\big)}\Bigg| \gg \delta^{26}XL,$$

where $f_{\xi_1}, f'_{\xi_2}, f_{\xi_1'}, f'_{\xi_2'}$ are twists of f, f', f, f' respectively by fixed additive characters (whose precise form is irrelevant).

Now we pass to real and imaginary parts, writing $z_1 = a + b\sqrt{-n}$, $z_1' = a' + b'\sqrt{-n}$ and $z_2 = x + y\sqrt{-n}$ where $a, b, a', b', x, y \in \mathbf{Z}$. This gives, with Γ, Γ' the lattices defined in (6.11),

$$\sum_{\substack{a,a',b,b'\in\mathbf{Z}\\ |a|,|a'|,|b|,|b'|\ll L^{1/2}}}\Bigg|\sum_{\substack{(x,y)\in\Gamma\\ |x|,|y|\leqslant C(X/L)^{1/2}}} f_{\xi_1}(\gamma(ax-nby))f'_{\xi_2}(\gamma(bx+ay))\times$$
$$\times\,\overline{f_{\xi_1'}(\gamma(a'x-nb'y))f'_{\xi_2'}(\gamma(b'x+a'y))}\Bigg| \gg \delta^{26}XL.$$

Henceforth, write $f_1 = f_{\xi_1}$, $f_2 = f'_{\xi_2}$, $f_3 = \overline{f_{\xi_1'}}$, $f_4 = \overline{f'_{\xi_2'}}$. Detecting the condition $(x,y) \in \Gamma'$ using (6.16), it follows by the triangle inequality that for some ζ, ζ' we have

$$\sum_{\substack{a,a',b,b'\in\mathbf{Z}\\ |a|,|a'|,|b|,|b'|\ll L^{1/2}}}\Bigg|\sum_{\substack{x,y\in\mathbf{Z}\\ |x|,|y|\leqslant C(X/L)^{1/2}}} f_1(\gamma(ax-nby))f_2(\gamma(bx+ay))\times$$
$$\times\, f_3(\gamma(a'x-nb'y))f_4(\gamma(b'x+a'y))e(\zeta x+\zeta' y)\Bigg| \gg \delta^{26}XL. \tag{7.2}$$

We now select a set $\mathcal{Q}$ of $\gg \delta^{26}L^2$ quadruples (a,a',b,b') for which the inner sum is $\gg \delta^{26}X/L$, and satisfying some further gcd, size and nondegeneracy conditions, the need for which will become apparent later, but which we need to select for now. The conditions are

$$\gcd(a,b), \gcd(a',b') \ll \delta^{-26}, \quad \delta^{26}L^{1/2} \ll |a|, |a'|, |b|, |b'| \ll L^{1/2}, \quad a'b - ab', aa' + nbb' \neq 0. \tag{7.3}$$

First note that each quadruple (a,a',b,b') contributes at most $O(X/L)$ to be sum (7.2). Therefore there are at least $\gg \delta^{26}L^2$ quadruples $\mathcal{Q}'$ each contributing at least $\gg \delta^{26}X/L$ to this sum. In order to construct $\mathcal{Q}$, we remove quadruples from $\mathcal{Q}'$ violating (7.3). To see this is possible leaving at least half of $\mathcal{Q}'$, note that the set of quadruples satisfying $\gcd(a,b) + \gcd(a',b') \gg \delta^{-26}$ is bounded by $\sum_{d\geqslant\delta^{-26}}(L^{1/2})^4/d^2 \ll \delta^{26}L^2$. Similarly, the number of quadruples satisfying $\min(|a|,|a'|,|b|,|b'|) \leqslant \delta^{26}L^{1/2}$ is bounded by $\ll \delta^{26}L^2$. Finally if a, b are non-zero, then given a, b, a' there are at most 2 choices of b' which cause $a'b - ab' = 0$ or $aa' + nbb' = 0$; therefore there are at most $L^{3/2}$ additional violating tuples here, which is $\ll \delta^{26}L^2$ provided that $L \gg \delta^{-104}$.

For the argument over the next couple of pages it is convenient to write

$$a_1 := \gamma a, \quad a_2 := \gamma b, \quad a_3 := \gamma a', \quad a_4 := \gamma b',$$
$$b_1 := -n\gamma b, \quad b_2 := \gamma a, \quad b_3 := -n\gamma b', \quad b_4 := \gamma a'. \tag{7.4}$$

It also saves on notation to write

$$I := [\pm C(X/L)^{1/2}]. \tag{7.5}$$

Thus, for $(a,a',b,b') \in \mathcal{Q}$ we have

$$\Big| \sum_{x,y\in I} e(\zeta x + \zeta' y) \prod_{i=1}^{4} f_i(a_i x + b_i y) \Big| \gg \delta^{26} XL. \tag{7.6}$$

We are now going to bound the LHS of (7.6) in terms of a Gowers-Peluse norm of f_4, in preparation for the application of concatenation results. Set

$$H := \lfloor \delta^{C_1} X^{1/2}/L \rfloor \tag{7.7}$$

for some C_1 to be specified. The important thing to note here is that if $L \sim X^{1/2-\kappa}$ for some $\kappa > 0$ then $|H| \sim \delta^{O(1)} X^{\kappa}$. The fact that this is a small power of X will be vital later (to apply Lemma 7.1 effectively), and is what constrains our Type II methods to the regime $L \ll X^{1/2-\kappa}$ for some $\kappa > 0$.

For $h \in [\pm H]$, substitute $x := x' + b_1 h$, $y := y' - a_1 h$. Then

$$\Big| \sum_{\substack{x'\in I - b_1 h \\ y' \in I + a_1 h}} e(\zeta x' + \zeta' y' + (\zeta b_1 - \zeta' a_1)h) f_1(a_1 x' + b_1 y') \prod_{j=2}^{4} f_j(a_j x' + b_j y' + (a_j b_1 - a_1 b_j)h) \Big| \gg \delta^{26} X/L.$$

We have $|a_1 h|, |b_1 h| \ll L^{1/2} H \ll \delta^{C_1} (X/L)^{1/2}$, so the error in replacing the shifted intervals $I - b_1 h$, $I + a_1 h$ by two copies of I is $\ll \delta^{C_1} X/L$, which is negligible for $C_1 > 28$. Therefore we have (replacing the dummy variables x', y' by x, y respectively)

$$\Big| \sum_{\substack{x\in I \\ y\in I}} e(\zeta x + \zeta' y + (\zeta b_1 - \zeta' a_1)h) f_1(a_1 x + b_1 y) \prod_{j=2}^{4} f_j(a_j x + b_j y + (a_j b_1 - a_1 b_j)h) \Big| \gg \delta^{26} X/L.$$

Summing over $h \in [\pm H]$ gives

$$\sum_{|h|\leqslant H} c_h \sum_{x,y\in I} e\big((\zeta b_1 - \zeta' a_1)h\big) \prod_{j=2}^{4} f_j(a_j x + b_j y + (a_j b_1 - a_1 b_j)h) \gg \delta^{26} XH/L, \tag{7.8}$$

where the c_h are complex numbers with unit modulus, chosen to make each term in the sum over h real and nonnegative. Switching the order of summation and applying Cauchy-Schwarz gives

$$\sum_{x,y\in I} \sum_{|h|,|h'|\leqslant H} c_h \overline{c}_{h'} e\big((\zeta b_1 - \zeta' a_1)h\big) e\big(-(\zeta b_1 - \zeta' a_1)h'\big) \times$$
$$\times \prod_{j=2}^{4} f_j(a_j x + b_j y + (a_j b_1 - a_1 b_j)h) \prod_{j=2}^{4} \overline{f_j(a_j x + b_j y + (a_j b_1 - a_1 b_j)h')} \gg \delta^{26} XH/L.$$

By the triangle inequality (and recalling the notation for difference operators as detailed at the beginning of Section 5) we obtain

$$\sum_{|h|,|h'|\leqslant H} \Big| \sum_{x,y\in I} \prod_{j=2}^{4} \Delta_{(a_j b_1 - a_1 b_j)(h,h')} f_j(a_j x + b_j y) \Big| \gg \delta^{52} XH^2/L.$$

Therefore for $\gg \delta^{52} H^2$ values of (h,h') we have

$$\sum_{x,y\in I} \prod_{j=2}^{4} \Delta_{(a_j b_1 - a_4 b_j)(h,h')} f_j(a_j x + b_j y) \gg \delta^{52} X/L.$$

Note that this statement is very similar to (7.6), only with a product of three functions rather than four, and with the exponent 26 replaced by 52. Consequently we may repeat the argument

two more times, obtaining at the end $\gg \delta^{208}H^6$ sextuples $(h_1, h_1', h_2, h_2', h_3, h_3')$, $h_i, h_i' \in [\pm H]$ such that

$$\sum_{x,y\in I} \Delta_{(a_4b_1-b_4a_1)(h_1,h_1')}\Delta_{(a_4b_2-b_4a_2)(h_2,h_2')}\Delta_{(a_4b_3-b_4a_3)(h_3,h_3')}f_4(a_4x+b_4y) \geqslant \delta^{208}X/L.$$

For these further two iterations of the argument to work, we need $C_1 > 104$, thus for definiteness set $C_1 := 200$. Henceforth we will not keep track of exponents explicitly.

We apply Lemma 5.4 to this statement, taking in that lemma $I_1 = I_2 = I$ with I as in (7.5), $N = X^{1/2}$, $Q = L^{1/2}$, M the denominator of γ and $\eta \sim \delta^{204}$. Most of the conditions of that lemma follow from (7.3); the condition $N \geqslant Q^2/\eta^3$ follows from $X^{1/2} \geqslant L\delta^{-1000}$, which follows from the assumptions of Proposition 4.3 if $C_{4.3}$ is large enough.

It follows that, for each of these sextuples,

$$\|\Delta_{(a_4b_1-b_4a_1)(h_1,h_1')}\Delta_{(a_4b_2-b_4a_2)(h_2,h_2')}\Delta_{(a_4b_3-b_4a_3)(h_3,h_3')}f_4\|^4_{U^2[X^{1/2}]} \gg \delta^{O(1)}.$$

We have the above inequality for a set of $\gg \delta^{O(1)}H^6$ sextuples $(h_1, h_1', h_2, h_2', h_3, h_3')$ from $[\pm H]^6$. However, $\|\Delta_{(a_4b_1-b_4a_1)(h_1,h_1')}\Delta_{(a_4b_2-b_4a_2)(h_2,h_2')}\Delta_{(a_4b_3-b_4a_3)(h_3,h_3')}f_4\|_{U^2[X^{1/2}]}$ is nonnegative and therefore we may average over all sextuples $(h_1, h_1', h_2, h_2', h_3, h_3') \in [\pm H]^6$ to obtain

$$\mathbb{E}_{(h_1,h_1',h_2,h_2',h_3,h_3')\in[\pm H]^6}\|\Delta_{(a_4b_1-b_4a_1)(h_1,h_1')}\Delta_{(a_4b_2-b_4a_2)(h_2,h_2')}\Delta_{(a_4b_3-b_4a_3)(h_3,h_3')}f_4\|^4_{U^2[X^{1/2}]} \gg \delta^{O(1)}$$

Unpacking the definition of the $U^2[X^{1/2}]$ norm via (4.11), this implies that

$$X^{-3/2}\sum_{x,h,h'\in\mathbb{Z}} \mathbb{E}_{(h_1,h_1',h_2,h_2',h_3,h_3')\in[\pm H]^6}\Delta_{(a_4b_1-b_4a_1)(h_1,h_1')}\Delta_{(a_4b_2-b_4a_2)(h_2,h_2')}$$
$$\Delta_{(a_4b_3-b_4a_3)(h_3,h_3')}\Delta_h\Delta_{h'}f_4(x) \gg \delta^{O(1)}. \tag{7.9}$$

Via switching the sum over x inward and recalling the expression in (5.2), we have that

$$X^{-1}\sum_{h,h'\in\mathbb{Z}}\sum_{q\in\mathcal{Q}}\|\Delta_h\Delta_{h'}f_4\|^8_{U_{\mathrm{GP}}[X^{1/2};\mu_{1,q},\mu_{2,q},\mu_{3,q}]} \gg \delta^{O(1)}L^2 \tag{7.10}$$

where, for $q = (a, a', b, b') \in \mathcal{Q}$ and for $j \in \{1, 2, 3\}$, $\mu_{j,q}$ is the uniform measure on $(a_4b_j - b_4a_j)[\pm H]$. Recalling the definitions (7.4) of the a_i, b_i in terms of $q = (a, a', b, b')$, we have

$$\mu_{1,q} \sim \mathrm{U}_{-(2n\gamma)^2(aa'+nbb')[\pm H]}, \quad \mu_{2,q} \sim \mathrm{U}_{(2n\gamma)^2(ab'-a'b)[\pm H]}, \quad \mu_{3,q} \sim \mathrm{U}_{-(2n\gamma)^2(a'^2+nb'^2)[\pm H]}. \tag{7.11}$$

Recall that we have this for a set $\mathcal{Q}$ of $\gg \delta^{O(1)}L^2$ quadruples $q = (a, a', b, b')$ satisfying the conditions (7.3). Summing (7.10) over all these tuples gives

$$X^{-1}\sum_{h,h'\in\mathbb{Z}}\sum_{q\in\mathcal{Q}}\|\Delta_h\Delta_{h'}f_4\|^8_{U_{\mathrm{GP}}[X^{1/2};\mu_{1,q},\mu_{2,q},\mu_{3,q}]} \gg \delta^{O(1)}L^2. \tag{7.12}$$

To drive our application of concatenation, we need to know that convolutions of different measures $\mu_{j,q}$ spread out. The following lemma encapsulates this.

Here, and for the remainder of the section, $E \lessapprox E'$ means $E \ll \delta^{-O(1)}(\log X)^{O(1)}E'$, and $E \lessapprox_m E'$ means $E \ll \delta^{-O_m(1)}(\log X)^{O_m(1)}E'$, and similarly for $E \gtrapprox E'$, $E \gtrapprox_m E'$.

Lemma 7.1. *Let $m \geqslant 1$ be an integer and let $j \in \{1, 2, 3\}$. Let H be as in (7.7), that is to say $H = \delta^{C_1}X^{1/2}/L$. Then we have*

$$\mathbb{E}_{q_1,\dots,q_m\in\mathcal{Q}}\Big\|\underset{i=1}{\overset{m}{*}}\mu_{j,q_i}\Big\|_2^2 \lessapprox_m \max(H^{-m}, X^{-1/2}).$$

Remark. The key point to note about the conclusion is that if $L = X^{1/2-\kappa}$ then the bound is $\lessapprox X^{-1/2}$ for $m > 1/\kappa$. Since all measures are supported on $[\pm O(X^{1/2})]$, this statement is asserting rough uniform distribution of the $*_{i=1}^m \mu_{j,q_i}$ on $[\pm O(X^{1/2})]$ on average over $q_1, \dots, q_m$. In particular, in the regime of interest in Proposition 4.3, where $L \leqslant X^{3/8}$, we can take $m = 8$.

