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

XIAOYU HE

Contents

1. Introduction

It is not surprising that he would have been [misled], unsuspicious as he presumably is of the diabolical malice inherent in the primes.

– Littlewood to Hardy, about Ramanujan

The representation of primes by integer polynomials is one of the chief preoccupations of number theory. The (closely related) questions are: which polynomials take prime values, and how frequent are these values? The Prime Number Theorem answers both questions for the polynomial $f(n) = n$; Dirichlet's theorem answers the first for linear polynomials of the form $an + b$. Already when we go to the simplest nontrivial quadratic polynomials $f(n) = n^2 + 1$ we find a major open problem, first proposed by Landau.

When f is allowed to have two or more variables, the theory becomes extremely rich, tying in intimately with algebraic number theory. The easiest case $x^2 + y^2$ already depends on prime factorization in $\mathbb{Z}[i]$ – counting such primes asymptotically is a special case of the Prime Number Theorem in arithmetic progressions. The representation of primes by many such special forms relate closely to prime factorization of number fields of the corresponding degree, and the ideal class group makes an appearance as soon as unique factorization breaks down.

The question is even richer and almost entirely intractable for nonhomogeneous forms. The representation of primes by the polynomial $f(x, y) = x^2 + y^3$, for example, is closely tied to the counting of elliptic curves with prime discriminant; as it stands $x^2 + y^3$ is considered hopeless with our current technology. Special values of discriminant polynomials have great value in arithmetic statistics – it is an active area of research today to count squarefree values of discriminants, something much more tractable.

Analytic number theory brings (at least) two extremely disparate approaches to these problems. Historically the first approach to prove the Prime Number Theorem worked via complex analysis using zero-free regions of the Riemann zeta function and similar L-functions. The complex analysis approach provides proofs of the prime number theorem and many vast generalizations thereof – for example to arithmetic progressions or number fields. When these problems are tractable purely via L-functions, the estimates are often far superior to those obtainable by any other means.

The historically second approach is the sieve, which always begins via the purely combinatorial inclusion-exclusion formula. It took a long time for people to recognize the power of sieve theory; beyond its technical difficulty was a central problem with sieves called the parity problem which number theorists long thought rendered a large class of problems completely out of reach.

Classical sieve theory estimates the number of prime values of f by reducing it to a much easier problem: how many values of f lie in any given arithmetic progression? In modern language being able to answer this question is the one of providing Type I estimates. Unfortunately, using Type I estimates alone is only enough to count almost-primes, numbers with at most 2 prime factors. As we will see, Type I estimates cannot distinguish between numbers with odd and even numbers of divisors. In the end these "classical" sieves failed to provide asymptotic results as the bounds they provide are off by at least, and often exactly, a factor of two.

The reason these sieves fail is because they cannot distinguish between numbers between numbers with an even or odd number of prime factors – this is called the parity problem. Here are three aspects of the Parity Problem.

(1) The Selberg sieve is off by a factor of 2 when trying to give an upper bound for prime numbers. If $\pi(X)$ is the number of primes up through X, then

$$\pi(X) \leq \frac{(2+o(1))X}{\log X},$$

which secretly comes from the fact that we can only use the primes up to $\sqrt{X}$ to sieve, so the bound is more transparently written as

$$\pi(X) \leq \frac{(1+o(1))X}{\log \sqrt{X}}.$$

(2) It is much easier to count k-almost primes (products of up to k primes) weighted by the generalized von Mangoldt functions (see Definition 10), as long as $k \geq 2$. For $k = 2$ the Selberg Symmetry Formula

$$\sum_{p \leq X} \log^2 p + 2 \sum_{p<q \leq X} \log p \log q = 2X \log X + O(X)$$

is disappointingly easy to prove.

(3) Consider the function $f(n)$ which is 2 when n has an even number of prime factors and 0 otherwise; from the perspective of Type I estimates, this sequence is indistinguishable from the constant function. In contrast if $f(n)$ is 3 when n has a multiple of three prime factors, the sequence is suddenly distinguishable!

We will discuss this phenomenon in much more depth in Section 2.4. Although the parity problem is now common knowledge, the most exact formulation remains in Selberg's original lecture notes on sieves [12]. Our goal is to fully highlight how surprising the parity problem is and also sketch why there is no corresponding mod n problem for any $n \neq 2$.

Only recently has the parity problem been broken in a handful of cases by injecting a second kind of estimate, the so-called bilinear forms or Type II estimates. Fouvry and Iwaniec were first to do this in order to show $x^2 + p^2$ represents infinitely many primes, when p must also be prime. This method has been extended with some success, most notably to count primes of the forms $x^2 + y^4$ [8] and $x^3 + 2y^3$ [10].

A Type II estimate is a guarantee that about an equal number of values of f have an odd or even numbers of divisors. The motivation for a Type II estimate is a common general principle in mathematics: prove that the known obstructions to solving a problem are the *only* obstructions. The success of the new sieve shows that while the parity problem cannot be solved by Type I estimates on their own, it is the *only* obstruction to sieving for primes.

Using these techniques we contribute a new asymptotic count for primes of the form $p^2 + Ny^2$ for fixed $N > 0$. Let $\Lambda(n)$ be the von Mangoldt function, and define $g(n)$ to be the multiplicative function which on prime powers is

$$g(p^\alpha) = \begin{cases} \frac{1+\left(\frac{-N}{p}\right)}{p^\alpha} & podd, p \nmid N \\ \frac{1}{2} & p = 2, p \nmid N, \alpha = 1 \\ \frac{1-\chi_4(N)}{2^\alpha} & p = 2, p \nmid N, \alpha \geq 2 \\ 0 & p|N \end{cases}$$

where χ_4 is the nontrivial Dirichlet character mod 4.

Theorem 1. *For any $A, X > 0$,*

$$\sum_{x^2+Ny^2 \leq X} \Lambda(x)\Lambda(x^2+Ny^2) = \frac{\pi H_N X}{4\sqrt{N}} + O_A(X(\log X)^{-A}),$$

where the implied constant depends only on A and

$$H_N = \prod_{p \nmid N} (1-g(p))(1-p^{-1})^{-1}.$$

We remark that the estimates we refer to as Type I and Type II have close analogies in the Hardy-Littlewood Circle Method. There Type I bounds refer to major arcs and Type II bounds refer to minor arcs. Although the analogy between the Asymptotic Sieve for Primes and the Circle Method is quite interesting, especially since both are related to Vaughan's identity, we will make no attempt to flesh out this connection.

1.1. **Outline.** The goal of this paper is twofold.

First, we revisit the history of the sieve. We trace the development of the earliest sieves – Brun's combinatorial sieve and Selberg's sieve – that fall within the framework of exploiting only Type I estimates to their maximum potential. From here we move to how the parity problem came to be understood as the central obstruction to sieve theory, and how Bombieri perfected the art of getting asymptotic counts for almost-prime values using only Type I estimates. Only then can we talk about the injection of Type II estimates to count prime values themselves and the breakthrough of the Friedlander-Iwaniec Asymptotic Sieve for Primes.

Second, we extend the methods of Fouvry and Iwaniec to prove Theorem 1. While it doesn't directly apply the Asymptotic Sieve for Primes, it uses exactly the same tools and philosophy. When $\mathbb{Z}[\sqrt{-N}]$ is still a unique factorization domain, the generalization is straightforward. However, to deal with quadratic number fields with nontrivial class number new tools must be introduced to prove both the Type I and Type II estimates required. As a consequence this section will prove much more technical than the first.

2. An Intuitive Account of Sieve Theory

A detailed and rigorous exposition of all of these results and much more can be found in *Opera de Cribro* by Iwaniec and Friedlander [6]. A good exposition of the prime-detecting sieve in particular can be found in Harman [9]. In contrast to these sources, we dispense with the full generality of sieve theory in favor of focusing on the core ideas of each sieve by trying to count primes with them. We define the prime-counting function as

$$\pi(X) = |\{p \text{ prime} : p \leq X\}|.$$

We have Atle Selberg to thank for the best ideas that follow. In particular, the powerful Λ^2 sieve, the Symmetry Formula, and the clarification of the parity problem are all due to Selberg.

Henceforth, we use the notations $A(X) \ll B(X)$ and $A(X) = O(B(X))$ interchangeably to mean the two functions A, B satisfy $B(X) \geq 0$ for all X and

$$\limsup_{X\to\infty} \frac{|A(X)|}{B(X)} < \infty.$$

The following table shows what each sieve proves about the asymptotic behavior of $\pi(X)$.

We define additionally the function

$$\psi_k(X) = \sum_{n\leq X} \Lambda_k(n),$$

where Λ_k is the generalized von Mangoldt function (see Definition 10). This function $\psi_k(X)$ counts the k-almost primes up through X weighted in a special way – note that in general it does NOT weight numbers with i prime factors the same, for each $1 \leq i \leq k$ (see Section 2.5.2).

Sieve	Bound on Prime Counting Function
Eratosthenes	$\pi(X) \ll \frac{X}{\log\log X}$
Brun's Pure Sieve	$\pi(X) \ll \frac{X \log\log X}{\log X}$
Selberg's Sieve	$\pi(X) \leq \frac{(2+o(1))X}{\log X}$
Bombieri's Asymptotic Sieve	$\psi_k(X) = \frac{(k+o(1))X}{\log X}$, $k \geq 2$
Asymptotic Sieve for Primes	$\pi(X) = \frac{(1+o(1))X}{\log X}$

Remark 2. The above table can be misleading. The more advanced form of Brun's sieve does better, getting $\pi(X) \ll \frac{X}{\log X}$ with a weaker constant than Selberg's sieve. Also, the Asymptotic Sieve for Primes, while it successfully counts the prime numbers, assumes a much stronger hypothesis than its conclusion (a form of the Prime Number Theorem on arithmetic progressions) because it requires Type II information. It is doubly unsuccessful in this case because it achieves a much inferior error term than it assumes.

2.1. **The Basics.** Sieve theory starts with the tension between what is easy to count: numbers in arithmetic progressions, and what we want to count: prime numbers.