Proof. We first handle the case $j = 1$, the other cases are analogous and we will indicate the minor modifications required at the end. First of all, note that

$$\Big\| \mathop{*}_{i=1}^m \mu_{j,q_i} \Big\|_2^2 = \Big(\mathop{*}_{i=1}^m (\mu_{j,q_i} * \mu_{j,q_i}) \Big)(0) \ll \Big(\mathop{*}_{i=1}^m \tilde{\mu}_{j,q_i} \Big)(0),$$

where $\tilde{\mu}_{j,q}$ is the variant of $\mu_{j,q}$ with H replaced by $2H$, since $\mu_{j,q} * \mu_{j,q} \ll \tilde{\mu}_{j,q}$ pointwise. It therefore suffices to show that

$$\mathbb{E}_{q_1,\dots,q_m \in \mathcal{Q}} \Big(\mathop{*}_{i=1}^m \tilde{\mu}_{j,q_i} \Big)(0) \lessapprox_m \max(H^{-m}, X^{-1/2}),$$

or equivalently

$$\nu^{(m)}(0) \lessapprox_m \max(H^{-m}, X^{-1/2}). \tag{7.13}$$

where ν is the probability measure $\mathbb{E}_{q \in \mathcal{Q}} \mu_{j,q}$, and $\nu^{(m)}$ denotes the mth convolution power of ν. Now $\nu \leqslant H^{-1}\delta_0 + \nu_*$, where ν_* is uniform on the multiset $\{-2n^2\gamma^2(aa' + nbb')h : (a, a', b, b') \in \mathcal{Q}, h \in [\pm 2H] \setminus \{0\}\}$. Note here that by the properties (7.3) of $\mathcal{Q}$ we have $\nu_*(0) = 0$. The idea now is to show that $\nu_*^{(2)}$ is somewhat close to uniform on $[\pm O(X^{1/2})]$ in the sense that

$$\|\nu_*^{(2)}\|_\infty \lessapprox X^{-1/2}. \tag{7.14}$$

Once this is proven, we can expand (since δ_0 is the identity in convolution)

$$\nu^{(m)} = H^{-m}\delta_0 + H^{-(m-1)} m \nu_* + \sum_{k=2}^m \binom{m}{k} H^{-(m-k)} \nu_*^{(k)}.$$

Each measure $\nu_*^{(k)}$, $k \geqslant 2$, is the convolution of $\nu_*^{(2)}$ with some probability measure, and hence by (7.14) is pointwise $\lessapprox X^{-1/2}$. Since $\nu_*(0) = 0$, the desired bound (7.13) then follows.

It remains to establish (7.14). First, note that uniformly for $t \neq 0$ we have

$$\#\{(a, a', b, b') \in S : aa' + nbb' = t\} \ll \sum_{|u| \ll L} \tau(u) \tau\Big(\frac{t-u}{n}\Big) \ll L(\log L)^3; \tag{7.15}$$

here the divisor function is extended so that $\tau(x) = \tau(|x|)$ and $\tau(x) = 0$ if $x = 0$ or $x \notin \mathbf{Z}$. It follows that $\nu_* \ll |\mathcal{Q}|^{-1} H^{-1} L(\log L)^3 \tau \lessapprox X^{-1/2}\tau$ pointwise, noting here that $|\mathcal{Q}| \gg \delta^{O(1)} L^2$ and $H = \delta^{O(1)} X^{1/2}/L$. Noting also that ν_* is supported on $[\pm O(X^{1/2})]$, it follows using Lemma D.1 in the case $m = 2$ that

$$\nu_*^{(2)}(t) \lessapprox X^{-1} \sum_{|x| \ll X^{1/2}} \tau(x)\tau(t-x) \ll X^{-1} \sum_{|x| \ll X^{1/2}} \tau(x)^2 \lessapprox X^{-1/2},$$

which is the required bound (7.14) in the case $j = 1$.

In the cases $j = 2$ and 3 we proceed entirely analogously, the only difference of any interest being the corresponding bound to (7.15). For $j = 3$ we have instead the bound

$$\#\{(a, a', b, b') \in \mathcal{Q} : a^2 + nb^2 = t\} \ll L\tau(t); \tag{7.16}$$

see Lemma D.2. This gives the bound $\nu_* \lessapprox X^{-1/2}\tau^2$ pointwise, and we can conclude as before but now using the case $m = 4$ of Lemma D.1.

The case $j = 2$ is almost identical to the case $j = 1$. Instead of (7.15) we have the bound

$$\#\{(a, a', b, b') \in \mathcal{Q} : a'b - ab' = t\} \ll L(\log L)^3, \tag{7.17}$$

which may be proven in the same manner as (7.15). □

From now on we will assume that $L \leqslant X^{3/8}$, which is one of the assumptions of Proposition 4.3. Thus, applying Lemma 7.1 with $m = 8$ (we take a power of two with the forthcoming application of Lemma 5.5 in mind), we obtain

$$\mathbb{E}_{q_1,\dots,q_8\in S}\Big\| \mathop{*}_{i=1}^{8}\mu_{j,q_i}\Big\|_2^2 \ll \Big(\frac{\log X}{\delta}\Big)^{C_0} X^{-1/2} \tag{7.18}$$

for some C_0 (beyond this point we will cease to use the $\lessapprox$ notation).

We now return to our main line of argument, specifically statement (7.12), and we are now ready to apply the key concatenation result of Kravitz, Kuca and Leng, Lemma 5.5. We apply that result with $s = 3$, with $I = \mathcal{Q}$, and measures given by $\mu_{1,i} = \mu_{1,q}$, $\mu_{1,i} = \mu_{2,q}$, $\mu_{3,i} = \mu_{3,q}$ and $N = O(X^{1/2})$ and with $m = 8$. The conclusion is that there is $t \ll_m 1$ such that

$$X^{-1}\sum_{h,h'\in\mathbb{Z}} \mathbb{E}_{q_1,\dots,q_t\in\mathcal{Q}}\|\Delta_h\Delta_{h'}f_4\|_{U_{\mathrm{GP}}[X^{1/2};\Omega(q_1,\dots,q_t)]} \gg \delta^{O(1)}, \tag{7.19}$$

where $\Omega(q_1,\dots,q_t)$ is the collection of probability measures

$$\Omega(q_1,\dots,q_t) = \Big(\mathop{*}_{i=1}^{8}\mu_{jq_{k_i}}\Big)_{\substack{1\leqslant k_1<k_2<\dots<k_8\leqslant t\\ j\in\{1,2,3\}}}.$$

Set $T := \big(\frac{\log X}{\delta}\big)^{C_1}$ for some very large constant $C_1 > C_0$. We say that a tuple $(q_1,\dots,q_t)\in\mathcal{Q}^{\otimes t}$ is *good* if $\|\mu\|_2^2 \leqslant T/X^{1/2}$ for all $\mu\in\Omega(q_1,\dots,q_t)$. Write $\mathcal{G}$ for the set of all good tuples. By (7.18) and Markov's inequality,

$$\mathbb{P}((q_1,\dots,q_t)\notin\mathcal{G}) \ll \frac{1}{T}\Big(\frac{\log X}{\delta}\Big)^{C_0}.$$

Therefore, from (7.19), if C_1 is large enough we have

$$X^{-1}\sum_{h,h'\in\mathbf{Z}} \mathbb{E}_{q_1,\dots,q_t\in\mathcal{Q}}1_{(q_1,\dots,q_t)\in\mathcal{G}}\|\Delta_h\Delta_{h'}f_4\|_{U_{\mathrm{GP}}[X^{1/2};\Omega(q_1,\dots,q_t)]} \gg \delta^{O(1)}. \tag{7.20}$$

In particular, there is at least one good tuple $(q_1,\dots,q_t)$ such that

$$X^{-1}\sum_{h,h'\in\mathbf{Z}} \|\Delta_h\Delta_{h'}f_4\|_{U_{\mathrm{GP}}[X^{1/2};\Omega(q_1,\dots,q_t)]} \gg \delta^{O(1)}. \tag{7.21}$$

Write $\delta(h,h')$ for the size of the inner norm, thus

$$X^{-1}\sum_{h,h'\in\mathbf{Z}} \delta(h,h') \gg \delta^{O(1)}. \tag{7.22}$$

Note that since $\mathrm{supp}(f_4)\subseteq[\pm X^{1/2}]$, we have that $\delta(h,h') = 0$ if $\max(|h|,|h'|)\geqslant 5X^{1/2}$ (say).

By Corollary 5.3, for each pair (h,h') we have

$$\|\Delta_h\Delta_{h'}f_4\|_{U^{2r}[X^{1/2}]}^{2^{2r}} \gg T^{-O(1)}\delta(h,h')^{O(1)},$$

where $r = 3\binom{t}{8}$ is the number of measures in each $\Omega(q_1,\dots,q_t)$. Averaging over $h,h'\in[\pm 5X^{1/2}]$ and applying Hölder and (7.22), we obtain

$$X^{-1}\sum_{h,h'\in\mathbf{Z}} \|\Delta_h\Delta_{h'}f_4\|_{U^{2r}[X^{1/2}]}^{2^{2r}} \gg (\delta/\log X)^{O(1)}.$$

By expanding out the Gowers norm using (4.11) and (4.12), the above is equivalent to

$$\|f_4\|_{U^{2r+2}[X^{1/2}]}^{2^{2r+2}} \gg (\delta/\log X)^{O(1)}.$$

This concludes the proof of Proposition 4.3.

8. Computing the asymptotic

We turn now to the last main task in the paper, the proof of Proposition 4.5, which asserts an asymptotic for the 'main term'

$$(\log X) \sum_{x,y\in\mathbf{Z}: x^2+ny^2\leqslant X} \chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda_{\text{Cramér}}(x)\Lambda_{\text{Cramér}}(y)1_{x^2+ny^2 \text{ prime}}. \tag{8.1}$$

8.1. **Hecke Größencharaktere on imaginary quadratic fields.** In this section we recall some basic facts about (Hecke) Größencharaktere, as well as key results concerning sums over prime ideals that we will require.

A good and fairly low-level introduction to the notion of Größencharakter on an imaginary quadratic field is [28, Section 3.8]. In the setting of imaginary quadratic fields we can be quite explicit, and the Größencharaktere we are interested in are defined as follows. Let $\mathfrak{m}$ be a non-zero integral ideal, that is to say an ideal in $\mathcal{O}_K$. By a *Dirichlet character* to modulus $\mathfrak{m}$ we mean the lift of a character on $(\mathcal{O}_K/\mathfrak{m})^*$ to $\mathcal{O}_K$, defined to be zero for elements of $\mathcal{O}_K$ which are not coprime to $\mathfrak{m}$ (the definition is analogous to that of a Dirichlet character on $\mathbf{Z}$).

Definition 8.1. Let K be an imaginary quadratic field embedded in $\mathbf{C}$. Let $\mathfrak{m}$ be a non-zero ideal in $\mathcal{O}_K$. Denote by $I_{\mathfrak{m}}$ the group of fractional ideals which are coprime to $\mathfrak{m}$. Let $\ell \in \mathbf{Z}$. Then a Größencharakter modulo $\mathfrak{m}$ with frequency ℓ is a homomorphism $\psi : I_{\mathfrak{m}} \to \mathbf{C}^*$ which satisfies $\psi((\alpha)) = \chi_\infty^{(\ell)}(\alpha)\chi(\alpha)$ for all $\alpha \in K^*$ coprime to $\mathfrak{m}$, where (as in (1.8)) $\chi_\infty^{(\ell)}(\alpha) = (\alpha/|\alpha|)^\ell$, and χ is a Dirichlet character to modulus $\mathfrak{m}$.

Remark. We use the term Größencharakter to emphasise that the frequency ℓ may not be zero. Some authors use the term Hecke character to mean a character of finite order, that is to say with $\ell = 0$. Others such as [28] use the terms Hecke character and (Hecke) Größencharakter interchangably. This can be (and was!) a source of considerable confusion. We will use the term Hecke character only when the frequency ℓ is definitely zero (or equivalently the character has finite order).

Note that if we have a Größencharakter ψ as above, then $\chi_\infty^{(\ell)}, \chi$ satisfy the 'units consistency condition'

$$\chi_\infty^{(\ell)}(u)\chi(u) = 1 \qquad \text{for all } u \in \mathcal{O}_K^*. \tag{8.2}$$

We now show that the converse is also true. Let I be the group of non-zero fractional ideals in K. By a *class group character* we mean a homomorphism $\psi : I \to \mathbf{C}$ which is trivial on principal ideals. Such a homomorphism factors through the class group of K, and so there are exactly h_K such characters.

Lemma 8.2. *Suppose that $\chi_\infty^{(\ell)}, \chi$ as above satisfy the units consistency condition (8.2). Then there is a Größencharakter ψ to modulus $\mathfrak{m}$ satisfying $\psi((\alpha)) = \chi_\infty^{(\ell)}(\alpha)\chi(\alpha)$ for all $\alpha \in K^*$ coprime to $\mathfrak{m}$. Moreover ψ is unique up to multiplication by class group characters.*

Proof. Write $I_{\mathfrak{m}}^0$ for the group of principal fractional ideals coprime to $\mathfrak{m}$. The units consistency condition is precisely what is needed to guarantee that there is a homomorphism $\psi_0 : I_{\mathfrak{m}}^0 \to \mathbf{C}^*$ with $\psi_0((\alpha)) = \chi_\infty^{(\ell)}(\alpha)\chi(\alpha)$ for all $\alpha \in K^*$. The lemma follows from the fact that ψ_0 extends to a homomorphism ψ on $I_{\mathfrak{m}}$. This is a special case of a general algebraic fact: if $G_0 \leqslant G$ are abelian groups, and Z is a divisible abelian group, then any homomorphism $\psi_0 : G_0 \to Z$ can be extended to a homomorphism $\psi : G \to Z$. In the case that $[G : G_0] < \infty$ (as in our situation) one may prove this fact in finitely many stages, first extending ψ_0 to a homomorphism on $G_1 := \langle G_0, x\rangle$ by defining $\psi_1(g_0x^r) := \psi_0(g_0)z^r$, where $z \in Z$ is any element with $z^d = \psi_0(x^d)$, where d is the order of x in G_1/G_0, and then proceeding similarly in further stages. (We leave it for the reader to check that ψ_1 is well-defined. A Zorn's lemma argument would handle the general case.)

It is clear that ψ is unique up to multiplication by characters $\xi : I_{\mathfrak{m}} \to \mathbf{C}^*$ which vanish on $I_{\mathfrak{m}}^0$. We claim that any such ξ is (the restriction of) a class group character. To see this, consider the natural injective homomorphism $\xi : I_{\mathfrak{m}}/I_{\mathfrak{m}}^0 \to I/I^0$ induced by the inclusion $I_{\mathfrak{m}} \hookrightarrow I$, where I^0 denotes the principal fractional ideals in K. It is a well-known fact that every ideal class in K contains infinitely many prime ideals (this can be proven using the prime ideal theorem for class group characters, which is a special case of Proposition 8.3 below) and in particular at least one ideal coprime to $\mathfrak{m}$. Therefore ξ is surjective and is hence an isomorphism. The claim follows. □

In this section most Größencharaktere will have the same frequency ℓ, which is the one in the statement of Proposition 4.5. Recall that we are making the assumption that $|\ell| \leqslant e^{\sqrt{\log X}}$.

We define the von Mangoldt function Λ_K on ideals by

$$\Lambda_K(\mathfrak{a}) = \begin{cases} \log N\mathfrak{p} & \text{if } \mathfrak{a} = \mathfrak{p}^k \text{ for some } k \geqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 8.3 below, a form of the prime ideal theorem for Größencharaktere, is the main input we require from classical multiplicative number theory. Before stating the result, we recall the notion of one Größencharakter being induced from another. If ψ is a Größencharakter to modulus $\mathfrak{m}$, and if $\tilde{\mathfrak{m}}$ is another ideal with $\mathfrak{m} \mid \tilde{\mathfrak{m}}$, then ψ induces a Größencharakter $\tilde{\psi}$ to modulus $\tilde{\mathfrak{m}}$ by setting $\tilde{\psi}(\mathfrak{a}) = \psi(\mathfrak{a})$ if $\mathfrak{a}$ and $\tilde{\mathfrak{m}}$ are coprime, and $\tilde{\psi}(\mathfrak{a}) = 0$ otherwise. The character which takes the value 1 for all nonzero prime ideals is called the *principal* character (to modulus $(1) = \mathcal{O}_K$). It induces a character to any modulus $\mathfrak{m}$, which we call the principal character to that modulus. A Größencharakter is called *primitive* if it is not induced from any Größencharakter of strictly smaller modulus.