For a given sequence of nonnegative reals $\{a_n\}$, which we can think of as the indicator sequence of a number-theoretically interesting sequence, and assume that the sequence is well-behaved in the sense that good estimates are available for the summatory functions

$$A(X) = \sum_{n\leq X} a_n$$

and

$$A_d(X) = \sum_{n\leq X, d|n} a_n,$$

asymptotically as $X \to \infty$, as long as d grows reasonably slowly compared to X. Usually the estimate takes the form

$$A_d(X) = g(d)A(X) + r_d(X)$$

where g is a multiplicative function and $r_d(X)$ is very small compared to $A(X)$.

Definition 3. A Type I estimate is a bound of the form

$$\sum_{d\leq D} |r_d(X)| \ll_C A(X)\log^{-C} X$$

which holds for every C. The function $D = D(X)$ is the level of distribution.

What we want to count is the amount of $\{a_n\}$ supported on the primes, that is,

$$S(X) = \sum_{p\leq X} a_p$$

using only Type I estimates. Henceforth the variables p, q are understood to range only through primes. Later on, it will be fruitful to consider as a proxy

$$S(X) = \sum_{n\leq X} a_n\Lambda(n)$$

where $\Lambda(n)$ is the von Mangoldt function, but from the perspective of the basic combinatorial sieve the first definition is more natural.

Our expectations cannot be too high; as we will see in Section 2.4, even with exact closed forms for $A(X)$ and $A_d(X)$ we can only estimate $S(X)$ to within a factor of two at best. We will concern ourselves only with deriving upper bounds for $S(X)$; there are many methods to convert upper bounds to lower bounds and vice versa, e.g. the Buchstab identities .

The idea is to take as small a linear combination of $A_d(X)$ as possible such that each a_n appears with a nonnegative coefficient and each a_p has coefficient at least one. It is helpful to add an extra parameter D, which we call the level of distribution, and estimate

$$S(D, X) = \sum_{D<p\leq X} a_p,$$

so then we can use the small primes $p \leq D$ to sieve. For example, if D is very small compared to X, we can bound

$$S(D, X) \leq A(X) - \sum_{p\leq D} A_p(X) + \sum_{p,q\leq D} A_{pq}(X) - \sum_{p,q,r\leq D} A_{pqr}(X) + \cdots$$

using the inclusion-exclusion principle. This bound is only good up to about $D \approx \log X$, just because so many error terms $r_d(X)$ accumulate in this expression when we let $d = p_1p_2 \dots p_k$ for any choice of distinct primes $p_i \leq D$.

How bad is this? Let's use it to bound the total number of primes, so $a_n = 1$ for all n. We have $A_d(X) = Xd^{-1} + O(1)$, so

$$\begin{aligned} S(D, X) &\leq A(X) - \sum_{p\leq D} A_p(X) + \sum_{p,q\leq D} A_{pq}(X) - \cdots \\ &= X\Big(1 - \sum_{p\leq D} p^{-1} + \sum_{p,q} p^{-1}q^{-1} - \cdots\Big) + O(2^{\pi(D)}), \end{aligned}$$

where $\pi(D)$ is the number of primes up through D. The first term has a nice Euler product, so we get

$$S(D, X) \leq X\prod_{p\leq D}(1 - p^{-1}) + O(2^{\pi(D)}).$$

Using elementary methods, Mertens was able to estimate the product above as $O((\log D)^{-1})$, and so we have the bound

$$\pi(X) \ll X(\log D)^{-1} + 2^{\pi(D)}.$$

The error term grows extremely quickly in D and we can only pick $D \approx \log X$ before it overwhelms the main term.

Proposition 4. *If π is the prime counting function, then*

$$\pi(X) \ll \frac{X}{\log\log X}. \tag{2.1}$$

This is very far from the Prime Number Theorem but still nontrivial.

2.2. **The Brun Sieve.** Viggo Brun was the first person to take the idea of the sieve as a competitive method for analytic number theory; the conventional wisdom was that bounds like (2.1) are the best that can be done. There are at least two sieves in the literature referred to as the Brun sieve; we will develop the simpler one, usually called Brun's pure sieve. Brun's sieve is notable for providing the first nontrivial bounds on the number of twin primes.

We return to the inclusion-exclusion formula

$$S(D, X) \leq A(X) - \sum_{p \leq D} A_p(X) + \sum_{p,q \leq D} A_{pq}(X) - \sum_{p,q,r \leq D} A_{pqr}(X) + \cdots,$$

which we can rewrite conveniently as

$$S(D, X) \leq \sum_{d | P_D} \mu(d) A_d(X)$$

where μ is the usual Möbius function, and P_D is the product of primes $p \leq D$. But instead of expanding out the entire inclusion-exclusion formula, we truncate it at some stage k and throw out all the terms with more than k prime factors. In general if we truncate at a step with k odd, we are under-counting, and if we truncate with k even, we are over-counting; in combinatorial literature these are known as Bonferroni's inequalities. The upshot is that the smallest terms, which also happen to be the most numerous, are thrown out. In this particular instance we only care for upper bounds for $S(D, X)$, so we pick k even.

Write $\omega(d)$ to be the number of (distinct) prime factors of d, so that by Bonferroni's inequality,

$$\begin{aligned} S(D, X) &\leq \sum_{d | P_D, \omega(d) \leq k} \mu(d) A_d(X) \\ &= X \sum_{d | P_D, \omega(d) \leq k} \frac{\mu(d)}{d} + O(D^k), \end{aligned}$$

happily noting that an error term that was once exponential in D is now polynomial. Now we add back in all the main terms with $\omega(d) > k$ that we just threw out, with the exception that they no longer have $O(1)$ error terms attached. This to get a nicer sum while adding something tiny:

$$S(D, X) \leq X \sum_{d | P_D} \frac{\mu(d)}{d} + O\Big(X \sum_{d | P_D, \omega(d) = k+1} \frac{1}{d}\Big) + O(D^k).$$

Note that the error bound only needs the terms with $\omega(d) = k+1$ exactly, by Bonferonni's inequalities again. The main term now has a nice Euler product form,

$$\sum_{d|P_D} \frac{\mu(d)}{d} = \prod_{p\le D} \Big(1 - \frac{1}{p}\Big),$$

and the first error term can be bounded by Rankin's trick, for a parameter $t > 1$ to be optimized later:

$$\begin{aligned} \sum_{d|P_D, \omega(d)=k+1} \frac{1}{d} &= t^{-(k+1)} \sum_{d|P_D,\omega(d)=k+1} \frac{t^{k+1}}{d} \\ &\le t^{-(k+1)} \sum_{d|P_D} \frac{t^{\omega(d)}}{d} \\ &= t^{-(k+1)} \prod_{p\le D} \Big(1 + \frac{t}{p}\Big). \end{aligned}$$

Putting this together we get a bound of the form

$$S(D,X) \le X \prod_{p\le D} \Big(1 - \frac{1}{p}\Big) + O\Big(Xt^{-(k+1)} \prod_{p\le D} \Big(1 + \frac{t}{p}\Big)\Big) + O(D^k). \tag{2.2}$$

Finally, to estimate the leftover products, we have

$$\begin{aligned} \log\Big(\prod_{p\le D} \Big(1 + \frac{t}{p}\Big)\Big) &= \sum_{p\le D} \frac{t}{p} + O(1) \\ &= t \log\log D + O(1) \end{aligned}$$

by an elementary estimate of Mertens. Thus, the bound (2.2) reduces to

$$\pi(X) \ll X(\log D)^{-1} + Xt^{-(k+1)}(\log D)^t + D^k,$$

and it remains to optimize the values of D, t, k as functions of X subject to the mild conditions that $D < X$ and k is an even natural number. Close to optimal values turn out to be

$$\begin{cases} D &= \exp\Big(\frac{C_1 \log X}{\log\log X}\Big) \\ t &= C_2 \\ k &= C_3 \log\log X \end{cases}$$

with some suitable constants C_i independent of X.

Proposition 5. *If π is the prime counting function, then*

$$\pi(X) \ll \frac{X \log\log X}{\log X}.$$

This is already close to the Prime Number Theorem. It is possible to optimize Brun's argument further by inhomogeneously truncating the inclusion-exclusion, cutting off terms $p_1 p_2 \cdots p_k$ depending not only on the number k of them but by the sizes of p_i as well – this sieve is also due to Brun and leads to a correct order of magnitude bound on the prime counting function.

2.3. **The Selberg Sieve.** The next evolution in sieve theory after Brun was the Selberg "quadratic" or Λ^2 sieve, which does away with the combinatorially motivated constraint that the coefficients of $A_d(X)$ should be in $\{0, \pm 1\}$ (such sieves are naturally called combinatorial sieves).

Without this constraint we are left with a pure optimization problem: pick weights $\rho_d \in \mathbb{R}$ for $A_d(X)$ so that for any $D < n \leq X$,

$$\sum_{d|n} \rho_d \geq 0,$$

and $\rho_1 = 1$, which together imply

$$S(D, X) \leq \sum_{d \leq D} \rho_d A_d(X). \tag{2.3}$$

We can think of $\{\rho_d\}$ as the optimal "projection" of the Möbius function onto the space of functions supported on $[1, D]$.

Selberg's insight was to guarantee the nonnegativity constraint by setting

$$\sum_{d|n} \rho_d = \Big(\sum_{d|n} \lambda_d \Big)^2$$

where λ_d are arbitrary real parameters with $\lambda_1 = 1$. In this situation, we can solve for $\rho_d = \sum_{[d_1,d_2]=d} \lambda_{d_1} \lambda_{d_2}$, where $[d_1, d_2]$ is the lcm of d_1, d_2. Now the optimal choice of λ_d is exhibited by simple Cauchy-Schwarz inequality.

Again, we begin with our example $a_n = 1$. Expanding out (2.3) with our choice of ρ_d,

$$\begin{aligned} S(D, X) &\leq \sum \sum_{d_1, d_2 \leq D} \lambda_{d_1} \lambda_{d_2} \Big(\frac{X}{[d_1, d_2]} + O(1) \Big) \\ &\leq X \sum \sum_{d_1, d_2 \leq D} \frac{1}{[d_1, d_2]} \lambda_{d_1} \lambda_{d_2} + O\Big(\Big(\sum_{d \leq D} |\lambda_d| \Big)^2 \Big). \end{aligned}$$

Thus it remains to minimize the quadratic form

$$Q(\lambda) = \sum \sum_{d_1, d_2 \leq D} \frac{1}{[d_1, d_2]} \lambda_{d_1} \lambda_{d_2}$$

subject to the constraint $\lambda_1 = 1$. Now this quadratic form can be diagonalized by repeatedly completing the square:

$$\begin{aligned} Q(\lambda) &= \sum \sum_{d_1, d_2 \leq D} \gcd(d_1, d_2) \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} \\ &= \sum_{g \leq D} \varphi(g) \sum_{g|d_1} \sum_{g|d_2} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} \\ &= \sum_{g \leq D} \varphi(g) \Big(\sum_{g|d, d \leq D} \frac{\lambda_d}{d} \Big)^2, \end{aligned}$$

where φ is the Euler totient function. Let

$$x(g) = \sum_{g|d, d \leq D} \frac{\lambda_d}{d},$$

so that we have essentially a free optimization of Q, subject to the single constraint

$$\lambda_1 = \sum_{g \leq D} \mu(g)x(g) = 1.$$

By Cauchy-Schwarz, the minimum is

$$\begin{aligned} Q(\lambda) &= \sum_{g \leq D} \varphi(g)x(g)^2 \\ &\geq \frac{(\sum_{g \leq D} \mu(g)x(g))^2}{(\sum_{g \leq D} \mu(g)^2 \varphi(g)^{-1})} \\ &= \Big(\sum_{g \leq D} \frac{\mu(g)^2}{\varphi(g)}\Big)^{-1}, \end{aligned}$$

the minimum easily achieved using the equality condition of Cauchy-Schwarz. The inner sum is supported on squarefree numbers, and can be estimated by expanding as a geometric series and multiplicativity

$$\frac{1}{\varphi(p)} = \frac{1}{p} + \frac{1}{p^2} + \cdots,$$

so that

$$\begin{aligned} \sum_{g \leq D} \frac{\mu(g)^2}{\varphi(g)} &\geq \sum_{n \leq D} \frac{1}{n}, \\ &= \log D + O(1) \end{aligned}$$

and so we are left with

$$S(D, X) \leq (1 + o(1))X(\log D)^{-1} + O\Big(\Big(\sum_{d \leq D} |\lambda_d|\Big)^2\Big).$$

The last thing to do is to figure out what the optimal values of λ_d were; it's not difficult to see that they have magnitude at most 1, so the error term is $O(D^2)$. Picking $D = X^{1/2-o(1)}$ is then optimal.

Proposition 6. *If $\pi(X)$ is the prime counting function, then*

$$\pi(X) \leq (2 + o(1))\frac{X}{\log X}.$$

Not only do we get the correct order of magnitude, the bound is off by exactly a factor of two! This is the parity problem at work.

Although the Selberg sieve is permanently handicapped by the parity problem, it achieves results of great uniformity. The most famous application is the Brun-Titchmarsh Theorem.

Theorem 7. *(Brun-Titchmarsh) If $\pi(X; q, a)$ is the number of primes $p \leq X$ satisfying $p \equiv a \pmod{q}$, then*

$$\pi(X; q, a) < \frac{(2 + o(1))X}{\varphi(q) \log(X/q)},$$

uniformly for all $q < X$.

This theorem was proved by Titchmarsh using the Brun sieve for a much weaker constant in place of 2, and refined to the above form by van Lint and Richert [13] using the Selberg sieve. For comparison, analytic methods involving L-functions can prove the Siegel-Walfisz theorem, the exact asymptotic for $\pi(X; q, a)$ with a much better error term, but it only holds up to $q \ll (\log X)^C$ for fixed C.

2.4. **The Parity Problem.** Much more can be said about sieve theory relying only on Type I estimates of the form

$$A_d(X) = g(d)A(X) + r_d(X).$$

Commonly, sieves are distinguished by two parameters: the level of distribution D they require, and the "sieve dimension" κ they are best suited for. The sieve dimension measures the average number of residue classes $\bmod p$ we are sieving out, defined as the constant κ (which almost always exists) for which

$$\sum_{p \le X} g(p) = \kappa \log\log X + O(1),$$

noting that the standard case $g(p) = 1/p$ is of dimension 1. Rosser's linear sieve, or beta sieve, provides bounds of the correct order of magnitude in the case of sieve dimension 1.

Sieves with finite dimension are known as small sieves. In contrast, the large sieve is a family of methods for dealing with the case that $g(p)$ is much larger than $1/p$, usually more like a constant; we will see an application of the large sieve inequality later on in Section 3.3. Selberg's sieve as presented above is unique in its generality, competitive with both small and large sieves.

All of these sieves cannot get asymptotic results because of the so-called parity problem, which tells us that Type I estimates alone cannot distinguish between primes and almost-primes, the products of two prime factors.

Remark 8. The exact statement is more delicate than this, as we can see from the Selberg sieve inequality

$$\pi(X) \ll (2 + o(1))X(\log X)^{-1}.$$

But there are

$$\pi_2(X) \approx \frac{X \log\log X}{\log X}$$

2-almost primes up to X, so most of them are eliminated by the Selberg sieve above. A correct statement requires either throwing out almost primes with a small prime factor or weighting the almost-primes by Λ_2, the second von Mangoldt function

$$\Lambda_2(n) = \sum_{d|n} \mu(d)(\log \frac{n}{d})^2.$$

Here is an explicit instance of the parity problem which we steal from the excellent exposition of Ford [3]. Let $\lambda(n)$ be the Liouville function $\lambda(n) = (-1)^{\Omega(n)}$, where $\Omega(n)$ is the number of prime factors of n, counting multiplicity. Define $a_n = 1 + \lambda(n)$, so that in particular $a_p = 0$ on all primes p, and $S(X) = 0$.

On the other hand, with $A(X) = X$ and $g(d) = 1/d$, the remainder terms $r_d(X)$ are very small because numbers in any given arithmetic progression have even or odd number of factors with equal probability. With the Riemann Hypothesis we can show $r_d(X) = O(\sqrt{X/d}\log(X/d))$, which is much smaller than needed for classical sieve methods. But

$a_n = 1$ also has $A(X) = X$ and $g(d) = 1/d$, while $S(X) = \pi(X)$. Hence, it is impossible to distinguish with only this information between $a_n = 1$ and $a_n = 1 + \lambda(n)$, one of which contains all the primes and the other none of them.

One of the consequences of the parity problem is that these sieves have a hard time proving the existence of primes in any given sequence – the numbers they found could equally well be almost-primes. No nontrivial information about the number of primes in a sequence a_n can be deduced from Type I information alone.

2.4.1. *Perspective From L-functions.* There is something very special going on with the parity problem – the Selberg sieve is fully capable of distinguishing primes from products pqr of three primes, for example. In fact, from the sequence $a_n = 1 + \lambda(n)$ above we can see that the parity problem comes from the square root cancellation of $\mu(n)$ (which has basically the same summatory function as $\lambda(n)$).

Because we are not prepared to introduce the whole subject of L-functions in detail, this section will work purely heuristically.

Dirichlet L-functions provide a general machinery for understanding growth rates of arithmetic functions; in particular, the summatory function of $\mu(n)$ can be related to a certain contour integral of the inverse of the Riemann zeta function

$$\zeta(s)^{-1} = \sum_{n \geq 1} \frac{\mu(n)}{n^s}.$$

The Riemann Hypothesis predicts that $\zeta(s)$ has no zeroes s with $\Re(s) > \frac{1}{2}$, which implies that $\zeta(s)^{-1}$ has no poles past this point. Roughly speaking, a pole of $\zeta(s)^{-1}$ at ρ contributes to $\sum_{n \leq X} \mu(n)$ a term of the form

$$\frac{1}{\zeta'(\rho)} \frac{X^\rho}{\rho},$$

which has order $X^{\mathrm{Re}(\rho)}$. By general principles the asymptotics of this summatory function can be deduced from the set of zeroes in a bounded region, of which the Riemann zeta function has only finitely many. Thus to assume the Riemann Hypothesis proves

$$\sum_{n \leq X} \mu(n) = O(X^{1/2} \log X).$$

Instead of $\mu(n)$, let ϵ_3 be a primitive cube root of unity and consider the function

$$\mu_3(n) = \begin{cases} \epsilon_3^\alpha & n = p_1 p_2 \cdots p_\alpha \\ 0 & n \text{ squarefree} \end{cases}.$$

If $\sum_{n \leq X} \mu_3(n)$ canceled at all significantly, then by a similar construction as in the previous section, we would have "mod 3" problem and expect all sieves to be off by at least a factor of 3. This is not the case; $\mu_3(n)$ cancels much more poorly than $\mu(n)$. We state a weaker form of the result of Selberg and Delange for simplicity.

Theorem 9. *If $\mu_3(n)$ is as above, then there exists a constant $C \neq 0$ for which*

$$\sum_{n \leq X} \mu_3(n) = \frac{(C + o(1))X}{(\log X)^{\epsilon_3}}$$

infinitely often.

Here is a heuristic (from communications with Terry Tao and Zeb Brady). Write $\epsilon_3 = e^{\frac{2\pi i}{3}}$. The L-function of $\mu_3(n)$ has an Euler product,

$$\begin{aligned} L(s) &= \sum_{n\geq 1} \frac{\mu_3(n)}{n^s} \\ &= \prod_p (1 + \frac{\epsilon_3}{p^s}) \\ &\approx \prod_p (1 + p^{-s})^{\epsilon_3} \\ &\approx \zeta(s)^{\epsilon_3}. \end{aligned}$$

Now complex exponentiation is multi-valued and there is no way to analytically continue this function past $\mathrm{Re}(s) = 1$; we must pick a branch cut past this, and the pole at $s = 1$ can be checked to contribute an oscillating main term that looks like $CX/(\log X)^{\epsilon_3}$.

The same can be said for any modulus except 2 in place of 3 – the Möbius function has significant cancellation in comparison and is the reason there is a parity problem.