Proposition 8.3. *Let $K = \mathbf{Q}(\sqrt{-n})$ and let $X > 1$. Then there is at most one 'exceptional' primitive Größencharakter $\chi_{\mathfrak{d}^*}$ to some modulus $\mathfrak{d}^*$ and associated real number $\beta \in (\frac{1}{2}, 1)$ such that the following is true uniformly for all Größencharaktere ψ on K whose modulus $\mathfrak{m}$ and frequency ℓ satisfy $|N\mathfrak{m}|, |\ell| \leqslant e^{\sqrt{\log X}}$:*

$$\sum_{N\mathfrak{a} \leqslant X} \Lambda_K(\mathfrak{a})\psi(\mathfrak{a}) = X 1_{\psi=\psi_0,\ell=0} - \frac{X^\beta}{\beta} 1_{\psi\sim\chi_{\mathfrak{d}^*},\ell=0} + O(Xe^{-c\sqrt{\log X}}).$$

Here ψ_0 is the principal character to modulus $\mathfrak{m}$. The exceptional charakter $\chi_{\mathfrak{d}^*}$, if it exists, is quadratic (that is to say $\chi_{\mathfrak{d}^*}^2$ is the principal charakter to modulus $\mathfrak{d}^*$). The notation $\psi \sim \chi_{\mathfrak{d}^*}$ means that ψ is induced from $\chi_{\mathfrak{d}^*}$, which in particular requires that $\mathfrak{d}^* \mid \mathfrak{m}$. Finally, β satisfies

$$1 - \beta \gg_\varepsilon (N\mathfrak{d}^*)^{-\varepsilon}. \tag{8.3}$$

Remark. The exceptional character depends on the scale X (which is fixed throughout the paper). The number β is a root of $L(s, \chi_{\mathfrak{d}^*})$. If no exceptional character exists (which is conjectured to be the case) then the term with $-X^\beta/\beta$ should be omitted.

Sketch. This result is well-known to experts and versions of it have been used in several places; see, for example, [37, Lemma 4.5] for an essentially identical statement. However it is very difficult to locate a proof in the literature. Ultimately, it can be demonstrated by applying [28, Theorem 5.13] to appropriate zero-free regions for Größencharaktere, all of which may be found either explicitly or with minor modifications (but sometimes with only skeleton proofs) across papers of Mitsui [40] and Fogels [5–7] from the late 1950s and early 1960s. The arguments are analogous to the proofs over $\mathbf{Q}$ (with Dirichlet characters) which are covered in complete detail in [3] or [42]. Due to the somewhat unsatisfactory nature of the literature here, the first-named author is preparing an expository document [15] with a complete proof. □

8.2. **Decomposing $\Lambda_{\text{Cramér}}$.** We now begin the proof of Proposition 4.5. The first step is to effect a decomposition of $\Lambda_{\text{Cramér}}$ into two parts. We will show that, of the resulting four terms contributing to (8.1), all but one are small. The evaluation of the remaining term is a more substantial task, and is deferred to subsequent sections.

Write $\mathcal{P}_Q$ for the set of primes $\leqslant Q$, and for $S \subseteq \mathcal{P}_Q$, write $P_S := \prod_{p\in S} p$.

Then by inclusion–exclusion we have

$$\Lambda_{\text{Cramér}}(x) = \prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-1}\sum_{S\subseteq\mathcal{P}_Q}(-1)^{|S|}1_{P_S|x}.$$

Let

$$t := 20\log\log X, \tag{8.4}$$

and split

$$\Lambda_{\text{Cramér}} = \Lambda^{\sharp}_{\text{Cramér}} + \Lambda^{\flat}_{\text{Cramér}} \tag{8.5}$$

where

$$\Lambda^{\sharp}_{\text{Cramér}} := \prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-1}\sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|\leqslant t}}(-1)^{|S|}1_{P_S|x} \tag{8.6}$$

and

$$\Lambda^{\flat}_{\text{Cramér}} := \prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-1}\sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|> t}}(-1)^{|S|}1_{P_S|x}. \tag{8.7}$$

The following lemma will be used to guarantee that the contribution of the $\Lambda^{\flat}_{\text{Cramér}}$ terms to (8.1) will be small.

Lemma 8.4. *We have*

$$\sum_{x\leqslant X^{1/2}}|\Lambda^{\flat}_{\text{Cramér}}(x)| \ll X^{1/2}(\log X)^{-8}. \tag{8.8}$$

and

$$\sum_{x\leqslant X^{1/2}}|\Lambda^{\sharp}_{\text{Cramér}}(x)| \ll X^{1/2}(\log X)^{2}. \tag{8.9}$$

Proof. We begin with (8.8). First note that $\prod_{p\leqslant Q}(1-\frac{1}{p})^{-1} \ll \log Q < \log X$. Therefore we have the pointwise bounds

$$\Lambda_{\text{Cramér}}(x) \ll \log X \tag{8.10}$$

and

$$|\Lambda^{\flat}_{\text{Cramér}}(x)| \ll (\log X)\sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|>t}}1_{P_S|x} \ll (\log X)\tau(x). \tag{8.11}$$

Moreover, $\Lambda^{\flat}_{\text{Cramér}}$ is supported on x with at least t distinct prime factors, and thus on x for which $|\tau(x)| \geqslant 2^t$. Thus, on the support of $\Lambda^{\flat}_{\text{Cramér}}$, we have $\tau(x) \leqslant 2^{-t}\tau(x)^2 < (\log X)^{-12}\tau(x)^2$. Substituting into (8.11) and summing gives

$$\sum_{x\leqslant X^{1/2}}|\Lambda^{\flat}_{\text{Cramér}}(x)| \ll (\log X)^{-11}\sum_{x\leqslant X^{1/2}}\tau(x)^2 \ll X^{1/2}(\log X)^{-8},$$

where in the last step we used Lemma D.1 in the case $m = 2$. This is (8.8).

Turning to (8.9), observe the bound

$$|\Lambda^{\sharp}_{\text{Cramér}}(x)| \ll (\log X)\tau(x), \tag{8.12}$$

which follows from (8.5), (8.10), and (8.11) and the triangle inequality. The desired bound (8.9) follows by summing over x. □

Substitute the decomposition (8.5) into (8.1). This gives a sum of four terms. Three of these terms are

$$(\log X) \sum_{x,y\in\mathbf{Z}:x^2+ny^2\leqslant X} \chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda'_{\text{Cramér}}(x)\Lambda''_{\text{Cramér}}(y)1_{x^2+ny^2\text{ prime}}$$

where $\Lambda'_{\text{Cramér}}, \Lambda''_{\text{Cramér}} \in \{\Lambda^\sharp_{\text{Cramér}}, \Lambda^\flat_{\text{Cramér}}\}$, and at least one of these functions is $\Lambda^\flat_{\text{Cramér}}$. Using Lemma 8.4, each such term can be bounded crudely by $X(\log X)^{-5}$, here ignoring the constraint that x^2+ny^2 is prime entirely. The remaining term involves two copies of $\Lambda^\sharp_{\text{Cramér}}$. We have therefore reduced the task of proving Proposition 4.5 to the task of showing the following asymptotic.

Proposition 8.5. *Suppose that n is even. Then we have*

$$(\log X) \sum_{\substack{x,y\in\mathbf{Z}\\ x^2+ny^2\leqslant X}} \chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)1_{x^2+ny^2\text{ prime}} = \frac{\kappa_n 1_{\ell=0}X}{\sqrt{n}} + O(X(\log X)^{-1}),$$

where κ_n is as given in (1.1).

This will occupy us for the rest of the section. In particular, we assume that n is even from now on.

8.3. **Writing as a sum over ideals.** We will ultimately evaluate the LHS of Proposition 8.5 by reference to the main analytic input of the paper, namely Proposition 8.3. To prepare for this application, we must rewrite this expression as an appropriate combination of sums $\sum_{\mathfrak{a}:N\mathfrak{a}\leqslant X}\Lambda_K(\mathfrak{a})\psi(\mathfrak{a})$.

We begin by expressing the LHS of Proposition 8.5 as a sum over ideals with a von Mangoldt weight. We claim that

$$\begin{aligned}(\log X) \sum_{x^2+ny^2\leqslant X} &\chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)1_{x^2+ny^2\text{ prime}}\\ &= \sum_{\mathfrak{a}:N\mathfrak{a}\leqslant X} \Lambda^\sharp_{\text{Cramér}} \boxtimes_\ell \Lambda^\sharp_{\text{Cramér}}(\mathfrak{a})\Lambda_K(\mathfrak{a}) + O\Big(\frac{X}{\log X}\Big),\end{aligned} \tag{8.13}$$

where the $\boxtimes_\ell$ notation is defined in Definition 2.2.

To see (8.13), note that the sum on the right counts x, y (weighted by $\Lambda^\sharp_{\text{Cramér}}$) for which $\mathfrak{a} = (x+y\sqrt{-n})$ is a prime power, weighted by $\Lambda_K(\mathfrak{a})$. For all of these, $x+y\sqrt{-n}$ is power of a rational prime, and the contribution from prime powers can be discarded. We have $\Lambda_K(\mathfrak{a})-\log X = \log(\frac{N\mathfrak{a}}{X})$, and so (8.13) is a consequence of the following lemma.

Lemma 8.6. *We have*

$$\sum_{x,y\in\mathbf{Z}:x^2+ny^2\leqslant X} \Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)\log\Big(\frac{x^2+ny^2}{X}\Big) \ll X.$$

Proof. We will use partial summation. We first claim that, uniformly for $X^{1/2}\leqslant Y\leqslant X$,

$$\sum_{x,y\in\mathbf{Z}:x^2+ny^2\leqslant Y} \Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y) = \frac{\pi}{\sqrt{n}}Y + O(X(\log X)^{-10}). \tag{8.14}$$

The proof of this claim is a nice warmup for the more complicated computations later on in the section. Opening up the definition (8.6) of $\Lambda^\sharp_{\text{Cramér}}$, the left-hand side is

$$\prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}}(-1)^{|S_1|+|S_2|}\sum_{x^2+ny^2\leqslant Y}1_{P_{S_1}|x,P_{S_2}|y}. \tag{8.15}$$

For S_1,S_2 appearing in the above sum, we have $P_{S_1},P_{S_2}\leqslant Q^t\ll\exp(\log^{1/8}X)$. Therefore, by an easy lattice point counting argument we have

$$\sum_{x^2+ny^2\leqslant Y}1_{P_{S_1}|x,P_{S_2}|y}=\frac{\pi}{\sqrt{n}}\frac{Y}{P_{S_1}P_{S_2}}+O(Y^{9/10}).$$

Substituting into (8.15) gives that the LHS of (8.14) is

$$\frac{\pi Y}{\sqrt{n}}\prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}}(-1)^{|S_1|+|S_2|}\frac{1}{P_{S_1}P_{S_2}}+O\Big(Y^{9/10}(\log Q)^2\binom{|\mathcal{P}_Q|}{t}^2\Big). \tag{8.16}$$

Since $\binom{|\mathcal{P}_Q|}{t}<Q^t<\exp(\log^{1/8}X)$, the error term here can be comfortably absorbed into the one claimed in (8.14). Now we remove the conditions $|S_1|,|S_2|\leqslant t$ in the sum. To bound the error in doing this, note that

$$\begin{aligned}\sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|>t}}\prod_{p\in S}\frac{1}{p-1}&\leqslant\sum_{s\geqslant t}\frac{1}{s!}\Big(\sum_{p\in\mathcal{P}_Q}\frac{1}{p-1}\Big)^s\leqslant\sum_{s\geqslant t}\frac{1}{s!}(\log\log Q+O(1))^s\\ &\leqslant\sum_{s\geqslant t}\Big(\frac{3\log\log Q}{s}\Big)^s<2\Big(\frac{3\log\log Q}{t}\Big)^t<\Big(\frac{1}{6}\Big)^{20\log\log X}<(\log X)^{-20},\end{aligned} \tag{8.17}$$

recalling here the choice (8.4) of t. (Actually, here we only need a bound with p rather than $p-1$ in the denominators but we will reuse these inequalities later on.) It follows that

$$\prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|\geqslant t}}(-1)^{|S_1|+|S_2|}\frac{1}{P_{S_1}P_{S_2}}\ll(\log Q)^2\prod_{p\leqslant Q}\Big(1+\frac{1}{p}\Big)\cdot\sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|>t}}\prod_{p\in S}\frac{1}{p}\ll(\log X)^{-10}.$$

Thus the error from dropping the conditions $|S_1|,|S_2|\leqslant t$ in (8.16) can be absorbed into the error term in (8.14). Finally, (8.14) itself follows form the observation that

$$\prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-1}\sum_{S\subseteq\mathcal{P}_Q}(-1)^{|S|}\frac{1}{P_S}=1.$$

It remains to actually carry out the partial summation and thereby establish the lemma. For this, we use

$$\sum_{x,y\in\mathbf{Z}:x^2+ny^2\leqslant X}\Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)\log\Big(\frac{x^2+ny^2}{X}\Big)=-\int_1^X\frac{dt}{t}\sum_{x^2+ny^2\leqslant t}\Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)$$

The contribution from the range $t<X^{1/2}$ may be bounded trivially via using the case $Y=X^{1/2}$ of (8.14). For the remaining integral we use (8.14) again, obtaining a contribution of

$$-\frac{\pi}{\sqrt{n}}(X-X^{1/2})+O(X(\log X)^{-9})$$

from this range. The result follows. □

Remark. The techniques in use here are those of Brun's sieve, as discussed for instance in [12, Chapter 6] or [29, Chapter 17].

8.4. **Expanding with Größencharaktere.** We have now established (8.13). To go further, we work with the sum over ideals on the RHS of this equation, with the aim being to write it in terms of sums $\sum_{\mathfrak{a}} \Lambda_K(\mathfrak{a})\psi(\mathfrak{a})$ for Größencharaktere ψ. Opening up the definitions of the two copies of $\Lambda^\sharp_{\text{Cramér}}$ in (8.13) (recalling the definition (8.6)), we may write

$$\sum_{\mathfrak{a}: N\mathfrak{a}\leqslant X} \Lambda^\sharp_{\text{Cramér}} \boxtimes_\ell \Lambda^\sharp_{\text{Cramér}}(\mathfrak{a})\Lambda_K(\mathfrak{a}) = \prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2} \sum_{\substack{S_1,S_2\subseteq \mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}} (-1)^{|S_1|+|S_2|} E(S_1,S_2), \tag{8.18}$$

where

$$E(S_1,S_2) := \sum_{\mathfrak{a}: N\mathfrak{a}\leqslant X} \varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a})\Lambda_K(\mathfrak{a}), \tag{8.19}$$

where

$$\varphi_{P_S}(x) := 1_{x\in\mathbf{Z}, P_S|x}.$$

To understand this better, write

$$\Gamma(S_1,S_2) := \{\gamma\in\mathcal{O}_K : \gamma = x+y\sqrt{-n} \text{ with } x,y\in\mathbf{Z}, P_{S_1}\mid x, P_{S_2}\mid y\}, \tag{8.20}$$

and observe that if $\mathfrak{a}=(\alpha)$ is principal then

$$\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a}) = \sum_{u\in\mathcal{O}_K^*} \chi_\infty^{(\ell)}(u\alpha) 1_{\Gamma(S_1,S_2)}(u\alpha). \tag{8.21}$$

The definition does not depend on the choice of α; also, by definition of $\boxtimes_\ell$, $\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a}) = 0$ if $\mathfrak{a}$ is not principal.

Let us also take the opportunity to remark that, for any P_S appearing in a term $E(S_1,S_2)$ involved in the sum (8.18), we have $P_S \leqslant Q^t < \exp(\log^{1/8} X)$.

We now expand the function $\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}$ using Größencharaktere. First we note that (see Definition 2.2) it is supported on principal ideals, a condition we detect using a sum over class group characters:

$$1_{[\mathfrak{a}]=1_{\mathrm{Cl}(K)}}(\mathfrak{a}) = h_K^{-1} \sum_{\chi_1\in\widehat{\mathrm{Cl}(K)}} \chi_1(\mathfrak{a}). \tag{8.22}$$

Define $\mathfrak{m} := (D)$ where $D := 2nP_{S_1}P_{S_2}$. We claim that $\Gamma(S_1,S_2)$ is a union of congruence classes mod $\mathfrak{m}$. To see this, suppose that $\gamma\in\Gamma(S_1,S_2)$ and that $\gamma'\in\gamma+\mathfrak{m}$. Then $\gamma' = \gamma + D\beta$ for some $\beta\in\mathcal{O}_K$. Recalling (from (2.1)) that $2n\mathcal{O}_K\subseteq\mathbf{Z}[\sqrt{-n}]$, one sees from the choice of D that $D\beta\in P_{S_1}P_{S_2}\mathbf{Z}[\sqrt{-n}]$, and the claim follows.