2.4.2. *The Way Forward.* Since the parity problem was formulated, one of the major goals of sieve theory has been to prove that the parity problem is the *only* obstruction to asymptotic sieves, sieves which give asymptotically correct bounds. There are two ways one can formulate this goal.

Bombieri's way was to construct an asymptotic sieve using only Type I information which correctly counts almost primes. By relaxing the question to counting

$$S_2(X) = \sum_{n\leq X} a_n \Lambda_2(n)$$

the parity problem goes away and sieve methods give asymptotic bounds correctly. We will describe Bombieri's Asymptotic Sieve in the next section.

Friedlander and Iwaniec were the first to break the parity problem altogether by injecting Type II information, creating their so-called Asymptotic Sieve for Primes. This sieve successfully counts prime values of $x^2 + y^4$ asymptotically, but requires a second condition, the bilinear forms condition, which is very difficult to prove in practice. Roughly speaking the bilinear forms condition asks that a_n never looks like $\mu(n)$ (or its cousin $\lambda(n)$) on any arithmetic progression.

2.5. **Bombieri's Asymptotic Sieve.** We first define the generalized von Mangoldt function to give the correct weighting on almost-primes.

Definition 10. The generalized von Mangoldt function $\Lambda_k(n)$ is defined by

$$\Lambda_k(n) = \sum_{d|n} \mu(d) \Big(\log \frac{n}{d} \Big)^k.$$

Of course the usual von Mangoldt function is just $\Lambda = \Lambda_1$. In fact Bombieri's actual result generalized this further to arbitrary Dirichlet convolutions of the Λ_k above, but we will make do with Λ_k.

Note that $\Lambda_k(n)$ satisfies the convolution recurrence

$$\Lambda_{k+1}(n) = \Lambda_k(n)\log n + \sum_{d|n} \Lambda(d)\Lambda_k(\frac{n}{d}),$$

so by induction Λ_k is supported on integers n with at most k distinct prime factors.

Bombieri's asymptotic sieve gives asymptotic results, allowing us to estimate

$$S_k(X) = \sum_{n\leq X} a_n \Lambda_k(n)$$

for any $k \geq 2$.

Assume for simplicity sieve dimension 1, so that $g(p) \approx 1/p$ on average. All that is required is a fairly strong Type I estimate of the form

$$\begin{aligned} A_d(X) &= g(d)A(X) + r_d(X) \\ \sum_{d\leq D} |r_d(X)| &\ll_{C,D} \frac{A(X)}{(\log X)^C}, \end{aligned} \tag{2.4}$$

for every level of distribution $D = X^{1-\varepsilon}$, $\varepsilon > 0$. For a simple exposition of Bombieri's Asymptotic Sieve, see *Opera de Cribro*, Chapter 3.

Theorem 11. *(Bombieri's Asymptotic Sieve) Assuming* (2.4)*,*

$$S_k(X) = (1 + o(1))kHA(X)(\log X)^{k-1},$$

where

$$H = \prod_p (1 - g(p))(1 - 1/p)^{-1}.$$

The product for H converges to a nonzero constant iff g has sieve dimension 1. Without the parity problem, we would expect this asymptotic to be true for $S(X) = S_1(X)$ as well.

2.5.1. *The Selberg Symmetry Formula.* In this section we prove the spiritual ancestor of Bombieri's Asymptotic Sieve, namely the Symmetry Formula of Selberg, which leads to the Erdős-Selberg elementary proof of the Prime Number Theorem. We follow the proof given by Balady [1].

Theorem 12. *Let $k \geq 2$, and $\Lambda_k(n)$ be the k-th von Mangoldt function. Then*

$$\sum_{n\leq X} \Lambda_k(n) = kX \log^{k-1} X + O(X \log^{k-2} X).$$

We first prove this for the case $k = 2$. Recall that $\Lambda_2(n)$ satisfies the identity

$$\Lambda_2(n) = \Lambda(n)\log n + \sum_{d|n} \Lambda(d)\Lambda(\frac{n}{d}),$$

so it weights primes p by $(\log p)^2$ and almost primes pq by $2\log p \log q$. By the Prime Number Theorem, the sum of $\Lambda(n)\log n$ should be asymptotic to $X \log X$, so the Symmetry Formula implies that $\Lambda_2(n)$ counts primes and almost primes in approximately equal measure.

The key lemma is an elementary cancellation estimate for sums of the Möbius function.

Lemma 13. *If $\mu(n)$ is the Möbius function, then*

$$\begin{aligned}\sum_{n\le X}\frac{\mu(n)}{n} &= O(1),\\ \sum_{n\le X}\frac{\mu(n)}{n}\log\frac{X}{n} &= O(1),\\ \sum_{n\le X}\frac{\mu(n)}{n}\log^2\frac{X}{n} &= 2\log X + O(1).\end{aligned}$$

Proof. The trick is to use the following transformation of the Möbius inversion formula which Balady [1] uses but never writes down. Let f, g be functions defined on $[1,\infty)$, for which

$$f(x) = \sum_{n\le x} g(x/n).$$

Then

$$\begin{aligned}\sum_{m\le x}\mu(m)f(\frac{x}{m}) &= \sum_{m\le x}\mu(m)\sum_{n\le x/m} g(x/mn)\\ &= \sum_{k\le x} g(x/k)\sum_{m|k}\mu(m)\\ &= g(x).\end{aligned}$$

We pick $g(1)\equiv 1$ first, so that $f(x) = \lfloor x\rfloor$. The identity becomes

$$\begin{aligned}\sum_{n\le X}\mu(n)\frac{X}{n} + O(X) &= 1\\ \sum_{n\le X}\frac{\mu(n)}{n} &= O(1),\end{aligned}$$

which is the first bound.

Next pick $g(x) = x$, so that

$$\begin{aligned}f(x) &= \sum_{n\le x}\frac{x}{n}\\ &= x\log x + C_1 x + O(1)\end{aligned}$$

by approximating the integral, where C_1 is Euler's constant, though its exact value is irrelevant for this purpose. After Möbius inversion, we get

$$\begin{aligned}\sum_{n\le X}\mu(n)\Big(\frac{X}{n}\log\frac{X}{n} + C_1\frac{X}{n} + O(1)\Big) &= X\\ X\sum_{n\le X}\frac{\mu(n)}{n}\log\frac{X}{n} &= O(X)\\ \sum_{n\le X}\frac{\mu(n)}{n}\log\frac{X}{n} &= O(1)\end{aligned}$$

using the previous bound.

Finally, pick $g(x) = x \log x$, so that

$$\begin{aligned} f(x) &= \sum_{n \le x} \frac{x}{n} \log \frac{x}{n} \\ &= \frac{x}{2} \log^2 x + C_2 x \log x + C_3 x + O(\log x) \end{aligned}$$

again by approximating the integral. Again C_2, C_3 are irrelevant constants. Möbius inversion gives

$$\begin{aligned} \sum_{n \le X} \mu(n) \Big(\frac{X}{2n} \log^2 \frac{X}{n} + C_2 \frac{X}{n} \log \frac{X}{n} + C_3 \frac{X}{n} + O(\log \frac{X}{n}) \Big) &= X \log X \\ \frac{1}{2} \sum_{n \le X} \frac{\mu(n)}{n} \log^2 \frac{X}{n} &= \log X + O(1), \end{aligned}$$

combining all the previous bounds, and the fact that the sum of $\log \frac{X}{n}$ is $O(X)$. □

Now we can control sums of $\Lambda_2(n)$ by directly expanding the convolution.

Proof. (of Theorem 12.) We expand

$$\begin{aligned} \sum_{n \le X} \Lambda_2(n) &= \sum_{n \le X} \mu(n) \sum_{m \le X/n} \log^2 m \\ &= \sum_{n \le X} \mu(n) \Big(\frac{X}{n} \log^2 \frac{X}{n} + C_4 \frac{X}{n} \log \frac{X}{n} + C_5 \frac{X}{n} + O(\log^2 \frac{X}{n}) \Big) \\ &= 2 \log X + O(X), \end{aligned}$$

as desired, by combining all the estimates in Lemma 13. It is in fact straightforward to continue the inductive application of Möbius inversion and prove for each $k \ge 2$,

$$\begin{aligned} \sum_{n \le X} \frac{\mu(n)}{n} \log^k \frac{X}{n} &= k \log^{k-1} X + O(\log^{k-2} X) \\ \sum_{n \le X} \Lambda_k(n) &= kX \log^{k-1} X + O(X \log^{k-2} X). \end{aligned}$$

□

The parity problem prevents us from finding such a simple computation of the Möbius sum for $k = 1$, since if we had any estimate of the form

$$\sum_{n \le X} \frac{\mu(n)}{n} \log \frac{X}{n} = 1 + o(1)$$

the Prime Number Theorem would immediately follow.

2.5.2. *An Easy Misconception about the Generalized von Mangoldt Function.* Looking at the formula

$$\sum_{n \le X} \Lambda_k(n) = kX \log^{k-1} X + O(X \log^{k-2} X)$$

we might expect that just as in the $k = 2$ case, approximately $X \log^{k-1} X$ of the sum comes from the j-almost primes, for each $1 \le j \le k$. This is decidedly false in general, since it would break the parity barrier! For k odd this would give an essentially sieve-theoretic way

to show that there are more k-almost primes with an odd number of prime factors than with an even number.

For $k = 3$ we have

$$\Lambda_3(n) = \Lambda_2(n)\log n + \sum_{d|n} \Lambda_2(d)\Lambda(\frac{n}{d}),$$

so since $\Lambda_2(p) = \log^2 p$ and $\Lambda_2(pq) = 2\log p \log q$, we get

$$\begin{aligned} \Lambda_3(p) &= \log^3 p \\ \Lambda_3(pq) &= 3\log p \log q \log pq \\ \Lambda_3(pqr) &= 6\log p \log q \log r. \end{aligned}$$

But the sum of $\log p \log q \log pq$ over almost-primes $pq \le X$ is approximately

$$\frac{1}{2}\log X \sum_{n\le X} \Lambda_2(n) - \Lambda(n)\log n = \frac{1}{2}X\log^2 X + O(X\log X),$$

and so the 2-almost primes pq actually contribute $\frac{3}{2}X\log^2 X$ to the sum of $\Lambda_3(n)$, which is exactly half the weight of the sum, just as the parity problem predicts. The numbers p contribute $X\log^2 X$ to the sum, the numbers pq contribute $\frac{3}{2}X\log^2 X$, and the numbers pqr contribute just $\frac{1}{2}X\log^2 X$.

2.6. **The Asymptotic Sieve for Primes.** Friedlander and Iwaniec [7] finally broke the parity barrier by injecting a very strong second condition, which they called the bilinear forms condition. The key tool used in their paper is Vaughan's identity, a simple combinatorial identity for the von Mangoldt function that separates what might be called its wavelengths.

The goal of this section is to state the simplest formulation of Vaughan's identity that we will need for producing primes of the form $p^2 + Ny^2$. The identity [14] allows us to separate the main term, Type I error term, and Type II error term directly out of $S(X)$.

Lemma 14. *(Vaughan's identity) Choose integers $y, z \ge 1$. For any $n > z$ we have*

$$\Lambda(n) = \sum_{b|n, b\le y} \mu(b)\log\frac{n}{b} - \sum_{bc|n, b\le y, c\le z} \mu(b)\Lambda(c) + \sum_{bc|n, b>y, c>z} \mu(b)\Lambda(c),$$

and the right hand side is zero if $n \le z$.

Proof. We have

$$\begin{aligned} \Lambda(n) &= \sum_{b|n} \mu(b)\log\frac{n}{b} \\ &= \sum_{b|n, b\le y} \mu(b)\log\frac{n}{b} + \sum_{bc|n, b>y} \mu(b)\Lambda(c) \\ &= \sum_{b|n, b\le y} \mu(b)\log\frac{n}{b} + \sum_{bc|n, b>y, c>z} \mu(b)\Lambda(c) + \sum_{bc|n, b>y, c\le z} \mu(b)\Lambda(c) \\ &= \sum_{b|n, b\le y} \mu(b)\log\frac{n}{b} - \sum_{bc|n, b\le y, c\le z} \mu(b)\Lambda(c) + \sum_{bc|n, b>y, c>z} \mu(b)\Lambda(c), \end{aligned}$$

since for any fixed $c \le z < n$, the whole second sum over b is $\Lambda(c)\sum_{b|nc^{-1}} \mu(b) = 0$. If $n \le z$, then the first two sums cancel and the last is empty. □

From here, what needs to be done depends on the application. For example, Friedlander and Iwaniec designed their Asymptotic Sieve for Primes with the application of $x^2 + y^4$ in mind, and broke the last sum in Vaughan's identity further down into three sums depending on the ranges of b and c. Heath-Brown also used Vaughan's identity to break the parity barrier for primes of the form $x^3 + 2y^3$, but the exact calculations were different – in fact to the author's knowledge the exact formulation of the Asymptotic Sieve for Primes in Friedlander-Iwaniec [7] has only been applied to the single case of $x^2 + y^4$, despite the fact that the general method works in a variety of settings.

In general, the Asymptotic Sieve for Primes expects the first two terms, containing the "small oscillations," to produce the main term of the sieve using only Type I estimates, whilst the last term must be bounded more delicately in terms of a bilinear forms condition (see Section 3.5 below).

If

$$S(X) = \sum_{n \leq X} a_n \Lambda(n)$$

as before, then substituting Vaughan's identity this sum resolves as follows.

Lemma 15. *Suppose $y, z \geq 1$ and $X > yz$. Then,*

$$S(X) = S(z) + A(X; y, z) + B(X; y, z)$$

where

$$\begin{aligned} A(X; y, z) &= \sum_{b \leq y} \mu(b) \Big(\sum_{b|n, n \leq X} a_n \log \frac{n}{b} - \sum_{c \leq z} \Lambda(c) \sum_{bc|n, n \leq X} a_n \Big) \\ B(X; y, z) &= \sum_{bd \leq X, b > y} \mu(b) a_{bd} \Big(\sum_{c|d, c > z} \Lambda(c) \Big). \end{aligned}$$

We will allow y, z to remain indeterminate for now. The term $S(z)$ will be negligible, $A(X; y, z)$ will be the main term asymptotically after we prove the level of distribution, and bounding $B(X; y, z)$ reduces almost immediately into the bilinear forms condition.

3. Primes of the Form $p^2 + Ny^2$

Fix squarefree $N > 0$. Our goal is to count asymptotically the number of primes $q = p^2 + Ny^2$ where p varies through primes and y through all integers. Each q is counted with multiplicity the number of times it occurs as $p^2 + Ny^2$. Write $\Lambda(n)$ to be the von Mangoldt function. We will prove Theorem 1, and in particular that there are infinitely many primes of the form $p^2 + Ny^2$.

This result generalizes the theorem of Fouvry and Iwaniec [5] which provides the case $N = 1$. We follow their presentation closely, altering the computations where necessary to accommodate the general case. In fact, Fouvry and Iwaniec proved a more general theorem for the case $N = 1$, replacing $\Lambda(x)$ with any reasonable sequence of complex numbers λ_x. However, we chose to specialize to the case $\Lambda(x)$ as it leads to a number of simplifications in the ensuing calculations.

Friedlander and Iwaniec [7] provided a general sieve called the Asymptotic Sieve for Primes to count primes of the form $x^2 + y^4$ [8]. In this case applying that sieve is unnecessarily burdensome and provides a much poorer error term. We will be able to obtain an extremely

high level of distribution $X^{1-\varepsilon}$, avoiding many of the difficult computations required by them. Nevertheless our computations are closely related.

We begin with a sequence $\{a_n\}$, which for us is

$$a_n = \sum_{x^2+Ny^2=n,(x,Ny)=1} \Lambda(x).$$

We add the condition $(x, Ny) = 1$ to simplify many of the ensuing computations. We wish to write

$$S(X) = \sum_{n\leq X} a_n \Lambda(n)$$

in terms of the much easier sums

$$A_d(X) \quad = \quad \sum_{n\leq X, d|n} a_n.$$

We write $A(X) = A_1(X)$. Each $A_d(X)$ can be approximated by $g(d)A(X) + r_d(X)$, where $g(d)$ is the multiplicative function defined on prime powers as:

$$g(p^\alpha) \quad = \quad \begin{cases} \frac{1+\left(\frac{-N}{p}\right)}{p^\alpha} & podd, p \nmid N \\ \frac{1}{2} & p = 2, p \nmid N, \alpha = 1 \\ \frac{1-\chi_4(N)}{2^\alpha} & p = 2, p \nmid N, \alpha \geq 2 \\ 0 & p|N \end{cases}$$

where χ_4 is the nontrivial Dirichlet character mod 4. The remainder term $r_d(X)$ is relatively small.

Write $F \ll G$ if $F = O(G)$. Classical sieve theory tells us that a remainder term bound (or level of distribution bound)

$$\sum_{d\leq D} |r_d(X)| \ll A(X)(\log X)^{-A}, \tag{3.1}$$

with level of distribution D large enough is sufficient for estimating

$$S_2(X) = \sum_{n\leq X} a_n \Lambda_2(n),$$

summing a_n over 2-almost primes weighted by the second von Mangoldt function Λ_2. The well-known "parity problem" in sieve theory prohibits estimating $S(X)$ directly from only the level of distribution (3.1). Using Vaughan's identity, Fouvry and Iwaniec are able to establish such estimates and break the parity problem with an additional bilinear forms condition:

$$\sum_{M<m\leq 2M} \Big| \sum_{N<n\leq(1+\epsilon)N} \mu(n) a_{mn} \Big| \ll A(X)(\log X)^{-A}, \tag{3.2}$$

which guarantees that a_{mn} does not "conspire" with the Möbius function on average.

It is natural to divide the sieve computation into three steps. First, we apply Vaughan's identity to bound $S(X)$ in terms of $A(X)$ and the sums (3.1) and (3.2), and compute the main term. Then, we separately prove the two bounds (3.1) and (3.2).

At the heart of the remainder term bound (3.1) for $x^2 + Ny^2$ is a simple equidistribution result, namely that the solutions (ν, d) to $\nu^2 + N \equiv 0(d)$ are well-spaced in the sense that the fractions ν/d are far apart when d is restricted to a short interval $D < d \leq (1 + \delta)D$.

The case $N = 1$ was proved by Duke, Friedlander, and Iwaniec [2]; the generalization is not difficult once we first divide the fractions $\{\nu/d\}$ into a finite number of families (depending only on N). Such a strong well-spacing result gives a correspondingly strong large sieve inequality. In Section 3.4, we show how to deduce the remainder term bound from this large sieve inequality.

The bilinear forms condition requires writing a_{mn} as a sum over ideals I, J in the ring of integers of $\mathbb{Q}[\sqrt{-N}]$ such that $m = N(I)$, $n = N(J)$. After reformulating the sum in terms of ideals and conditioning on the ideal class group representative, the sum is essentially identical to the one Fouvry and Iwaniec treat. It requires a delicate application of the Cauchy-Schwarz inequality, reducing the inequality to a standard Siegel-Walfisz type bound on Möbius sums for $\mathbb{Q}[\sqrt{-N}]$.

3.1. **The Value of $A(X)$.** We first compute asymptotically the sum

$$\begin{aligned} A(X) &= \sum_{n\leq X} a_n \\ &= \sum_{x^2+Ny^2\leq X,(x,Ny)=1} \Lambda(x). \end{aligned}$$

This is elementary, depending only on the prime number theorem.

Lemma 16. *Let $A(X)$ be as above. Then,*

$$A(X) = \frac{\pi X}{4\sqrt{N}} + O\Big(X \exp\Big(\frac{-c(\log X)^{3/5}}{(\log\log X)^{1/5}}\Big)\Big)$$

for a positive constant c.