We will only care about $1_{\Gamma(S_1,S_2)}(\gamma)$ at values of γ which are coprime to $\mathfrak{m}$. Restricted to such values we have an expansion

$$1_{\gamma\in\Gamma(S_1,S_2)} = \sum_{\chi (\mathrm{mod}\,\mathfrak{m})} c_\chi^{S_1,S_2}\chi(\gamma) \tag{8.23}$$

in Dirichlet characters χ to modulus $\mathfrak{m}$ for some coefficients $c_\chi^{S_1,S_2}$.

Substituting into (8.21) gives

$$\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a}) = \sum_{\chi (\mathrm{mod}\,\mathfrak{m})} c_\chi^{S_1,S_2} \sum_{u\in\mathcal{O}_K^*} \chi_\infty^{(\ell)}(u\alpha)\chi(u\alpha).$$

Note that the inner sum vanishes unless $\chi_\infty^{(\ell)}, \chi$ satisfy the units consistency condition described at the start of Section 8.1, since $\chi_\infty^{(\ell)}, \chi$ are both characters and

$$\sum_u \chi_\infty^{(\ell)}(u)\chi(u) = \sum_{j=0}^{|\mathcal{O}_K^*|-1} (\chi_\infty^{(\ell)}(u_0)\chi(u_0))^j,$$

where u_0 is a generator for the group of units. If the units consistency condition is satisfied then it follows from Lemma 8.2 that $\chi_\infty^{(\ell)}\chi$ is the restriction to principal ideals of some Größencharakter ψ to modulus $\mathfrak{m}$, with frequency ℓ. Therefore, for principal ideals $\mathfrak{a}$, coprime to $\mathfrak{m}$, we have

$$\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a}) = |\mathcal{O}_K^*| \sum_\psi c_\psi^{S_1,S_2}\psi(\mathfrak{a}), \tag{8.24}$$

where ψ ranges over a complete set of Größencharaktere (mod $\mathfrak{m}$) with frequency ℓ modulo equivalence by class group characters, and we have abused notation by writing $c_\psi^{S_1,S_2} = c_\chi^{S_1,S_2}$. Finally, combining with (8.22) we obtain an expression valid for *all* ideals coprime to $\mathfrak{m}$ (not just principal ones) namely

$$\varphi_{P_{S_1}} \boxtimes_\ell \varphi_{P_{S_2}}(\mathfrak{a}) = |\mathcal{O}_K^*|h_K^{-1} \sum_\psi c_\psi^{S_1,S_2}\psi(\mathfrak{a}), \tag{8.25}$$

for all $\mathfrak{a}$ coprime to $\mathfrak{m}$, where now the sum is over all Größencharaktere to modulus $\mathfrak{m}$ with frequency ℓ, and again we have abused notation by writing $c_\psi^{S_1,S_2}$ for $c_\chi^{S_1,S_2}$, where χ is the Dirichlet character associated to ψ.

For use later on note that we have the bound

$$|c_\psi^{S_1,S_2}| \leqslant c_{\psi_0}^{S_1,S_2} \leqslant 1, \tag{8.26}$$

where ψ_0 is the principal character; this follows from the proof of (8.23), $c_\chi^{S_1,S_2}$ being the average of $1_{\gamma\in\Gamma(S_1,S_2)}\overline{\chi(\gamma)}$ over $\gamma \in (\mathcal{O}_K/\mathfrak{m})^*$. The right-hand bound in (8.26) is crude and will be improved later.

8.5. **Contribution of the principal character.** From this point on we will use the quantities n_* (squarefree part of n), r (defined to be $\sqrt{n/n_*}$) and $\omega \in \{\frac{1}{2}, 1\}$ (depending on the value of n_* mod 4, see (2.4)).

We now turn to the key computation which will lead to the constant in our asymptotic formula, which is that of the coefficient of the principal character ψ_0 to modulus $\mathfrak{m}$, namely

$$\sigma(S_1, S_2) := |\mathcal{O}_K^*|h_K^{-1}c_{\psi_0}^{S_1,S_2}. \tag{8.27}$$

Note that this section is only relevant in the case $\ell = 0$, since if $\ell \neq 0$ then the sum (8.25) does not contain the principal character. Assume, then, that $\ell = 0$. The idea is to sum (8.23) for γ in a set of representatives for $(\mathcal{O}_K/\mathfrak{m})^*$. By orthogonality of Dirichlet characters, this implies that

$$\#\{\gamma \in \Gamma(S_1, S_2)/\mathfrak{m},\ \gcd(\gamma, \mathfrak{m}) = 1\} = |(\mathcal{O}_K/\mathfrak{m})^*|c_{\psi_0}^{S_1,S_2}. \tag{8.28}$$

We have (recalling the definition (8.20) of $\Gamma(S_1, S_2)$)

$$\Gamma(S_1, S_2) = \{\alpha = x + ry\sqrt{-n_*} : x, y \in \mathbf{Z}, P_{S_1} \mid x, P_{S_2} \mid y\}. \tag{8.29}$$

Using this we can almost immediately observe that, in certain cases, $\sigma(S_1, S_2)$ is zero. In the following lemma, and for the rest of the section, set

$$T := \{p \leqslant Q : p \mid n\}.$$

Lemma 8.7. *Suppose that $\sigma(S_1, S_2) \neq 0$. Then $S_1 \cap S_2 = S_1 \cap T = \emptyset$.*

Proof. If $p \in S_1 \cap S_2$ then clearly p divides both $\mathfrak{m}$ and every element of $\Gamma(S_1, S_2)$, so in this case we have $\sigma(S_1, S_2) = 0$ by (8.27) and (8.28).

Alternatively suppose $p \in S_1 \cap T$. We have either (i) $p \mid r$ or (ii) $p \mid n_*$ (or both). In case (i), p divides both $\mathfrak{m}$ and every element of $\Gamma(S_1, S_2)$, and we conclude as before. In case (ii), p ramifies (since n_* divides the discriminant Δ; see Section 2), thus $p = \mathfrak{p}^2$ for some prime ideal $\mathfrak{p}$. We must then have $\mathfrak{p} \mid \sqrt{-n_*}$. Since $\mathfrak{p} \mid p \mid P_{S_1}$, we see that $\mathfrak{p}$ divides both $\mathfrak{m}$ and every element of $\Gamma(S_1, S_2)$, and we conclude as before. □

Assume henceforth that $S_1 \cap S_2 = S_1 \cap T = \emptyset$. To progress in this case, we must evaluate the two sides of (8.28). Looking first at the right-hand side, we have

$$|(\mathcal{O}_K/\mathfrak{m})^*| = (N\mathfrak{m}) \prod_{\mathfrak{p}|\mathfrak{m}} \Big(1 - \frac{1}{N\mathfrak{p}}\Big) = (N\mathfrak{m}) \prod_{p|\mathfrak{m}} \Big(1 - \frac{1}{p}\Big)\Big(1 - \frac{(\Delta \mid p)}{p}\Big), \tag{8.30}$$

where $\mathfrak{p}$ ranges over the distinct prime ideals dividing $\mathfrak{m}$, and p ranges over the distinct rational primes dividing $\mathfrak{m}$. (To see the second expression, consider the possible splitting types of p, with reference to the remarks in Section 2.)

Using the fact that $\mathcal{O}_K = \mathbf{Z}[\sqrt{-n_*}]$ in the case $\omega = 1$ and $\mathbf{Z}[\frac{1}{2}(1 + \sqrt{-n_*})]$ in the case $\omega = \frac{1}{2}$, one may check that a complete set of representatives for $\Gamma(S_1, S_2)$ mod $\mathfrak{m}$ is

$$\Gamma(S_1, S_2) = \{\alpha = x + ry\sqrt{-n_*} : 0 \leqslant x < 2nP_{S_1}P_{S_2},\ 0 \leqslant y < 2\omega nP_{S_1}P_{S_2}/r,\ P_{S_1} \mid x,\ P_{S_2} \mid y\},$$

or equivalently

$$\Gamma(S_1, S_2) = \{\alpha = aP_{S_1} + brP_{S_2}\sqrt{-n_*} : 0 \leqslant a < 2nP_{S_2},\ 0 \leqslant b < 2\omega nP_{S_1}/r\}. \tag{8.31}$$

Lemma 8.8. *Suppose that $S_1 \cap S_2 = S_1 \cap T = \emptyset$. Then the elements of (8.31) having a common factor with $\mathfrak{m}$ are precisely:*

(1) Those with $p \mid a$ for some $p \in S_2 \cup T$;
(2) Those with $p \mid b$ for some $p \in S_1$.

Remark. Note that, in our main theorem, n *is* even.

Proof. The set of primes factors of $\mathfrak{m}$ is $S_1 \cup S_2 \cup T$ (here we use the assumption that n is even in Proposition 8.5).

Suppose that $\mathfrak{p}$ is a prime ideal common factor of $\alpha = aP_{S_1} + brP_{S_2}\sqrt{-n_*}$ and $\mathfrak{m}$. Let p be the rational prime above $\mathfrak{p}$. Then we have exactly one of (i) $p \in S_2 \cup T$ or (ii) $p \in S_1$.

In case (i), we break in further cases based on whether $p \in S_2$, $p \mid n_*$, or $p \mid r$. Suppose first that $p \in S_2$. Then $\mathfrak{p} \mid p \mid P_{S_2}$, whilst $\mathfrak{p}$ is coprime to P_{S_1} since S_1, S_2 are disjoint. Thus the condition that $\mathfrak{p} \mid \alpha$ is then precisely that $\mathfrak{p} \mid a$, or equivalently $p \mid a$. Next suppose (still in case (i)) that $p \mid n_*$. Then p is ramified and $p = \mathfrak{p}^2$, and $\mathfrak{p} \mid \sqrt{-n_*}$. Thus $\mathfrak{p} \mid aP_{S_1}$ which, since we are assuming $p \notin S_1$, means that $\mathfrak{p} \mid a$ and hence $p \mid a$ is again the condition. Finally suppose (still in case (i)) that $p \mid r$, so $\mathfrak{p} \mid r$. Then, since $\mathfrak{p} \nmid P_{S_1}$, we immediately see that the condition is again $p \mid a$.

In case (ii), since $p \in S_1$ we have $\mathfrak{p} \mid P_{S_1}$ and so $\mathfrak{p} \mid brP_{S_2}\sqrt{-n_*}$. Now $\mathfrak{p}$ is coprime to r (since $S_1 \cap T = \emptyset$), to P_{S_2} (since $S_1 \cap S_2 = \emptyset$) and to $\sqrt{-n_*}$ (since otherwise $p \mid n$, as explained above, but this contradicts $S_1 \cap T = \emptyset$). Therefore $p \mid b$, as claimed.

All the above implications are reversible, as may easily be checked. □

As a consequence of this lemma and (8.31) it follows that

$$\#\{\gamma \in \Gamma(S_1, S_2)/\mathfrak{m},\ \gcd(\gamma, \mathfrak{m}) = 1\} = \frac{4\omega n^2 P_{S_1}P_{S_2}}{r} \prod_{p|\mathfrak{m}} \Big(1 - \frac{1}{p}\Big) = \frac{\omega N\mathfrak{m}}{rP_{S_1}P_{S_2}} \prod_{p|\mathfrak{m}} \Big(1 - \frac{1}{p}\Big). \tag{8.32}$$

Comparing (8.27), (8.28), (8.30), and (8.32) gives

$$\sigma(S_1,S_2)=\frac{\omega|\mathcal{O}_K^*|}{rh_KP_{S_1}P_{S_2}}\prod_{p|\mathfrak{m}}\Big(1-\frac{(\Delta\mid p)}{p}\Big)^{-1}. \tag{8.33}$$

It is convenient to write this as

$$\sigma(S_1,S_2)=\frac{\omega|\mathcal{O}_K^*|P_{T\setminus S_2}}{rh_K}\prod_{p\in T\cup S_1\cup S_2}\nu(p), \tag{8.34}$$

where

$$\nu(p):=(p-(\Delta\mid p))^{-1}, \tag{8.35}$$

and recall that this is under the assumption that $S_1\cap S_2=S_1\cap T=\emptyset$ (otherwise Lemma 8.7 applies).

8.6. **Applying the prime ideal theorem.** Now we return to our main task of proving Proposition 8.5. Let us briefly recap the progress made so far. We proved (8.13) and (8.18), which combine together to give

$$\begin{aligned}(\log X)\sum_{x^2+ny^2\leqslant X}\chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda^\sharp_{\text{Cramér}}(x)\Lambda^\sharp_{\text{Cramér}}(y)1_{x^2+ny^2\text{ prime}}\\ =\prod_{p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}}(-1)^{|S_1|+|S_2|}E(S_1,S_2)+O\Big(\frac{X}{\log X}\Big),\end{aligned} \tag{8.36}$$

where

$$E(S_1,S_2):=\sum_{\mathfrak{a}:N\mathfrak{a}\leqslant X}\varphi_{P_{S_1}}\boxtimes_\ell\varphi_{P_{S_2}}(\mathfrak{a})\Lambda_K(\mathfrak{a}), \tag{8.37}$$

(see around (8.19) for the details). Moreover, in (8.25) we obtained an expansion for $\varphi_{P_{S_1}}\boxtimes_\ell\varphi_{P_{S_2}}(\mathfrak{a})$ in characters, in the case that $\mathfrak{a}$ is coprime to $\mathfrak{m}=\mathfrak{m}^{S_1,S_2}=(2nP_{S_1}P_{S_2})$. Combining this with (8.37) gives

$$E(S_1,S_2)=\frac{|\mathcal{O}_K^*|}{h_K}\sum_\psi c_\psi^{S_1,S_2}\sum_{\substack{\mathfrak{a}:N\mathfrak{a}\leqslant X\\ \gcd(\mathfrak{a},\mathfrak{m}^{S_1,S_2})=1}}\psi(\mathfrak{a})\Lambda_K(\mathfrak{a})+O(\log X),$$

where the sum is over Größencharaktere ψ to modulus $\mathfrak{m}$ with frequency ℓ, and where the $O(\log X)$ error term comes from those $\mathfrak{a}$ which are not coprime to $\mathfrak{m}^{S_1,S_2}$, the contribution of these being

$$\ll\sum_{\mathfrak{p}^j|\mathfrak{m}^{S_1,S_2}}\log N\mathfrak{p}\ll\log X.$$

In the following $c_1,c_2,c_3>0$ are absolute constants. Applying Proposition 8.3 gives

$$E(S_1,S_2)=\frac{|\mathcal{O}_K^*|}{h_K}\Big(Xc_{\psi_0}^{S_1,S_2}-\frac{X^\beta}{\beta}c_{\psi^*}^{S_1,S_2}\Big)+O\Big(Xe^{-c_1\sqrt{\log X}}\Big(1+\sum_\psi|c_\psi^{S_1,S_2}|\Big)\Big), \tag{8.38}$$

where here all Größencharaktere ψ are to modulus $\mathfrak{m}=\mathfrak{m}^{S_1,S_2}$ and have frequency ℓ, ψ_0 is the principal character to modulus $\mathfrak{m}$ (which is only present if $\ell=0$) and the 'Siegel term' (the one involving X^β/β) is only present if $\ell=0$ and $\mathfrak{d}^*\mid\mathfrak{m}$, which is the condition for there to be some Größencharakter ψ_* to modulus $\mathfrak{m}$ with frequency ℓ which is induced from the exceptional charakter $\chi_{\mathfrak{d}^*}$.

In particular, the whole main term here is absent unless $\ell=0$.