Proof. We compute:

$$\begin{aligned} A(X) &= \sum_{x\leq\sqrt{X},(x,N)=1} \Lambda(x) \sum_{(y,x)=1,Ny^2\leq X-x^2} 1 \\ &= \sum_{p\leq\sqrt{X},p\nmid N} \log p \sum_{p\nmid y,Ny^2\leq X-p^2} 1 + O(X^{3/4}\log X) \\ &= \sum_{p\leq\sqrt{X}} \log p\Big(\sqrt{\frac{X-p^2}{N}} + O\Big(\frac{\sqrt{X}}{p}\Big)\Big) + O(X^{3/4}\log X) \\ &= \frac{1}{\sqrt{N}} \sum_{x\leq\sqrt{X}} \Lambda(x)\sqrt{X-x^2} + O(X^{3/4}\log X). \end{aligned}$$

We were free to include and exclude the prime powers $p^\alpha, \alpha \geq 2$ at will. Also the finite set of primes $p|N$ fall into the error term. Let

$$E(X) = O\Big(X \exp\Big(\frac{-c(\log X)^{3/5}}{(\log\log X)^{1/5}}\Big)\Big)$$

be the best known error term on the prime number theorem, due to Ford [4]. Applying summation by parts to the main term, we get

$$\begin{aligned}\sum_{x\le\sqrt{X}}\Lambda(x)\sqrt{X-x^2} &= \sum_{x\le\sqrt{X}}\Lambda(x)\int_x^{\sqrt{X}}\frac{tdt}{\sqrt{X-t^2}}\\ &= \int_0^{\sqrt{X}}\sum_{x\le t}\Lambda(x)\frac{tdt}{\sqrt{X-t^2}}\\ &= \int_0^{\sqrt{X}}(t+O(E(t)))\frac{tdt}{\sqrt{X-t^2}}\\ &= \frac{\pi}{4}X+O(\sqrt{X}E(\sqrt{X})).\\ &= \frac{\pi}{4}X+O\Big(X\exp\Big(\frac{-c(\log X)^{3/5}}{(\log\log X)^{1/5}}\Big)\Big).\end{aligned}$$

The constant c in the last line is not necessarily the same as in $E(X)$. □

3.2. **Vaughan's Identity.** Using Vaughan's identity (Lemma 15), Iwaniec and Fouvry are able to split the sum $S(X)$ into three terms: the main term, a remainder term controlled by (3.1), and a bilinear term controlled by (3.2). We get

$$S(X) = S(z) + A(X;y,z) + B(X;y,z)$$

where

$$\begin{aligned}A(X;y,z) &= \sum_{b\le y}\mu(b)\Big(\sum_{b|n,n\le X}a_n\log\frac{n}{b}-\sum_{c\le z}\Lambda(c)\sum_{bc|n,n\le X}a_n\Big)\\ B(X;y,z) &= \sum_{bd\le X,b>y}\mu(b)a_{bd}\Big(\sum_{c|d,c>z}\Lambda(c)\Big).\end{aligned}$$

3.2.1. *Computation of the Main Term.* To treat $A(X;y,z)$, we need the level of distribution estimate (3.1) proved in Section 3.4.

Lemma 17. *If* (3.1) *holds and there exists* $\varepsilon>0$ *for which* $y\ll X^{1-\varepsilon}$, *then*

$$A(X;y,z) = \frac{\pi H_N X}{4\sqrt{N}} + O_A(X(\log X)^{-A})$$

for any $A>0$, *the implicit constant depending only on* A.

Proof. We first express $A(X;y,z)$ in terms of $A_d(X)$:

$$A(X;y,z) = \sum_{b\le y}\mu(b)\Big(A_b(X)\log X - A_b(X)\log b - \int_1^X A_b(t)\frac{dt}{t} - \sum_{c\le z}\Lambda(c)A_{bc}(X)\Big),$$

Now, we have the estimate $A_d(X) = g(d)A(X)+r_d(X)$, so we wish to approximate $A(X;y,z)$ by

$$M(X;y,z) = A(X)\sum_{b\le y}\mu(b)\Big(g(b)\log(X/b)-\sum_{c\le z}\Lambda(c)g(bc)\Big) - \Big(\int_1^X A(t)\frac{dt}{t}\Big)\sum_{b\le y}\mu(b)g(b).$$

Since $A(X)$ is approximately linear, integrating against dt/t does nothing except change constants in the error term:

$$M(X;y,z)=\frac{\pi X}{4\sqrt{N}}\Big(1+O\Big(\exp\Big(\frac{-c(\log X)^{3/5}}{(\log\log X)^{1/5}}\Big)\Big)\Big)\sum_{b\le y}\mu(b)\Big(g(b)\log(X/b)-g(b)-\sum_{c\le z}\Lambda(c)g(bc)\Big).$$

To deal with the sum over b, we first extend over all b and show that most of the sum vanishes:

$$\begin{aligned}\sum_{b\ge 1}\mu(b)g(b) &= \prod_p(1-g(p))\\ &= 0\end{aligned}$$

since for a positive proportion of primes, $g(p)=2p^{-1}$. Similarly,

$$\begin{aligned}\sum_{b\ge 1}\mu(b)\sum_{c\le z}\Lambda(c)g(bc) &= \sum_{c\le z}\Lambda(c)\sum_{b\ge 1}\mu(b)g(bc)\\ &= \sum_{c\le z}\Lambda(c)\prod_{p\nmid c}(1-g(p))\prod_{p\mid c}(g(c)-g(pc))\\ &= 0.\end{aligned}$$

For the last sum left, the following identity holds:

$$-\sum_{b\ge 1}\mu(b)g(b)\log b=\prod_p(1-g(p))(1-p^{-1})^{-1},$$

assuming only that

$$\sum_{p\le X}g(p)=\log\log X+C+O(\log^{-10}X)$$

for all X [7]. Note that the product on the right is exactly H_N, so we get

$$M(X;y,z)=\frac{\pi H_N X}{4\sqrt{N}}+O\Big(X\exp\Big(\frac{-c(\log X)^{3/5}}{(\log\log X)^{1/5}}\Big)\Big)+O\Big(X\sum_{b>y}\mu(b)\Big(g(b)\log(X/b)-g(b)-\sum_{c\le z}\Lambda(c)g(bc)\Big)\Big).$$

We assume that the second error term is $O((\log X)^{-A})$ for any A, for suitable choice of y [5].
It follows that

$$M(X;y,z)=\frac{\pi H_N X}{4\sqrt{N}}+O_A((\log X)^{-A})$$

as desired. It remains to handle the remainder term

$$\begin{aligned}R(X;y,z) &= A(X;y,z)-M(X;y,z)\\ &= \sum_{b\le y}\mu(b)\Big(r_b(X)\log(X/b)-\int_1^X r_b(t)\frac{dt}{t}-\sum_{c\le z}\Lambda(c)r_{bc}(X)\Big).\end{aligned}$$

In Section 3.4 we show that

$$\begin{aligned}R_D(X) &= \sum_{d\le D}|r_d(X)|\\ &\ll_\varepsilon X^{1-\varepsilon}\end{aligned}$$

for all $A \geq 1$, as long as $D \ll X^{1-3\varepsilon}$. It follows that

$$\begin{aligned} |R(X;y,z)| &\leq 2R_y(X)\log X + \int_1^X R_y(t)\frac{dt}{t} \\ &\ll_\varepsilon X^{1-\varepsilon}, \end{aligned}$$

as desired. □

3.3. **A Large Sieve Inequality.**

3.3.1. *Fractions Well-Spaced Mod 1.* From the reduction of level-of-distribution results to large-sieve type inequalities, we are led to consider, for a fixed integer $N > 0$, the roots of $\nu^2 + N \equiv 0(d)$ as d ranges through all positive integers for which $-N$ has a square root mod d. In fact any well-spacing of these fractions ν/d better than the trivial $1/d^2$ spacing of distinct rationals will give a nontrivial large sieve bound. Iwaniec and Friedlander were able to show an almost perfect well-spacing of ν/d in the case $N = 1$, in fact that they are separated by at least about $1/4d$ when $d \in [(1-\delta)D, D]$ lies in a short interval.

We will try to show the general case, getting a slightly weaker bound. The first step is to associate to each ν/d a solution (x, y, T) to

$$x^2 + Ny^2 = Td$$

satisfying $(x, y) = 1$, $x \equiv \nu y(d)$, and $0 < x \leq |y|\sqrt{N}$. We call such a triple (x, y, T) a standard solution for ν/d. The relationship between well-spacedness of ν/d and these standard solutions is the following lemma.

Lemma 18. *If (x, y, T) is a standard solution for ν/d, then*

$$\frac{\nu}{d} \equiv \frac{x}{yd} - \frac{T\bar{x}}{y} \pmod 1,$$

where $\bar{x}$ is the multiplicative inverse of x modulo y. In particular ν/d is within $\sqrt{N}/d$ of a fraction with denominator $y \leq \sqrt{Td/N}$.

Proof. The identity is just combining

$$\begin{aligned} \nu y &\equiv x \pmod d \\ Td\bar{x} &\equiv x \pmod y, \end{aligned}$$

to find the residue class of νy modulo dy. Note that $(x, y) = 1$ implies $(d, y) = 1$.

The first fraction x/yd is very small: since $0 < x \leq |y|\sqrt{N}$, it is at most $\sqrt{N}/d$. On the other hand the second fraction has denominator at most y and $Ny^2 \leq Td$. □

Using Lemma (18), it is possible to partition the fractions ν/d into a small number of families, each of which is well-spaced.

Lemma 19. *Suppose that for every ν/d satisfying $\nu^2 + N \equiv 0(d)$ with $(d, N) = 1$ there exists a standard solution (x, y, T) for ν/d for which*

$$T < T_{max}.$$

Then there exists a constant $C > 0$ such that for every dyadic interval $I = [D/2, D]$ of moduli d, the set of fractions $\{\nu/d, d \in I\}$ can be partitioned into at most $C\max(T_{max}^2, T_{max}\sqrt{N})$

classes and within each class any two fractions $\nu/d, \nu'/d'$ satisfy

$$\left|\frac{\nu}{d} - \frac{\nu'}{d'}\right| > \frac{\sqrt{N}}{D}.$$

Proof. Every fraction ν/d is within $\sqrt{N}/d$ of a fraction with denominator at most $d_{max} = \sqrt{T_{max}D/N}$, say $f(\nu/d)$. As before, the fibers $f^{-1}(a/b)$ have size bounded by $T_{max}\sqrt{N}$.