Recall (see (8.27)) that we gave the name $\sigma(S_1,S_2)$ to the coefficient $|\mathcal{O}_K^*|h_K^{-1}c_{\psi_0}^{S_1,S_2}$ appearing here, and derived expressions for it (Lemma 8.7 and (8.34)). Thus, from (8.38) (and bounding the error term there fairly trivially using (8.26) and the bound $N\mathfrak{m} = (2nP_{S_1}P_{S_2})^2 \ll \exp(O(\log^{1/8} X))$), we derive from (8.38) that

$$E(S_1,S_2) = X\sigma(S_1,S_2) - \frac{|\mathcal{O}_K^*|}{h_K}\frac{X^\beta}{\beta}c_{\psi^*}^{S_1,S_2} + O\big(Xe^{-c_2\sqrt{\log X}}\big). \tag{8.39}$$

Now it is time to substitute this into the formula (8.36) for the expression of interest in Proposition 8.5. Doing this, we obtain

$$(\log X)\sum_{x^2+ny^2\leqslant X}\chi_\infty^{(\ell)}(x+y\sqrt{-n})\Lambda_{\text{Cramér}}^\sharp(x)\Lambda_{\text{Cramér}}^\sharp(y)1_{x^2+ny^2\text{prime}}$$
$$= E_{\text{main}} + E_{\text{Siegel}} + E_{\text{error}} + O\Big(\frac{X}{\log X}\Big), \tag{8.40}$$

where

$$E_{\text{main}} := X\prod_{p\in\mathcal{P}_Q}\Big(1-\frac{1}{p}\Big)^{-2}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}}(-1)^{|S_1|+|S_2|}\sigma(S_1,S_2), \tag{8.41}$$

$$E_{\text{Siegel}} \ll (\log Q)^2X^\beta\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_1|,|S_2|\leqslant t}}\sum_{\substack{\psi^*\pmod{\mathfrak{m}^{S_1,S_2}}\\ \psi_*\sim\chi_{\mathfrak{d}^*}}}|c_{\psi^*}^{S_1,S_2}|, \tag{8.42}$$

and

$$E_{\text{error}} \ll X(\log Q)^2\binom{|\mathcal{P}_Q|}{t}^2e^{-c_2\sqrt{\log X}} \ll Xe^{-c_3\sqrt{\log X}} \tag{8.43}$$

can be absorbed into the error term in (8.40). This in fact completes the proof of Proposition 8.5 in the case $\ell\neq 0$, since in this case both E_{main} and E_{Siegel} are absent.

Suppose from now on that $\ell = 0$. In the remaining sections, we evaluate E_{main} asymptotically, and show that E_{Siegel} is small; this will then conclude the proof of Proposition 8.5 in all cases.

8.7. **Estimating the Siegel term.** In this subsection we estimate the Siegel term E_{Siegel}. Write $d_* := N\mathfrak{d}^*$. First note that a pair S_1,S_2 of subsets of $\mathcal{P}_Q$ only contributes to the sum if $d_* \mid 4n^2P_{S_1}P_{S_2} = N\mathfrak{m}^{S_1,S_2}$, since $\mathfrak{d}^*\mid\mathfrak{m}$ for any exceptional modulus $\mathfrak{m}$.

To bound the contribution of these terms, we first upgrade the crude bound (8.26) using the knowledge about $c_{\psi_0}^{S_1,S_2}$ gained in previous sections. Specifically, from (8.26), (8.27), and (8.33) we have

$$|c_{\psi^*}^{S_1,S_2}| \leqslant c_{\psi_0}^{S_1,S_2} \ll \frac{\log Q}{P_{S_1}P_{S_2}}.$$

It follows that

$$E_{\text{Siegel}} \ll (\log Q)X^\beta\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ d_*\mid 4n^2P_{S_1}P_{S_2}}}\frac{1}{P_{S_1}P_{S_2}}. \tag{8.44}$$

To bound this sum, note that the condition $d_*\mid 4n^2P_{S_1}P_{S_2}$ implies that $d_1\mid P_{S_1}$ and $d_2\mid P_{S_2}$ for some divisors d_1,d_2 of d_* with $d_1d_2\geqslant d_*/4n^2$. Therefore

$$E_{\text{Siegel}} \ll (\log Q)^2X^\beta\tau(d_*)^2\sup_{\substack{d_1\mid d_*,d_2\mid d_*\\ d_1d_2\geqslant d_*/4n^2}}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ d_1\mid P_{S_1},d_2\mid P_{S_2}}}\frac{1}{P_{S_1}P_{S_2}} \ll (\log Q)^2X^\beta\tau(d_*)^2\cdot(\log Q)^2\cdot d_*^{-1}$$
$$\ll (\log X)X^\beta d_*^{2\eta-1} \ll X(\log X)X^{-C_\varepsilon d_*^{-\varepsilon}}d_*^{2\eta-1},$$

where in the last step we applied Siegel's theorem (8.3), and $\eta > 0$ is any positive constant coming from the divisor bound $\tau(d_*) \ll_\eta d_*^\eta$. Therefore

$$E_{\text{Siegel}} \ll X(\log X)\left(C_\varepsilon \frac{\log\log X}{\log X}\right)^{(1-2\eta)/\varepsilon} \ll \frac{X}{\log X} \tag{8.45}$$

no matter the value of d_*, taking any $\varepsilon, \eta < \frac{1}{4}$ for the second bound.

Remark. By adjusting the value (1.7) of Q to $\exp(\log^c X)$ for very small c, and if we were prepared to settle for a weaker error term $O(X(\log X)^{-1-c'})$ in our main theorem, we could take any value of ε here. In particular, we could take ε close to 1 (and in particular at least $\frac{1}{2}$), at which point one could apply the Landau–Page theorem, thereby rendering this part of the argument effective. (By contrast the implied constant in Siegel's bound (8.3) is ineffective.) However, we have already used Siegel's theorem in Proposition 4.4, so the argument as a whole is ineffective in its current form.

If one wanted to make the argument fully effective then one could work with Λ_{Siegel} (see the proof of Proposition 4.4) instead of $\Lambda_{\text{Cramér}}$, and truncated versions of it. This would introduce even more complexity to the present section.

8.8. **Evaluation of E_{main}.** We now come to the final task, which is the asymptotic evaluation of E_{main}, whose definition is (8.41). First, note that by Lemma 8.7 we may restrict the sum to $S_1 \cap S_2 = S_1 \cap T = \emptyset$, that is to say

$$E_{\text{main}} := X \prod_{p\in\mathcal{P}_Q}\Big(1-\frac{1}{p}\Big)^{-2} \sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ S_1\cap S_2=S_1\cap T=\emptyset\\ |S_1|,|S_2|\leqslant t}} (-1)^{|S_1|+|S_2|}\sigma(S_1,S_2). \tag{8.46}$$

For the remaining terms in the sum we have the formula (8.34) and the accompanying definition (8.35) of $\nu(p)$, which in particular gives the bound

$$\sigma(S_1,S_2) \ll \prod_{p\in S_1\cup S_2}\frac{1}{p-1}. \tag{8.47}$$

Using this, we show using another Brun sieve-type computation that we may remove the constraints $|S_1|, |S_2| \leqslant t$ in (8.46) with little penalty. Indeed, by (8.47) the error in doing this is

$$\begin{aligned}\ll X(\log Q)^2 \sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ |S_2|>t\\ S_1\cap S_2=\emptyset}} \prod_{p\in S_1\cup S_2}\frac{1}{(p-1)} &\ll X(\log Q)^2 \prod_{p\in\mathcal{P}_Q}\Big(1+\frac{1}{p-1}\Big)\cdot \sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|>t}}\prod_{p\in S}\frac{1}{p-1}\\ &\ll X(\log Q)^3 \sum_{\substack{S\subseteq\mathcal{P}_Q\\ |S|>t}}\prod_{p\in S}\frac{1}{p-1} \ll X(\log X)^{-10},\end{aligned}$$

where the last step follows from (8.17). Thus

$$E_{\text{main}} = X \prod_{p\in\mathcal{P}_Q}\Big(1-\frac{1}{p}\Big)^{-2} \sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ S_1\cap S_2=S_1\cap T=\emptyset}} (-1)^{|S_1|+|S_2|}\sigma(S_1,S_2) + O(X(\log X)^{-10}). \tag{8.48}$$

The main expression here is of a rather algebraic nature and we can evaluate it using (8.34). To do this, set

$$S_1' := S_1, \quad S_2' := S_2 \setminus T, \quad S_3' := T \setminus S_2, \quad S_4' := T \cap S_2.$$

Note that these sets are all disjoint and that S_2', S_4' give a partition of S_2, whilst $\bigcup_{i=1}^4 S_i'$ is a partition of $S_1 \cup S_2 \cup T$. Then we obtain from (8.34) that

$$\begin{aligned}\frac{rh_K}{\omega|\mathcal{O}_K^*|}\sum_{\substack{S_1,S_2\subseteq\mathcal{P}_Q\\ S_1\cap S_2=S_1\cap T=\emptyset}}(-1)^{|S_1|+|S_2|}\sigma(S_1,S_2) &= \sum_{\substack{S_1',S_2'\subseteq\mathcal{P}_Q\setminus T, S_1'\cap S_2'=\emptyset\\ S_3'\subseteq T, S_4'=T\setminus S_3'}}(-1)^{|S_1'|+|S_2'|+|S_4'|}P_{S_3'}\prod_{p\in\bigcup_{i=1}^4 S_i'}\nu(p)\\ &= \prod_{p\in\mathcal{P}_Q\setminus T}(1-2\nu(p))\cdot\prod_{p\in T}(p-1)\nu(p)\\ &= \nu(2)\cdot\prod_{3\leqslant p\leqslant Q}(1-2\nu(p))\cdot\prod_{\substack{p\mid n\\ p\geqslant 3}}\frac{(p-1)\nu(p)}{1-2\nu(p)}. \end{aligned} \tag{8.49}$$

(We are using here the fact that n is even, so $2\in T$; note that $1-2\nu(p)\neq 0$ for $p\geqslant 3$.)

Therefore from (8.48) we have

$$E_{\mathrm{main}} = X\frac{\omega r h_K}{|\mathcal{O}_K^*|}\cdot 4\nu(2)\cdot\prod_{3\leqslant p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}(1-2\nu(p))\cdot\prod_{\substack{p\mid n\\ p\geqslant 3}}\frac{(p-1)\nu(p)}{1-2\nu(p)} + O(X(\log X)^{-10}). \tag{8.50}$$

The product over p here is absolutely convergent, with each term having size $1-O(p^{-2})$, and so we may extend it to all p with a multiplicative loss of $1+O(1/Q)$, the effect of which may comfortably be absorbed into the $O(X(\log X)^{-10})$ error term.

Finally, we apply the class number formula (2.6), which in our notation gives

$$\frac{2\pi h_K}{|\mathcal{O}_K^*||\Delta|^{1/2}} = \prod_p p\nu(p). \tag{8.51}$$

This, the formula $\Delta = -4\omega^2 n_*$ and (8.50) combine to give

$$E_{\mathrm{main}} = \frac{\pi}{\sqrt{n}}\cdot 2\cdot\Bigg(\prod_{3\leqslant p\leqslant Q}\Big(1-\frac{1}{p}\Big)^{-2}\frac{1-2\nu(p)}{p\nu(p)}\Bigg)\cdot\prod_{\substack{p\mid n\\ p\geqslant 3}}\frac{(p-1)\nu(p)}{1-2\nu(p)} + O(X(\log X)^{-10}). \tag{8.52}$$

We claim that the main expression here is precisely $\frac{\pi}{\sqrt{n}}\kappa_n$ with κ_n as in (1.1), which is exactly the constant in Proposition 4.1. We verify this one prime p at a time (the contribution of the prime $p=2$ being 2). Here, once more, we use the fact that n is even.

If $(-n \mid p) = 1$ then $(\Delta \mid p) = 1$ and $p \nmid n$, so the contribution from p is

$$\Big(1-\frac{1}{p}\Big)^{-2}\frac{1-2\nu(p)}{p\nu(p)} = \frac{p(p-3)}{(p-1)^2}.$$

If $(-n \mid p) = -1$ then $(\Delta \mid p) = -1$ and $p \nmid n$, so the contribution from p is

$$\Big(1-\frac{1}{p}\Big)^{-2}\frac{1-2\nu(p)}{p\nu(p)} = \frac{p}{p-1}.$$

If $(-n \mid p) = 0$ then $p \mid n$, and in this case we also get a contribution from p of $p/(p-1)$, both when $p\geqslant 3$ and $p=2$. This verifies the claim, and therefore

$$E_{\mathrm{main}} = \frac{\pi\kappa_n}{\sqrt{n}}X + O(X(\log X)^{-10}).$$

From this, (8.40), (8.43), and (8.45), we obtain the conclusion of Proposition 8.5 in the case $\ell = 0$. Recall that the case $\ell \neq 0$ has already been established.

This concludes the proof of all the statements in the paper.

Appendix A. Properties of the Gowers and Gowers–Peluse norms

In this section, we collect some basic properties of Gowers–Peluse norms which were defined in Definition 5.1. Let $g : \mathbf{Z} \to \mathbf{C}$ be a finitely-supported function. Recall that $\Delta_h g(x) = g(x)\overline{g(x+h)}$ and that $\Delta_{(h,h')} g(x) = g(x+h)\overline{g(x+h)}$ as in (4.10) and (5.1) respectively. We begin by noting that

$$\sum_{x,h\in\mathbf{Z}} \Delta_h g(x) = \Big|\sum_x g(x)\Big|^2 \geqslant 0 \tag{A.1}$$

and that, for any probability measure μ on $\mathbf{Z}$,

$$\mathbb{E}_{h,h'\sim\mu} \sum_{x\in\mathbf{Z}} \Delta_{(h,h')} g(x) = \sum_{x\in\mathbf{Z}} \Big|\mathbb{E}_{h\sim\mu} g(x+h)\Big|^2 \geqslant 0. \tag{A.2}$$

We first record that the Gowers–Peluse norm is nonnegative and that we have a version of the Gowers–Cauchy–Schwarz inequality. To the latter end, if we have a collection $(f_\omega)_{\omega\in\{0,1\}^k}$ of functions, we define the Gowers–Peluse inner product

$$\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]} := \frac{1}{N}\sum_{x\in\mathbf{Z}} \mathbb{E}_{h_i,h_i'\sim\mu_i} \prod_{\omega\in\{0,1\}^k} \mathcal{C}^{|\omega|} f_\omega(x + (\mathbf{1}-\omega)\cdot h + \omega\cdot h'), \tag{A.3}$$

where here $\mathbf{1}-\omega = (1-\omega_i)_{i=1}^k$ and $h = (h_i)_{i=1}^k$, $h' = (h_i')_{i=1}^k$. Note in particular that if $f_\omega = f$ for all ω then

$$\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]} = \|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]},$$

with notation as in Definition 5.1.