Now, sort the fractions up to denominator d_{max} mod 1; each consecutive pair has difference at least $1/d_{max}^2$, so a pair of two such fractions that are k apart are at least k/d_{max}^2 apart in value. If we choose k so that

$$\frac{k}{d_{max}^2} > \frac{4\sqrt{N}}{D},$$

then we can partition our set of fractions into $O(k)$ classes first, so that within each class the fractions either correspond to the same x/y or else are at least $2\sqrt{N}/D$ apart. Split each of these classes further into $O(T_{max}\sqrt{N})$ classes so no two fractions correspond to the same x/y, and we get the desired result with $O(kT_{max}\sqrt{N})$ classes. Choosing $k = \max(1, 4T_{max}/\sqrt{N})$, we get the exact bound $O(T_{max}^2)$ on the number of classes. □

With this lemma in hand, we can begin to prove the large sieve inequality we need. All that is needed now is the construction of standard solutions with bounded T.

3.3.2. *Construction of Standard Solutions.* Numbers representable as $Td = x^2 + Ny^2$ correspond to norms of principal ideals in the ring of integers in $\mathbb{Q}[\sqrt{-N}]$. Write $K = \mathbb{Q}[\sqrt{-N}]$ and $\mathcal{O} = \mathcal{O}_K$ for its ring of integers.

Lemma 20. *There exists a constant $T_N > 0$ depending only on N such that for each $(d, 2N) = 1$ for which $-N$ has square roots* mod d*, and each solution ν to $\nu^2 + N \equiv 0(d)$, there exists a corresponding standard solution (x, y, T) for which $T \leq T_N$.*

Proof. Because $(d, 2N) = 1$ we don't need to deal with ramified primes. Let G be the ideal class group of $\mathcal{O}$, and pick a set of generators as a finite abelian group $\{[I_1], [I_2], \ldots, [I_m]\}$, so that $[I_i]$ has order ℓ_i and

$$G \simeq \bigoplus_{i \leq m} \mathbb{Z}/\ell_i\mathbb{Z},$$

where the generator in each component $\mathbb{Z}/\ell_i\mathbb{Z}$ is $[I_i]$. Let I_i be arbitrary prime representatives of $[I_i]$ (there are infinitely many primes in any ideal class).

Now, for a given d, a standard solution $x^2 + Ny^2 = Td$ is just a principal ideal $(x + y\sqrt{-N})$ with norm a multiple of d. We can factor $d = \prod q_i^{e_i} \bar{q}_i^{e_i}$ over $\mathcal{O}$. Every integer prime factor splits since $-N$ has square roots mod d and is coprime to d. A standard solution then corresponds to an integral multiple of

$$z = \prod_i p_i^{e_i},$$

where p_i is either q_i or $\bar{q}_i$. Regardless of the choice of p_i, it is possible to pick $J_i \in \{I_i, \bar{I}_i\}$ so that $J_i \neq \bar{p}_{i'}$ for any i, i', so that $\{J_i\}_{i \leq m}$ is a set of generators for G. It follows that there is a unique product

$$z \prod_{i \leq m} J_i^{\alpha_i} = x + y\sqrt{-N}$$

which is principal, where $0 \le \alpha_i < \ell_i$. We need to verify that (x, y) are coprime, that distinct choices of z correspond to distinct $\nu \equiv x/y \pmod d$, and that all possible ν are attained by some such $x + y\sqrt{-N}$.

We chose J_i to be nonprincipal prime ideals which are not conjugate to any of the prime factors of z, and not conjugate to each other. Therefore, $x + y\sqrt{-N}$ has no nontrivial real integral factors and $(x, y) = 1$.

If two distinct z, z' correspond to the same solution $\nu^2 + N \equiv 0(d)$, then the corresponding standard solutions satisfy $x + y\sqrt{-N} \equiv u(x' + y'\sqrt{-N}) \pmod d$ for a real integer u coprime to d. But factoring d and using the Chinese Remainder Theorem, this means that the same choice of conjugate of q_i divides both z and z' for every i, and so z, z' are the same.

There are $2^{\omega(d)}$ distinct solutions $\nu^2 + N \equiv 0(d)$, where $\omega(d)$ counts the number of distinct (real) prime factors of d, and there are $2^{\omega(d)}$ choices of z. It follows that they are in bijection, as desired.

In general, elements $x + y\sqrt{-N}$ of $\mathcal{O}_N$ may have x, y half-integers; since d is odd we may multiply by 2 to guarantee $x, y \in \mathbb{Z}$.

Finally, to make them all into standard solutions, we need to apply the transformation $(x, y) \mapsto (Ny, x)$ to those (x, y) with $|x| > |y|\sqrt{N}$. If $y < 0$ we multiply this by -1. The only common factors that can be introduced are ramified primes over $\mathcal{O}_N$, so we are free to divide them out from (Ny, x) until we have a standard solution. Thus, we have constructed standard solutions satisfying

$$T \le 4N(\prod_{i \le m} J_i^{\ell_i})N,$$

and this is the value of T_N we take. □

3.3.3. *The Large Sieve Inequality.* We have shown that when $D/2 < d \le D$ and $(d, 2N) = 1$, the maximum value of T_{max} for any given d is T_N, independent of D. Now we need the large sieve inequality of Montgomery and Vaughan [11]. We say that a set of points $\alpha_r \in \mathbb{R}/\mathbb{Z}$ is δ-spaced if their pairwise distances $\bmod 1$ are at least δ. We state it in the same form as Theorem 9.1 from *Opera de Cribro* [6].

Theorem 21. *(The Large Sieve Inequality.) For any set of δ-spaced points $\alpha_r \in \mathbb{R}/\mathbb{Z}$ and any complex numbers a_n with $n \le N$, where $0 < \delta \le \frac{1}{2}$ and N is a positive integer, we have*

$$\sum_r \Big| \sum_{n \le N} a_n e(\alpha_r n) \Big| \le (\delta^{-1} + N - 1) \sum_{n \le N} |a_n|^2.$$

This inequality is a statement about average cancellation of exponential sums, but originates from the study of sieving problems where the number of residue classes to sieve mod p is comparatively large, hence the name. Here we do not use it directly as a sieve inequality.

Lemma 22. *For any squarefree $N > 0$,*

$$\sum_{d \le D} \sum_{\nu^2 + N \equiv 0(d)} \Big| \sum_{m \le M} \alpha_m e(\nu m/d) \Big| \ll D^{1/2}(D + M)^{1/2} \log D \|\alpha\|_2$$

for any complex numbers $\{\alpha_m\}_{m \le M}$. The first sum is over $(d, N) = 1$.

Proof. By the large sieve inequality, we have immediately

$$\sum_{D/2 < d' \le D} \sum_{\nu^2 + N \equiv 0(d')} \Big| \sum_{m \le M} \alpha_m e(\nu m/d') \Big|^2 \ll (D + M)\|\alpha\|_2^2$$

when summed over $(d', 2N) = 1$. Summing over dyadic intervals, we get

$$\sum_{d' \le D} \sum_{\nu^2 + N \equiv 0 (d')} \Big| \sum_{m \le M} \alpha_m e(\nu m/d') \Big|^2 \ll (D+M) \log D \|\alpha\|_2^2,$$

and it remains to remove the parity condition on d, assuming N is odd. In this case, any given d can be written $d = 2^t d'$ where d' is odd, and an inner sum for d can be split into 2^t sums for d':

$$\begin{aligned}
\sum_{d \le D} \sum_{\nu^2 + N \equiv 0 (d)} \Big| \sum_{m \le M} \alpha_m e(\nu m/d) \Big|^2 &= \sum_{d' \le D} \sum_{2^t d' \le D} \sum_{\nu^2 + N \equiv 0 (2^t d')} \Big| \sum_{m \le M} \alpha_m e(\nu m/2^t d') \Big|^2 \\
&\le \sum_{t \le \log_2 D} \sum_{d' \le D} \sum_{\nu^2 + N \equiv 0 (d')} \Big| \sum_{m \le M} \alpha_m e(\nu m/2^t d') \Big|^2 \\
&\le \sum_{t \le \log_2 D} \sum_{k \le 2^t} \sum_{d' \le D} \sum_{\nu^2 + N \equiv 0 (d')} \Big| \sum_{m' \le M/2^t} \alpha_{2^t m' + k} e(\nu k/2^t d') e(\nu m'/d') \Big|^2 \\
&\ll \sum_{t \le \log_2 D} \sum_{k \le 2^t} (D+M) \log D \sum_{m' \le M/2^t} |\alpha_{2^t m' + k}|^2 \\
&\ll (D+M)(\log D)^2 \|\alpha\|_2^2.
\end{aligned}$$

Finally, Cauchy-Schwarz gives the desired inequality. □

In the level of distribution calculation, however, the harmonics we use will not be $e(\nu m/d)$ but instead the related sum

$$\rho(k, \ell; d) = \sum_{N y_0^2 + \ell^2 \equiv 0 (d)} e\Big(\frac{y_0 k}{d}\Big).$$

For $(d, N) = 1$, we can multiply the condition by N^{-1} and write

$$\begin{aligned}
\rho(k, \ell; d) &= \sum_{\nu^2 + N (N^{-1}\ell)^2 \equiv 0 (d)} e\Big(\frac{\nu k}{d}\Big) \\
&= \sum_{\nu^2 + N \equiv 0 (d)} e\Big(\frac{\nu N^{-1} k \ell}{d}\Big).
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{d \le D} \Big| \sum_{k \le K} \sum_{\ell \le L} \alpha_{k,\ell} \rho(k, \ell; d) \Big| &= \sum_{d \le D} \Big| \sum_{k \le K} \sum_{\ell \le L} \alpha_{k,\ell} \sum_{\nu^2 + N \equiv 0 (d)} e\Big(\frac{\nu N^{-1} k \ell}{d}\Big) \Big| \\
&\le \sum_{d \le D} \sum_{\nu^2 + N \equiv 0 (d)} \Big| \sum_{k \le K} \sum_{\ell \le L} \alpha_{k,\ell} e\Big(\frac{\nu N^{-1} k \ell}{d}\Big) \Big|.
\end{aligned}$$

Write

$$\tilde{\alpha}_n = \sqrt{\tau(n)} \sum_{k\ell = n} \alpha_{k,\ell},$$

so that $\|(\alpha_{k,\ell})\|_2 \le \|(\tilde{\alpha}_n)\|_2$. Thus, we get a large sieve bound on sums involving $\rho(k, \ell, d)$.

$$\sum_{d \le D} \Big| \sum_{k \le K} \sum_{\ell \le L} \alpha_{k,\ell} \rho(k, \ell; d) \Big| \ll D^{1/2} (D + KL)^{1/2} \log D \|(\tilde{\alpha}_n)\|_2.$$

On the other hand,

$$\begin{aligned}\|(\tilde{\alpha}_n)\|_2^2 &= \sum_k \sum_\ell \tau(k\ell)|\alpha_{k,\ell}|^2 \\ &\ll \log(KL)\|(\alpha_{k,\ell})\|_2^2.\end{aligned}$$

We write $\|\alpha\|_2 = \|(\alpha_{k,\ell})\|_2$ henceforth.