Lemma A.1. *Let $\mu_1,\dots,\mu_k : \mathbf{Z} \to [0,1]$ be probability measures, let $N \geqslant 1$, and suppose that $f : \mathbf{Z} \to \mathbf{C}$ is a function. Then for any function $f : \mathbf{Z} \to \mathbf{C}$ we have*

$$\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]} \geqslant 0. \tag{A.4}$$

Moreover if $(f_\omega)_{\omega\in\{0,1\}^k}$ are functions from $\mathbf{Z}$ to $\mathbf{C}$ then we have that

$$\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]} \leqslant \prod_{\omega\in\{0,1\}^k} \|f_\omega\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}. \tag{A.5}$$

Proof. Statement (A.4) is immediate from (A.2) and the definition (5.2), taking

$$g = \mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\ 1\leqslant i\leqslant k-1}} \Delta_{(h_1,h_1')}\cdots\Delta_{(h_{k-1},h_{k-1}')} f, \qquad \mu = \mu_k,$$

in (A.2). For (A.5), we proceed by induction on k. First note that by splitting $\omega = (\tilde\omega,\varepsilon)$ with $\tilde\omega \in \{0,1\}^{k-1}$ and $\varepsilon\in\{0,1\}$, and writing $\tilde h = (h_i)_{i=1}^{k-1}$ and $\tilde h' = (h_i')_{i=1}^{k-1}$, we may write the Gowers–Peluse inner product $\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}$ as

$$\frac{1}{N}\cdot\sum_{x\in\mathbf{Z}} \mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\ 1\leqslant i\leqslant k-1}} \prod_{\varepsilon\in\{0,1\}} \Big(\mathbb{E}_{h_k\sim\mu_k} \prod_{\tilde\omega\in\{0,1\}^{k-1}} \mathcal{C}^{|\tilde\omega|+\varepsilon} f_{(\tilde\omega,\varepsilon)}(x + (\mathbf{1}-\tilde\omega)\cdot\tilde h + \tilde\omega\cdot\tilde h' + h_k)\Big).$$

By Cauchy-Schwarz, this is bounded above by

$$\prod_{\varepsilon\in\{0,1\}} \Big(\frac{1}{N}\cdot\sum_{x\in\mathbf{Z}} \mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\ 1\leqslant i\leqslant k-1}} \Big|\mathbb{E}_{h_k\sim\mu_k} \prod_{\tilde\omega\in\{0,1\}^{k-1}} \mathcal{C}^{|\tilde\omega|+\varepsilon} f_{(\tilde\omega,\varepsilon)}(x + (\mathbf{1}-\tilde\omega)\cdot\tilde h + \tilde\omega\cdot\tilde h' + h_k)\Big|^2\Big)^{1/2},$$

which upon expanding the square and introducing a dummy variable h_k' may be rewritten as

$$\prod_{\varepsilon\in\{0,1\}} \Big(\frac{1}{N}\cdot\sum_{x\in\mathbf{Z}} \mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\ 1\leqslant i\leqslant k-1}} \mathbb{E}_{h_k,h_k'\sim\mu_k} \prod_{\tilde\omega\in\{0,1\}^{k-1}} \Delta_{(h_k,h_k')}\mathcal{C}^{|\tilde\omega|} f_{(\tilde\omega,\varepsilon)}(x + (\mathbf{1}-\tilde\omega)\cdot\tilde h + \tilde\omega\cdot\tilde h')\Big)^{1/2}.$$

By induction (that is, by (A.5) in the case $k-1$) this is bounded by

$$\prod_{\varepsilon\in\{0,1\}}\left(\mathbb{E}_{h_k,h_k'\sim\mu_k}\prod_{\tilde{\omega}\in\{0,1\}^{k-1}}\|\Delta_{(h_k,h_k')}f_{(\tilde{\omega},\varepsilon)}\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_{k-1}]}\right)^{1/2}.$$

Using Hölder's inequality, this is at most

$$\prod_{\varepsilon\in\{0,1\}}\prod_{\tilde{\omega}\in\{0,1\}^{(k-1)}}\left(\mathbb{E}_{h_k,h_k'\sim\mu_k}\|\Delta_{(h_k,h_k')}f_{(\tilde{\omega},\varepsilon)}\|^{2^{k-1}}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_{k-1}]}\right)^{1/2^k},$$

which equals

$$\prod_{\varepsilon\in\{0,1\}}\prod_{\tilde{\omega}\in\{0,1\}^{(k-1)}}\|f_{(\tilde{\omega},\varepsilon)}\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}=\prod_{\omega\in\{0,1\}^k}\|f_\omega\|_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}.$$

This concludes the proof. □

The next lemma asserts a kind of monotonicity of Gowers–Peluse norms with respect to the sequence of measures. It is essentially identical to [32, Lemma 3.5 (iii)], but formulated using measures.

Lemma A.2. *Fix $\delta\in(0,1/2)$, $N\geqslant 1$ and let $f:\mathbf{Z}\to\mathbf{C}$ be a 1-bounded function supported on $[\pm N]$. Let $\mu_1,\dots,\mu_k$ be probability measures supported on $[\pm N]$. Then*

$$\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}\geqslant\frac{1}{2k+3}\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_{k-1}]}.$$

Proof. We have

$$\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}=\frac{1}{N}\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant k-1}}\sum_{x\in\mathbf{Z}}\left|\mathbb{E}_{h_k\sim\mu_k}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{k-1},h_{k-1}')}f(x+h_k)\right|^2.$$

Observe that, since f is supported on $[\pm N]$, the sum over x is supported on $|x|\leqslant(k+1)N$. By Cauchy-Schwarz, it follows that

$$\begin{aligned}\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_k]}&\geqslant((2k+3)N^2)^{-1}\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant k-1}}\left|\sum_{x\in\mathbf{Z}}\mathbb{E}_{h_k\sim\mu_k}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{k-1},h_{k-1}')}f(x+h_k)\right|^2\\&=((2k+3)N^2)^{-1}\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant k-1}}\left|\sum_{x\in\mathbf{Z}}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{k-1},h_{k-1}')}f(x)\right|^2\\&\geqslant(2k+3)^{-1}\left|\frac{1}{N}\cdot\mathbb{E}_{\substack{h_i,h_i'\sim\mu_i\\1\leqslant i\leqslant k-1}}\sum_{x\in\mathbf{Z}}\Delta_{(h_1,h_1')}\cdots\Delta_{(h_{k-1},h_{k-1}')}f(x)\right|^2\\&=(2k+3)^{-1}\|f\|^{2^k}_{U_{\mathrm{GP}}[N;\mu_1,\dots,\mu_{k-1}]}.\end{aligned}$$

□

Appendix B. Proof of concatenation estimates

We give a complete proof of Lemma 5.5 in this section. Our proof follows that in [32, Section 6], which in turn closely follows [33, Section 2], with certain modifications. The reader may wish to review the basic notation for probability measures in Section 1.4. In this section, all probability measures are additionally assumed symmetric, that is to say $\mu(x)=\mu(-x)$.

Managing notation is problematic in the arguments in this section. To ease matters at least a little, we write $\|f\|^\bullet_{U_{\mathrm{GP}}[N;\Omega]}$, we mean $\|f\|^{2^d}_{U_{\mathrm{GP}}[N;\Omega]}$, where $d=|\Omega|$. (That is, the Gowers–Peluse norm is always raised to the 'obvious' power.)

A convenient piece of notation is to write $\binom{[N]}{r}$ for the collection of r-element subsets of $[N]$, and more generally $\binom{[M,N]}{r}$ for the collection of r-element subsets of the discrete interval $[M,N]$.

We begin by showing that the general case of Lemma 5.5 follows from the case $m=2$. This is a minor modification of [33, Corollary 2.5] or [32, Corollary 6.4].

Proof of Lemma 5.5, assuming the case $m=2$. For the proof, it is convenient to restate the hypothesis and conclusion of Lemma 5.5 in a slightly more abstract form. We define a *measure tuple* $\mathcal{T}=(\Omega,I,s)$ to be the following data: I is an indexing set, $s\geqslant 1$ is an integer, and for each $i\in I$, we have a sequence $\Omega_i=(\mu_{1i},\dots,\mu_{si})$ of symmetric probability measures on $\mathbf{Z}$. Given such a tuple, and given $t,m\geqslant 1$, we define the tuple $\wedge^{m,t}\mathcal{T}=(\Omega',I',s')$, where $I'=I^t$, $s'=s\binom{t}{m}$, and

$$\Omega'_{i_1,\dots,i_t}:=\Big(\mathop{*}_{r\in R}\mu_{ji_r}\Big)_{j\in[s],R\in\binom{[t]}{m}},$$

where we place an arbitrary order here in order to create a sequence rather than a just a set.

With this notation, the hypothesis of Lemma 5.5 is that we have a measure tuple $\mathcal{T}=(\Omega,I,s)$, with all the measures in each Ω_i supported on $[\pm N]$, and a 1-bounded function $f:\mathbf{Z}\to\mathbf{C}$ such that

$$\mathbb{E}_{i\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega_i]}\geqslant\delta. \tag{B.1}$$

The conclusion for a given value of m, which we call Lemma 5.5(m), is that there exists $t=t_0(m,s)$ such that, if we write $\mathcal{T}'=\wedge^{m,t}\mathcal{T}=(\Omega',I',s')$, we have

$$\mathbb{E}_{i'\in I'}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega'_{i'}]}\geqslant\delta^{O_{m,s}(1)}.$$

For technical reasons which will be apparent shortly, we work with the equivalent statement

$$\mathbb{E}_{i'\in I'}\|f\|^{\bullet}_{U_{\mathrm{GP}}[mN;\Omega'_{i'}]}\geqslant\delta^{O_{m,s}(1)}. \tag{B.2}$$

(The equivalence is clear from the definition of the Gowers–Peluse norm (5.2); the factor of m may be absorbed into a power $\delta^{O_m(1)}$.)

Suppose we have established Lemma 5.5(m) and Lemma 5.5(2). Set $t:=t_0(m,s)$ and $s'=s\binom{t}{m}$. Thus, we have (B.2) for this value of m. As can be seen from the similarity in form between (B.1) and (B.2), we can feed this into Lemma 5.5(2), taking the value $N'=mN$ in this application of the lemma. Note that the underlying measures, which are all m-fold convolutions of the μ_{ji}, are all supported on $[\pm N']$ and so the lemma applies.

Setting $t':=t_0(2,s')$, the conclusion is that if we write $\wedge^{2,t'}\mathcal{T}'=(\Omega'',I'',s'')$ then

$$\mathbb{E}_{i''\in I''}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N';\Omega''_{i''}]}\geqslant\delta^{O_{m,s}(1)}. \tag{B.3}$$

Unpacking the definitions, we see that $I''=(I^t)^{t'}$, and if $i''=\big((i_u)_v\big)_{u\in[t],v\in[t']}\in I''$ then

$$\Omega''_{i''}=\Big(\mathop{*}_{a\in A}\mathop{*}_{b\in B}\mu_{j,(i_b)_a}\Big)_{j\in[s],A\in\binom{[t']}{2},B\in\binom{[t]}{m}}.$$

We enlarge these collections of measures to

$$\tilde{\Omega}''_{i''}=\Big(\mathop{*}_{r\in R}\mu_{j,i_r}\Big)_{j\in[s],R\in\binom{[tt']}{2m}},$$

here abusing notation by identifying $(I^t)^{t'}$ with $I^{tt'}$, and noting the natural inclusion $\binom{[t']}{2}\times\binom{[t]}{m}\hookrightarrow\binom{[tt']}{2m}$. By Lemma A.2, this enlargement cannot decrease the Gowers–Peluse norms by more than a factor $O_{m,s}(1)$, arising from the inconsequential factors of $2k+3$ in that lemma. Thus, (B.3) implies that

$$\mathbb{E}_{i''\in I''}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N';\tilde{\Omega}''_{i''}]}\geqslant\delta^{O_{m,s}(1)}. \tag{B.4}$$

However, unpicking the notation we see that $\wedge^{2m,tt'}\mathcal{T} = (\tilde{\Omega}'', I'', s'')$, so (B.4) is precisely the conclusion of Lemma 5.5 with m replaced by $2m$, and with N replaced by N'. Much as before, we can trivially change N' to N; in fact, this increases the Gowers–Peluse norm.

We have therefore shown that Lemma 5.5(m) and Lemma 5.5(2) imply Lemma 5.5($2m$). By induction the whole of Lemma 5.5 (for all powers of two m) follows from the (as yet unestablished) base case $m = 2$. Moreover, we have shown that we can take

$$t_0(2m, s) = t_0\left(2, s\binom{t_0(m,s)}{m}\right)t_0(m, s). \tag{B.5}$$

□

Remark. We will show below that we can take $t_0(2, s) = 2^s$. In the main part of the paper, the Gowers norm we require is the U^k-norm with $k = 2 + 6\binom{t_0(8,3)}{8}$; by repeated use of (B.5) one may see that this is roughly on the order of $2^{2^{345}}$, that is to say surprisingly close to a googolplex. If one wanted control of Type II sums in Proposition 4.3 up to $L \leqslant X^{1/2-\kappa}$ using a Gowers U^k-norm then, via this argument, k would need to be a tower of twos of height $\sim \log(1/\kappa)$).

We are now free to focus exclusively on the case $m = 2$ of Lemma 5.5, and consequently it is somewhat natural to use the language of graphs. Suppose throughout the following that we have sequences of measures $\Omega_i = (\mu_{1i}, \dots, \mu_{si})$, $i \in I$, as in the statement of Lemma 5.5.

Definition B.1. Let $t \geqslant 1$ be an integer. Then a *graph system* Γ of level t is the following data. For each $j \in [s]$ we have a set $V_j \subseteq [t]$ of vertices and a set $E_j \subseteq \binom{[t]}{2}$ of edges. (The edges do not have to be between the vertices in V_j.) To a graph system of level t, and for any tuple $(i_1, \dots, i_t) \in I^t$, we associate the following collection $\Omega^{\Gamma}_{i_1,\dots,i_t}$ of probability measures: μ_{ji_v} for $v \in V_j$, and $\mu_{ji_v} * \mu_{ji_w}$ for $(v, w) \in E_j$, $j = 1, \dots, s$.

With this definition, the case $m = 2$ of Lemma 5.5 (with $t_0(2, s) = 2^s$) can be rephrased as follows.

Lemma B.2. *Let $\delta \in (0, \frac{1}{2})$, and let s, I, Ω_i be as above. Let $f : \mathbf{Z} \to \mathbf{C}$ be a 1-bounded function such that*

$$\mathbb{E}_{i\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega_i]} \geqslant \delta. \tag{B.6}$$

Then

$$\mathbb{E}_{i_1,\dots,i_M\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega^{\Gamma}_{i_1,\dots,i_M}]} \gg \delta^{O_s(1)}, \tag{B.7}$$

where $M = 2^s$ and Γ is the graph system at level $M := 2^s$ such that $V_j = \emptyset$ and $E_j = \binom{[M]}{2}$ for all $j \in [s]$.

Thus our remaining task is to prove Lemma B.2. The proof idea is as follows: we note that the assumption (B.6) can be written

$$\mathbb{E}_{i_1\in I^1}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega^{\Gamma^1}_{i_1}]} \geqslant \delta \tag{B.8}$$

where $V^1_j = \{1\}$ and $E^1_j = \emptyset$. Thus, our aim is to move from (B.8) (with the graphs 'vertex-complete' but 'edge-empty') to the conclusion (B.7) (with the graphs 'vertex-empty' but 'edge-complete'). We will do this by incrementing the level parameter of the graph system from 1 to $M = 2^s$ in single steps.

The following key lemma is the driver for this process. A version of this lemma where the μ_{si} are uniform measures on subgroups in an ambient abelian group appears in the work of Kuca [33, Lemma 2.2]. A slight variant of the precise lemma we prove appears as [32, Lemma 6.1].

Lemma B.3. *Fix $\delta \in (0,1/2)$, $k \geqslant 1$, and I an indexing set. For each $i \in I$, let $\nu_{1i}, \dots, \nu_{ki}$ be symmetric probability measures supported on $[\pm N]$. Let $f : \mathbf{Z} \to \mathbf{C}$ be 1-bounded such that* $\mathrm{supp}(f) \subseteq [\pm N]$. *Suppose that*

$$\mathbb{E}_{i\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{ki}]} \geqslant \delta.$$

Then

$$\mathbb{E}_{i,i'\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'},\nu_{ki}*\nu_{ki'}]} \geqslant \delta^{O_k(1)}.$$

Proof. Written out, the hypothesis is

$$\mathbb{E}_{i\in I}\sum_{x'\in\mathbf{Z}}\sum_{h'_1,\dots,h'_k\in\mathbf{Z}}\sum_{h''_1,\dots,h''_k\in\mathbf{Z}}\Big(\prod_{j=1}^{k}\nu_{ji}(h'_j)\nu_{ji}(h''_j)\Big)\Delta_{(h'_1,h''_1)}\cdots\Delta_{(h'_k,h''_k)}f(x') \geqslant \delta N.$$

Substituting $h_j := h''_j - h'_j$ and $x := x' + \sum_{i=1}^{k} h'_i$, this is equivalent to

$$\mathbb{E}_{i\in I}\sum_{x\in\mathbf{Z}}\sum_{h_1,\dots,h_k\in\mathbf{Z}}\Big(\prod_{j=1}^{k}\nu_{ji}^{(2)}(h_j)\Big)\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_k}f(x) \geqslant \delta N, \tag{B.9}$$

where here $\nu_{ji}^{(2)} = \nu_{ji} * \nu_{ji}$. In what follows, for $t \geqslant 1$ and $\ell = (\ell_1, \dots, \ell_t) \in \mathbf{Z}^t$ we denote

$$(\Delta_{\ell_1}\cdots\Delta_{\ell_t})^* f(x) = \prod_{\omega\in\{0,1\}^t\setminus\{0\}}\mathcal{C}^{|\omega|}f(x+\omega\cdot\ell),$$

where as usual $\mathcal{C}$ denotes complex conjugation. With this notation, (B.9) becomes

$$\sum_{x\in\mathbf{Z}}f(x)\cdot\mathbb{E}_{i\in I}\sum_{h_1,\dots,h_k\in\mathbf{Z}}\Big(\prod_{j=1}^{k}\nu_{ji}^{(2)}(h_j)\Big)(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_k})^* f(x) \geqslant \delta N.$$

Via Cauchy–Schwarz, we have that

$$\sum_{x\in\mathbf{Z}}\mathbb{E}_{i,i'\in I}\sum_{\substack{h_1,\dots,h_k\in\mathbf{Z}\\ h'_1,\dots,h'_k\in\mathbf{Z}}}\Big(\prod_{j=1}^{k}\nu_{ji}^{(2)}(h_j)\nu_{ji'}^{(2)}(h'_j)\Big)(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_k})^* f(x)(\Delta_{h'_1}\Delta_{h'_2}\cdots\Delta_{h'_k})^*\overline{f}(x) \gg \delta^2 N.$$

We may write this as

$$\sum_{x\in\mathbf{Z}}\mathbb{E}_{i,i'\in I}\sum_{\substack{h_1,\dots,h_k\in\mathbf{Z}\\ h'_1,\dots,h'_k\in\mathbf{Z}}}\Big(\prod_{j=1}^{k}\nu_{ji}^{(2)}(h_j)\nu_{ji'}^{(2)}(h'_j)\Big)\times$$
$$f(x+h_k)\overline{f(x+h'_k)}(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_{k-1}})^*(\Delta_{h_k}f)(x)(\Delta_{h'_1}\Delta_{h'_2}\cdots\Delta_{h'_{k-1}})^*(\Delta_{h'_k}\overline{f})(x) \gg \delta^2 N.$$

We now isolate the terms corresponding to $\nu_{ki}^{(2)}$ and $\nu_{ki'}^{(2)}$; in particular the above is equivalent to

$$\mathbb{E}_{i,i'\in I}\sum_{h_k,h'_k\in\mathbf{Z}}\nu_{ki}^{(2)}(h_k)\nu_{ki'}^{(2)}(h'_k)\sum_{x\in\mathbf{Z}}\sum_{\substack{h_1,\dots,h_{k-1}\in\mathbf{Z}\\ h'_1,\dots,h'_{k-1}\in\mathbf{Z}}}\Big(\prod_{j=1}^{k-1}\nu_{ji}^{(2)}(h_j)\nu_{ji'}^{(2)}(h'_j)\Big)\times$$
$$f(x+h_k)\overline{f(x+h'_k)}(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_{k-1}})^*\Delta_{h_k}f(x)(\Delta_{h'_1}\Delta_{h'_2}\cdots\Delta_{h'_{k-1}})^*\Delta_{h'_k}\overline{f}(x) \gg \delta^2 N. \tag{B.10}$$

For fixed i, i', h_k, h'_k, we now interpret the inner sum over $x, h_1, \dots, h_{k-1}, h'_1, \dots, h'_{k-1}$ as a Gowers–Peluse inner product of dimension $2k-2 = (k-1)+(k-1)$ with respect to the measures $\nu_{1i}, \dots, \nu_{(k-1)i}, \nu_{1i'}, \dots, \nu_{(k-1)i'}$. This is the key idea of the proof. As preparation, consider

an arbitrary such inner product

$$\langle (F_{\omega,\omega'})_{\omega,\omega'\in\{0,1\}^{k-1}}\rangle_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'}]} \tag{B.11}$$

for some functions $F_{\omega,\omega'}$ (here of course we identify $\{0,1\}^{2k-2}$ with the product of two copies of $\{0,1\}^{k-1}$). Written out in full, this is

$$\frac{1}{N}\sum_{\substack{x'\in\mathbf{Z}\\ a_j,b_j\in\mathbf{Z}\\ a'_j,b'_j\in\mathbf{Z}}}\prod_{j=1}^{k-1}\nu_{ji}(a_j)\nu_{ji}(b_j)\nu_{ji'}(a'_j)\nu_{ji'}(b'_j)\prod_{\substack{\omega\in\{0,1\}^{k-1}\\ \omega'\in\{0,1\}^{k-1}}}\mathcal{C}^{|\omega|+|\omega'|}F_{\omega,\omega'}(x'+(\mathbf{1}-\omega)\cdot a+\omega\cdot b+(\mathbf{1}-\omega')\cdot a'+\omega'\cdot b').$$

Substitute $h_j := b_j - a_j$, $h'_j := b'_j - a'_j$, $x := x' + \mathbf{1}\cdot a + \mathbf{1}\cdot a'$, and we see that this equals

$$\frac{1}{N}\sum_{x\in\mathbf{Z}}\sum_{\substack{h_1,\dots,h_{k-1}\in\mathbf{Z}\\ h'_1,\dots,h'_{k-1}\in\mathbf{Z}}}\Big(\prod_{j=1}^{k-1}\nu_{ji}^{(2)}(h_j)\nu_{ji'}^{(2)}(h'_j)\Big)\prod_{\omega,\omega'\in\{0,1\}^{k-1}}\mathcal{C}^{|\omega|+|\omega'|}F_{\omega,\omega'}(x+\omega\cdot h+\omega'\cdot h').$$

(The computation here is very similar to the one leading to (B.9).) One may now see that the inner sum in (B.10) can be written as N times the Gowers–Peluse inner product (B.11), where $F_{\mathbf{0},\mathbf{0}} = \Delta_{(h_k,h'_k)}f$, $F_{\omega,\omega'}(x) = \Delta_{h_k}f$ when only ω' is zero, $F_{\omega,\omega'}(x) = \Delta_{h'_k}\overline{f}(x)$ when only ω' is zero and all the other $F_{\omega,\omega'}$ are $1_{|x|\leqslant 10kN}$. Note that, due to the support properties of f and the measures ν, the insertion of these cutoffs is harmless and makes no difference to the expressions; note also that h_k, h'_k are being considered as fixed.

From this rewriting of the inner sum in (B.10) and the inequality (A.5), it follows that

$$\mathbb{E}_{i,i'\in I}\sum_{h_k,h'_k\in\mathbf{Z}}\nu_{ki}^{(2)}(h_k)\nu_{ki'}^{(2)}(h'_k)\|\Delta_{(h_k,h'_k)}f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'}]}\gg\delta^{O_k(1)}.$$

Via shift invariance of the Gowers–Peluse norm, we have

$$\mathbb{E}_{i,i'\in I}\sum_{t\in\mathbf{Z}}(\nu_{ki}^{(2)}*\nu_{ki'}^{(2)})(t)\|\Delta_t f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'}]}\gg\delta^{O_k(1)}.$$

This is equivalent to

$$\mathbb{E}_{i,i'\in I}\sum_{t_1,t_2\in\mathbf{Z}}(\nu_{ki}*\nu_{ki'})(t_1)(\nu_{ki}*\mu_{ki'})(t_2)\|\Delta_{(t_1,t_2)}f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'}]}\gg\delta^{O_k(1)}.$$

Finally, this is equivalent to

$$\mathbb{E}_{i,i'\in I}\|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\nu_{1i},\dots,\nu_{(k-1)i},\nu_{1i'},\dots,\nu_{(k-1)i'}\nu_{ki}*\nu_{ki'}]}\geqslant\delta^{O_k(1)},$$

as desired. □

We return now to the proof of Lemma B.2, which is an equivalent rephrasing of the case $m = 2$ of the main result of interest, Lemma 5.5. The task now is to explain how repeated use of Lemma B.3 leads to the proof of that result. In this we follow the arguments of [33, pages 10–11] or the proof of [32, Proposition 6.2]. Here is the statement of the steps we will take.

Definition B.4. Suppose we have a graph system Γ of level t (see Definition B.1). Let $k\in[s]$ and let $u\in[t]$, and suppose that $u\in V_k$. Then the *duplication* of Γ along (k,u) is the graph system $\tilde{\Gamma}$ at level $t+1$ defined as follows:

- if $j\in[s]\setminus\{k\}$ then include all of $V_j\setminus\{u\}$ in $\tilde{V}_j$;
- if $j\in[s]\setminus\{k\}$ and if $u\in V_j$, include both u and $t+1$ in $\tilde{V}_j$;
- Set $\tilde{V}_k = V_k\setminus\{u\}$;
- If $j\in[s]$ then include all edges $(v,w)\in E_j$ with $w\neq u$ in $\tilde{E}_j$;

- If $j \in [s]$ and if $(v,u) \in E_j$, include both (v,u) and $(v,t+1)$ in $\tilde{E}_j$;
- Include $(u,t+1)$ in $\tilde{E}_k$.

An *enlargement* of Γ is simply any graph system of level t obtained by adding further edges (on the same underlying vertex set $[t]$).

Lemma B.5. *Let $\delta \in (0,\frac{1}{2})$, and let s, I, Ω_i be as above. Let $t \geqslant 1$ be an integer, and suppose that Γ is a graph system at level t. Let $f : \mathbf{Z} \to \mathbf{C}$ be a 1-bounded function such that*

$$\mathbb{E}_{i_1,\dots,i_t \in I} \|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega^{\Gamma}_{i_1,\dots,i_t}]} \geqslant \delta. \tag{B.12}$$

Then, if $\tilde{\Gamma}$ is obtained from Γ by duplication and enlargement we have

$$\mathbb{E}_{i_1,\dots,i_{t+1} \in I} \|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega^{\tilde{\Gamma}}_{i_1,\dots,i_{t+1}}]} \geqslant \delta^{O(1)}. \tag{B.13}$$

Proof. The notion of duplication (Definition B.4) is symmetric with respect to permuting $[s]$ and $[t]$, so we may assume that the duplication is with respect to (s,t). In particular, $t \in V_s$.

We start with (B.12) and write out the Gowers–Peluse norm in terms of derivatives. Thus, for a fixed index tuple $(i_1,\dots,i_t)$, we have an average

$$\frac{1}{N}\sum_{x \in \mathbf{Z}} \mathbb{E}_{h_{(v,w)},h'_{(v,w)} \sim \mu_{ji_v} * \mu_{ji_w}} \mathbb{E}_{h_v,h'_v \sim \mu_{ji_v}} \left\{\Delta_{h_{(v,w)},h'_{(v,w)}}\right\}_{(v,w)\in E_j, j\in[s]} \left\{\Delta_{(h_v,h'_v)}\right\}_{v\in V_j, j\in[s]} f(x),$$

where here the notation means that we take *all* the indicated derivatives (the order being immaterial), and the h,h'-averages are over all these derivatives. (For instance, $\mathbb{E}_{h_i\sim\mu_i}\{\Delta_{h_i}\}_{i\in[3]}$ means the same as $\mathbb{E}_{h_1\sim\mu_1,h_2\sim\mu_2,h_3\sim\mu_3}\Delta_{h_1}\Delta_{h_2}\Delta_{h_3}$.) Now take the average over indices $(i_1,\dots,i_t)$, and rearrange this so that only terms involving i_t are on the inside. To notate this, it is convenient to write E'_j for the subset of E_j consisting of edges (v,t), and V'_j for the subset of V_j which is either $\{t\}$ if $t \in V_j$, or $\emptyset$ otherwise. Thus we get

$$\mathop{\mathbb{E}}_{\substack{i_1,\dots,i_{t-1}\in I\\ h_{(v,w)},h'_{(v,w)}\sim\mu_{ji_v}*\mu_{ji_w}\\ h_v,h'_v\sim\mu_{ji_v}}} \mathbb{E}_{i_t\in I} \left\| \left\{\Delta_{h_{(v,w)},h'_{(v,w)}}\right\}_{\substack{(v,w)\in E_j\setminus E'_j\\ j\in[s]}} \left\{\Delta_{h_v,h'_v} f\right\}_{\substack{v\in V_j\setminus V'_j\\ j\in[s]}} \right\|^{\bullet}_{U_{\mathrm{GP}}[N;\tilde{\Omega}_{i_1,\dots,i_t}]} \gg \delta^{O(1)},$$

where here $\tilde{\Omega}_{i_1,\dots,i_t}$ are the measures which do involve i_t, namely the $\mu_{ji_v} * \mu_{ji_t}$ for $(v,t) \in E_j$ for all $j \in [s]$, and the μ_{ji_t} for j such that $t \in V_j$. In particular, since $t \in V_s$, we see that the measure μ_{si_t} itself lies in $\tilde{\Omega}_{i_1,\dots,i_t}$.

We now apply Lemma B.3 with the ν's being the elements of $\tilde{\Omega}_{i_1,\dots,i_t}$ and with ν_{ki} in that lemma (that is, the measure to be 'duplicated') being μ_{si_t}, and with N replaced by $2N$ so that the support condition holds. Writing i_{t+1} for the resulting i' variables, the conclusion is then that

$$\mathop{\mathbb{E}}_{\substack{i_1,\dots,i_{t-1}\in I\\ h_{(v,w)},h'_{(v,w)}\sim\mu_{ji_v}*\mu_{ji_w}\\ h_v,h'_v\sim\mu_{ji_v}}} \mathop{\mathbb{E}}_{i_t,i_{t+1}\in I} \left\| \left\{\Delta_{h_{(v,w)},h'_{(v,w)}}\right\}_{\substack{(v,w)\in E_j\setminus E'_j\\ j\in[s]}} \left\{\Delta_{h_v,h'_v} f\right\}_{\substack{v\in V_j\setminus V'_j\\ j\in[s]}} \right\|^{\bullet}_{U_{\mathrm{GP}}[2N;\tilde{\Omega}^*_{i_1,\dots,i_t,i_{t+1}}]} \gg \delta^{O(1)},$$

where now $\Omega^*_{i_1,\dots,i_t,i_{t+1}}$ consists of:

- $\mu_{ji_v} * \mu_{ji_t}$ and $\mu_{ji_v} * \mu_{ji_{t+1}}$ for $(v,t) \in E_j$ for all $j \in [s]$;
- μ_{ji_t} and $\mu_{ji_{t+1}}$ for $j \in [s-1]$ such that $t \in V_j$;
- $\mu_{si_t} * \mu_{si_{t+1}}$.

Folding the outer derivatives back into the Gowers–Peluse norm notation (and replacing $2N$ by N, which loses only a factor of 2) one sees that

$$\mathbb{E}_{i_1,\dots,i_{t+1}\in I} \|f\|^{\bullet}_{U_{\mathrm{GP}}[N;\Omega^{\tilde{\Gamma}}]} \gg \delta^{O(1)},$$

where $\tilde{\Gamma}$ (of level $t+1$) consists of the following edges and vertices:

- $\tilde{V_j}$ contains $V_j \setminus \{t\}$ for $j \in [s-1]$;
- $\tilde{V_j}$ contains t and $t+1$ if $t \in V_j$ for $j \in [s-1]$;
- $\tilde{V_s} = V_s \setminus \{t\}$;
- $\tilde{E_j}$ contains $(v,w) \in E_j$ for $w < t$, for all $j \in [s]$;
- $\tilde{E_j}$ contains $(v,t), (v,t+1)$ if $(v,t) \in E_j$, for $j \in [s]$;
- $\tilde{E_s}$ contains $(t,t+1)$.

One may check that $\tilde{\Gamma}$ is precisely the duplication of Γ^t along (s,t); indeed the six bullets above are in 1-1 correspondence with those in Definition B.4, if one takes $(k,u) = (s,t)$ there.

Any enlargement of $\tilde{\Gamma}$ cannot decrease the Gowers–Peluse norm by more than harmless losses of $O(1)$ factors, by Lemma A.2. □

We now show how, by a sequence of duplications and enlargements, we may move from an edge-empty graph to an edge-complete one. This essentially combines the proof of [32, Proposition 6.2, Corollary 6.3] into a double induction.

Lemma B.6. *There is a sequence of graph systems Γ^t, $t = 1, 2, 3 \dots, 2^s$ such that Γ^1 is the 'vertex-complete, edge-empty' system in (B.8), and Γ^{2^s} is the 'edge-complete, vertex-empty' system in (B.7), and moreover Γ^{t+1} is obtained from Γ^t by a duplication and an enlargement.*

Proof. For $r = 0, 1, 2, \dots$ and for $0 \leqslant \ell < 2^r$ write $t = t(r,\ell) := 2^r + \ell$, and define the graph system Γ^t in terms of its edges and vertices, as follows:

$$E^t_{s-r+1} = \dots = E^t_s = \binom{[t]}{2}, \quad E^t_{s-r} = \binom{[2^r - \ell + 1, t]}{2}, \quad E^t_1 = \dots = E^t_{s-r-1} = \emptyset,$$

and

$$V^t_1 = \dots = V^t_{s-r-1} = [t], \quad V^t_{s-r} = [2^r - \ell], \quad V^t_{s-r+1} = \dots = V^t_s = \emptyset.$$

Note in particular that when $t = 1$ (and so $r = 0$ and $\ell = 0$) these graphs are exactly the graphs Γ^1 in (B.8). Duplicate Γ^t along $(s-r, 2^r - \ell)$. The resulting graph then has

$$E^{t+1}_{s-r+1} = \dots = E^{t+1}_s = \binom{[t+1]}{2}, \quad E^{t+1}_{s-r} = \binom{[2^r - \ell + 1, t]}{2} \cup \{(2^r - \ell, t+1)\},$$

and $E^t_1 = \dots = E^t_{s-r-1} = \emptyset$, and

$$V^{t+1}_1 = \dots = V^{t+1}_{s-r-1} = [t+1], \quad V^{t+1}_{s-r} = [2^r - \ell - 1], \qquad V^t_{s-r+1} = \dots = V^t_s = \emptyset.$$

Enlarging E^{t+1}_{s-r} to $\binom{[2^r - \ell, t+1]}{2}$ then gives Γ^{t+1}. (Note that this is true for $0 \leqslant \ell \leqslant 2^r - 2$ and also when $\ell = 2^r - 1$, when $t+1 = t(r+1, 0)$.)

□

Appendix C. A large sieve bound in several dimensions

In this appendix we give a brief discussion of the higher-dimensional large sieve, for which it is hard to find a suitable reference.

Proposition C.1. *Let θ_j be δ-well spaced points in the ℓ^∞ distance on $(\mathbf{R}/\mathbf{Z})^k$, let $(a_n)_{n \in [N]^k}$ be a sequence of complex numbers, and write*

$$S(\theta) := \sum_{n \in [N]^k} a_n e(\theta \cdot n). \tag{C.1}$$

Then

$$\sum_j |S(\theta_j)|^2 \leqslant (N^{1/2} + \delta^{-1/2})^{2k} \|a\|^2,$$

where $\|a\| := \big(\sum_n a_n\big)^{1/2}$.

Proof. See [27, Theorem 1]. We note that, for our purposes (where we take $k = 2$), it would suffice to have $\ll_k$ instead of $\leqslant$ in this inequality. If one's aim is a result of this type, the need for special trigonometric constructions as used in [27] (who quotes a construction of Bombieri and Davenport [1]) can be avoided and replaced with cruder bounds. See the remarks in [28, p177–178]. Note also that the proof of [28, Theorem 7.7] (who use Hilbert's inequality, following Montgomery and Vaughan [41]) does not adapt so straightforwardly to the high-dimensional case. For further discussion see [30, Chapter 4]. □

Corollary C.2. *With notation as above, we have*

$$\sum_{q_1,\dots,q_k\in[N^{1/2}]}\sum_{a_i\,(\mathrm{mod}\,q_i)}{}' \big|S\big((\tfrac{a_i}{q_i})_{i=1}^k\big)\big|^2 \leqslant (2N)^k\|a\|^2,$$

where the dash means that the sum is taken over a_i coprime to q_i.

Proof. This follows immediately from Proposition C.1 and the fact that the points $(\frac{a_i}{q_i})_{i=1}^k$ are N^{-1}-well spaced in the ℓ^∞ distance on $(\mathbf{R}/\mathbf{Z})^k$. □

Lemma C.3. *For each prime p, suppose that we have a set $\Omega_p \subseteq (\mathbf{Z}/p\mathbf{Z})^k$ of density α_p. Then*

$$\#\{|x|,|y|\leqslant N : \prod_p 1_{\Omega_p^c}(x,y) = 1\} \leqslant (2N)^k\Big(\sum_{\substack{q\leqslant N^{1/2}\\ \mu^2(q)=1}} h(q)\Big)^{-1}, \tag{C.2}$$

where h is the multiplicative function satisfying $h(p) = \alpha_p/(1-\alpha_p)$.

Proof. This is a routine generalisation of [28, Theorem 7.14] to several variables, using Corollary C.2 in place of [28, Theorem 7.11]; note that [28, Lemma 7.15], which is the main argument, goes over essentially verbatim to the multidimensional setting. □

Appendix D. Number-theoretic bounds

In this appendix we collect some more-or-less standard bounds of a number-theoretic flavour.

Lemma D.1. *For $Y \geqslant 1$ and for integer $m \geqslant 1$, we have $\sum_{|y|\leqslant Y}\tau(y)^m \ll_m Y(\log Y)^{2^m-1}$.*

Proof. This is standard; see e.g. [28, Equation (1.80)]. □

Lemma D.2. *Let $n \geqslant 1$ be fixed and let $r(t)$ be the number of representations of t by x^2+ny^2 with $x,y\in\mathbf{Z}$. Then $r(t)\leqslant 6\tau(t)$.*

Proof. We work with ideals in $K = \mathbf{Q}(\sqrt{-n})$. We have $t = x^2+ny^2$ if and only if $N((x+y\sqrt{-n})) = t$. Since a principal ideal is generated by $|\mathcal{O}_K^*| \leqslant 6$ different elements of $\mathcal{O}_K$, we see that $r(t)$ is bounded above by 6 times the number of ideals of norm t. To count the number of such ideals, factor $t = p_1^{a_1}\cdots p_k^{a_k}$ as a product of prime powers. The number of ideals of norm p^a is either $a+1$ (if p splits in K) or $1_{2|a}$ (if p is inert in K) or 1 if p ramifies. In all cases, the number of such ideals is at most $a+1$, and so the number of ideals of norm t is $\leqslant (a_1+1)\cdots(a_k+1) = \tau(t)$. □

Once in the main text, and also in Lemma D.4 below, we need the following classical estimate on the number of ideals of norm up to a given threshold.

Lemma D.3. *For any number field K, the number of ideals in $\mathcal{O}_K$ of norm $\leqslant X$ is $\ll_K X$. In the case of an imaginary quadratic field K, the number of ideals in $\mathcal{O}_K$ of norm $\leqslant X$ is $C_K X + O(X^{1/2+o(1)})$.*

Proof. Results of this form date back to work of Dedekind, Weber, and Hecke and take the form $C_K X + O_K(X^{1-[K:\mathbf{Q}]^{-1}})$. Here $C_K = |\mathcal{O}_K^*|^{-1}|\Delta|^{-1/2}2^{r_1+r_2}\pi^{r_2}R$ where R is the regulator and r_1, r_2 the number of real and pairs of conjugate complex embeddings. For a sketch proof, see [26, Theorem 1]. □

Finally, we need a weak Mertens-type upper bound in number fields.

Lemma D.4. *Let K be an algebraic number field. Then*

$$\sum_{\mathfrak{p}: N\mathfrak{p}\leqslant X} (N\mathfrak{p})^{-1} \ll_K \log\log X.$$

Proof. Consider the Dedekind ζ-function

$$\zeta_K(s) = \sum_{\mathfrak{a}} (N\mathfrak{a})^{-s} = \prod_{\mathfrak{p}} \big(1 - (N\mathfrak{p})^{-s}\big)^{-1}$$

for $s = 1 + (\log X)^{-1}$. On the one-hand, partial summation using Lemma D.3 gives

$$\zeta_K(s) = \int_1^\infty s x^{-s-1} \#\{\mathfrak{a} : N\mathfrak{a} \leqslant x\} dx \ll_K \frac{s}{s-1} \ll \log X.$$

On the other hand, using $(1-x)^{-1} \geqslant e^x$,

$$\zeta_K(s) = \prod_{\mathfrak{p}} \big(1 - (N\mathfrak{p})^{-s}\big)^{-1} \geqslant \exp\Big(\sum_{\mathfrak{p}} (N\mathfrak{p})^{-s}\Big) \geqslant \exp\Big(\frac{1}{e} \sum_{\mathfrak{p}: N\mathfrak{p}\leqslant X} (N\mathfrak{p})^{-1}\Big).$$

Combining these inequalities gives the stated bound. □

References

[1] Enrico Bombieri and Harold Davenport, *On the large sieve method*, Number Theory and Analysis (Papers in Honor of Edmund Landau), Plenum, New York, 1969, pp. 9–22. 57

[2] David A. Cox, *Primes of the form $x^2 + ny^2$*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1989, Fermat, class field theory and complex multiplication. 2

[3] Harold Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Revised and with a preface by Hugh L. Montgomery. 38

[4] William Duke, John B. Friedlander, and Henryk Iwaniec, *Equidistribution of roots of a quadratic congruence to prime moduli*, Ann. of Math. (2) **141** (1995), 423–441. 4, 8

[5] Ernests Fogels, *On the zeros of Hecke's L-functions. I*, Acta Arith. **7** (1961/62), 87–106. 38

[6] Ernests Fogels, *On the zeros of Hecke's L-functions. II*, Acta Arith. **7** (1961/62), 131–147. 38

[7] Ernests Fogels, *über die Ausnahmenullstelle der Heckeschen L-Funktionen*, Acta Arith. **8** (1962/63), 307–309. 38

[8] Kevin Ford and James Maynard, *On the theory of prime producing sieves*, arXiv:2407.14368. 3, 4, 13

[9] Étienne Fouvry and Henryk Iwaniec, *Gaussian primes*, Acta Arith. **79** (1997), 249–287. 2

[10] John B. Friedlander and Henryk Iwaniec, *Asymptotic sieve for primes*, Ann. of Math. (2) **148** (1998), 1041–1065. 2

[11] John B. Friedlander and Henryk Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes*, Ann. of Math. (2) **148** (1998), 945–1040. 2

[12] John B. Friedlander and Henryk Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, vol. 57, American Mathematical Society, Providence, RI, 2010. 42

[13] John B. Friedlander and Henryk Iwaniec, *Coordinate distribution of Gaussian primes*, J. Eur. Math. Soc. (JEMS) **24** (2022), 737–772. 2

[14] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588. 19

[15] Ben Green, *Classical analytic theory of Hecke Größencharaktere*, expository paper, in preparation. 38

[16] Ben Green and Mehtaab Sawhney, *Corners in the Gaussian primes*, Forthcoming. 3, 20

[17] Ben Green and Terence Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), 1753–1850. 4, 20

[18] Ben Green and Terence Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), 541–566. 4

[19] Ben Green and Terence Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), 465–540. 4

[20] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), 1231–1372. 4

[21] Glyn Harman, *Prime-detecting sieves*, London Mathematical Society Monographs Series, vol. 33, Princeton University Press, Princeton, NJ, 2007. 27

[22] Xiaoyu He, *Primes of the form p^2+ny^2*, Harvard Senior Thesis 2016, https://legacy-www.math.harvard.edu/theses/senior/he/he.pdf. 2

[23] D. R. Heath-Brown, *Primes represented by x^3+2y^3*, Acta Math. **186** (2001), 1–84. 3

[24] D. R. Heath-Brown and Xiannan Li, *Prime values of a^2+p^4*, Invent. Math. **208** (2017), 441–499. 2

[25] D. R. Heath-Brown and B. Z. Moroz, *Primes represented by binary cubic forms*, Proc. London Math. Soc. (3) **84** (2002), 257–288. 3, 5, 9

[26] Hans Heilbronn, *Zeta-functions and L-functions*, Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965), Academic Press, London, 1967, pp. 204–230. 58

[27] Martin N. Huxley, *The large sieve inequality for algebraic number fields*, Mathematika **15** (1968), 178–187. 57

[28] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004. 3, 15, 21, 37, 38, 57

[29] Dimitris Koukoulopoulos, *The distribution of prime numbers*, Graduate Studies in Mathematics, vol. 203, American Mathematical Society, Providence, RI, [2019] ©2019. 42

[30] Emmanuel Kowalski, *The large sieve and its applications*, Cambridge Tracts in Mathematics, vol. 175, Cambridge University Press, Cambridge, 2008, Arithmetic geometry, random walks and discrete groups. 57

[31] Noah Kravitz, Borys Kuca, and James Leng, *Corners with polynomial side length*, arXiv:2407.08637. 23

[32] Noah Kravitz, Borys Kuca, and James Leng, *Quantitative concatenation for polynomial box norms*, arXiv:2407.08636. 4, 5, 23, 26, 50, 51, 52, 54, 56

[33] Borys Kuca, *Multidimensional polynomial patterns over finite fields: bounds, counting estimates and Gowers norm control*, Adv. Math. **448** (2024), Paper No. 109700, 61. 4, 50, 51, 52, 54

[34] James Leng, *Efficient Equidistribution of Nilsequences*, arXiv:2312.10772. 4, 6, 21

[35] James Leng, Ashwin Sah, and Mehtaab Sawhney, *Quasipolynomial bounds on the inverse theorem for the Gowers $U^{s+1}[N]$-norm*, arXiv:2402.17994. 4, 19, 21

[36] Freddie Manners, *Quantitative bounds in the inverse theorem for the Gowers U^{s+1}-norms over cyclic groups*, arXiv:1811.00718. 4

[37] James Maynard, *Primes represented by incomplete norm forms*, Forum Math. Pi **8** (2020), e3, 128. 3, 5, 9, 38

[38] Jori Merikoski, *On gaussian primes in sparse sets*, arXiv:2302.11331. 2

[39] Jori Merikoski, *The polynomials $X^2+(Y^2+1)^2$ and $X^2+(Y^3+Z^3)^2$ also capture their primes*, Proc. Lond. Math. Soc. (3) **127** (2023), 1057–1133. 2

[40] Takayoshi Mitsui, *Generalized prime number theorem*, Jpn. J. Math. **26** (1956), 1–42. 38

[41] Hugh L. Montgomery and Robert C. Vaughan, *Hilbert's inequality*, J. London Math. Soc. (2) **8** (1974), 73–82. 57

[42] Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, Cambridge, 2007. 16, 38

[43] Sarah Peluse, *Bounds for sets with no polynomial progressions*, Forum Math. Pi **8** (2020), e16, 55. 4, 5, 23, 26

[44] Sarah Peluse and Sean Prendiville, *Quantitative bounds in the nonlinear Roth theorem*, arXiv:1903.02592. 4

[45] Sarah Peluse, Ashwin Sah, and Mehtaab Sawhney, *Effective bounds for Roth's theorem with shifted square common difference*, arXiv:2309.08359. 23

[46] Kevin Pratt, *Primes from sums of two squares and missing digits*, Proc. Lond. Math. Soc. (3) **120** (2020), 770–830. 2

[47] Terence Tao and Joni Teräväinen, *Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions*, J. Eur. Math. Soc. (2023), 1–64. 4, 6, 19, 21

[48] Terence Tao and Tamar Ziegler, *Concatenation theorems for anti-Gowers-uniform functions and Host-Kra characteristic factors*, Discrete Anal. (2016), Paper No. 13, 60. 4

[49] Heinrich M. Weber, *Beweis des satzes, dass jede eigentlich primitive quadratische form unendlich viele primzahlen darzustellen fähig ist*, Math. Ann. **20** (1882), 301–329. 2

Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6QW, UK
E-mail address: ben.green@maths.ox.ac.uk

Department of Mathematics, Columbia University, New York, NY 10027
E-mail address: m.sawhney@columbia.edu