Lemma 23. *For any squarefree $N > 0$,*

$$\sum_{d\le D}\Big|\sum_{k\le K}\sum_{\ell\le L}\alpha_{k,\ell}\rho(k,\ell;d)\Big| \ll D^{1/2}(D+KL)^{1/2}(\log(KL))^{1/2}\log D\|\alpha\|_2.$$

for any complex numbers $\alpha_{k,\ell}$. Here the sum is over $(d,N)=1$.

3.4. **The Level of Distribution.** The sums $A_d(X)$ should be approximable by a multiplicative function times $A(X)$:

$$A_d(X) = g(d)A(X) + r_d(X).$$

In this section we show that $|r_d(X)|$ is small on average; write

$$R_D(X) = \sum_{d\le D}|r_d(X)|.$$

Lemma 24. *For any $\varepsilon > 0$, if $D = X^{1-3\varepsilon}$, then*

$$R_D(X) \ll X^{1-\varepsilon}, \tag{3.3}$$

where the implicit constant depends only on ε.

3.4.1. *Preliminaries.* Pick $\delta > 0$. It is possible to construct $f : \mathbb{R} \to [0,1]$ to be a smooth function approximating the indicator function of $[0,X]$ satisfying the following conditions: f is supported on $[0,X]$, identically 1 on $[\delta, X-\delta]$, and the n-th derivative of f scales inversely with δ:

$$\frac{d^n f}{dt^n} \ll \frac{1}{\delta^n},$$

uniformly in t, the implicit constant depending only on n. Define

$$A_d(f) = \sum_{d|n} a_n f(n).$$

But expanding the definition of a_n, the sum $A_d(f)$ can be divided into many sums of smooth functions over arithmetic progressions. Thus we can apply Poisson summation; write $e(t) = e^{2\pi i t}$.

$$\begin{aligned}A_d(f) &= \sum_{(x,N)=1}\Lambda(x)\sum_{Ny_0^2+x^2\equiv 0(d)}\sum_{y\equiv y_0(d)} f(x^2+Ny^2) \\ \sum_{y\equiv y_0(d)} f(x^2+ny^2) &= \frac{1}{d}\sum_k e\Big(\frac{y_0 k}{d}\Big)\int_{-\infty}^{\infty} f(x^2+Nt^2)e\Big(\frac{tk}{d}\Big)dt \\ A_d(f) &= \frac{1}{d}\sum_{(x,N)=1}\Lambda(x)\sum_{Ny_0^2+x^2\equiv 0(d)}\sum_k e\Big(\frac{y_0 k}{d}\Big)\int_{-\infty}^{\infty} f(x^2+Nt^2)e\Big(\frac{tk}{d}\Big)dt. \\ &= \frac{1}{d}\sum_{(x,N)=1}\Lambda(x)\sum_k \rho(k,x;d)I(k/d,x),\end{aligned}$$

where

$$\begin{aligned}\rho(k,x;d) &= \sum_{Ny_0^2+x^2\equiv 0(d)} e\Big(\frac{y_0 k}{d}\Big)\\ I(\alpha,x) &= \int_{-\infty}^{\infty} f(x^2+Nt^2)e(t\alpha)dt.\end{aligned}$$

The frequency $k=0$ is the main term, as usual.

$$M_d(f)=\frac{1}{d}\sum_{(x,N)=1}\Lambda(x)\rho(0,x;d)I(0,x).$$

The other frequencies form the remainder term.

$$r_d(f)=\frac{1}{d}\sum_{k\neq 0}\sum_{(x,N)=1}\Lambda(x)\rho(k,x;d)I(k/d,x).$$

Here we show that $A_d(f)$ and $M_d(f)$ (and therefore $r_d(f)$) are good approximations for $A_d(X), M_d(X)$ (and $r_d(X)$) respectively, at least on average. We take

$$M_d(X)=g(d)A(X)$$

to be the expected value of $A_d(X)$.

Lemma 25. *Let $X^{1/2+\varepsilon}\ll\delta\le X$ and $D\le X$. Then*

$$\begin{aligned}\sum_{d\le D}|A_d(f)-A_d(X)| &\ll \delta X^{\varepsilon}\\ \sum_{d\le D}|M_d(f)-M_d(X)| &\ll \delta^{1/2}X^{1/2}\log^2 X.\end{aligned}$$

Proof. For the first claim, note that $|a_n f(n)-a_n|$ appears in at most $\tau(n)$ (where τ is the divisor function) terms, and each is bounded by $a_n\ll n^{\varepsilon}$ on the two intervals where $f\neq 1$.

To estimate $M_d(f)$, we note that for most x, $(x,d)=1$ and therefore $\rho(0,x;d)d^{-1}=g(d)$. The others are powers of primes $p|d$, of which there are at most $O(\omega(d)\log X)=O(\log^2 X)$. Further, for individual x, $I(0,x)\ll\sqrt{X}$. Thus

$$M_d(f)=g(d)\sum_{(x,d)=1}\Lambda(x)I(0,x)+O\Big(\frac{1}{d}X^{1/2}\log^2 X\Big).$$

Meanwhile, $I(0,x)$ is a good approximation for the number of y satisfying $x^2+ny^2\le X$:

$$\begin{aligned}I(0,x) &= \int_{-\infty}^{\infty} f(x^2+Nt^2)dt\\ &= \sum_{x^2+Ny^2\le X} 1+O(\delta(X+\delta-x^2)^{-1/2}),\end{aligned}$$

implying

$$\begin{aligned}M_d(f) &= g(d)A(X)+O\Big(g(d)\delta^{1/2}X^{1/2}\log X+\frac{1}{d}X^{1/2}\log^3 X\Big)\\ \sum_{d\le D}|M_d(f)-M_d(X)| &\ll \delta^{1/2}X^{1/2}\log^2 X+X^{1/2}\log^4 X,\end{aligned}$$

as desired. □

For a final simplification, we show that the tail of the sum $r_d(f)$ is negligible. This is the reason why we need a smoothed approximation with small derivatives; picking δ as large as possible will allow us to cut off as large a tail as possible in $r_d(f)$.

Lemma 26. *If $\varepsilon > 0$ and K satisfies $K \geq DX^{1/2+\varepsilon}\delta^{-1}$, then for any $k \geq K$,*

$$\Lambda(x)\rho(k, x; d)I(k/d, x) \ll X^{-j\varepsilon}$$

for any $j \geq 1$, the implicit constant depending only on j.

Proof. Note that

$$\begin{aligned} \frac{\partial^j}{\partial y^j} f(x^2 + Ny^2) &= \sum_{0 \leq 2i \leq j} c_{ij} y^{j-2i} f^{(j-i)}(x^2 + Ny^2) \\ &\ll \Big(\frac{\sqrt{X}}{\delta}\Big)^j \end{aligned}$$

for some positive constants c_{ij}. It follows by repeated integration by parts that

$$I(k/d, x) \ll \sqrt{X}\Big(\frac{d\sqrt{X}}{k\delta}\Big)^j,$$

so if $k \geq DX^{1/2+\varepsilon}\delta^{-1}$ then $I(k/d, x)$ is smaller than $X^{-j\varepsilon}$ for any $j \geq 0$. □

3.4.2. *The Smoothed Remainder Term.* Using Lemmas 25 and 26, it suffices to show that the truncated smoothed remainder term is small. Note that $\rho(k, x; d)$ and $I(k/d, x)$ are both even in k.

Lemma 27. *If*

$$r_d^*(f) = \frac{2}{d} \sum_{0 < k \leq K} \sum_x \Lambda(x)\rho(k, x; d)I(k/d, x)$$

is small, where $K = DX^{1/2+\varepsilon}\delta^{-1}$, then

$$\sum_{d \leq D} |r_d^*(f)| \ll D^{1/2}X^{3/2+\varepsilon}\delta^{-1}(\log X)^2.$$

Proof. Now it is time to change variables in $I(k/d, x)$ to remove dependence on k from the phase of the integral, so that we can interchange summation and integration:

$$\begin{aligned} I(k/d, x) &= \int_{-\infty}^{\infty} f(x^2 + Nt^2)e(td^{-1}k)dt \\ &= k^{-1}\sqrt{X} \int_{-\infty}^{\infty} f(x^2 + Nk^{-2}Xt^2)e(td^{-1}\sqrt{X})dt. \\ r_d^*(f) &= \frac{2\sqrt{X}}{d} \int_0^K \sum_{0 < k \leq K} \sum_{x \leq \sqrt{X}} \Lambda(x)k^{-1} f(x^2 + Nk^{-2}Xt^2)\rho(k, x; d)e(td^{-1}\sqrt{X})dt \\ &\leq \frac{2\sqrt{X}}{d} \int_0^K \Big| \sum_{0 < k \leq K} \sum_{x \leq \sqrt{X}} \Lambda(x)k^{-1} f(x^2 + Nk^{-2}Xt^2)\rho(k, x; d)\Big| dt. \end{aligned}$$

Averaging over d,

$$\sum_{d\leq D} d|r_d^*(f)| \leq 2\sqrt{X}\int_0^K \sum_{d\leq D}\Big|\sum_{0<k\leq K}\sum_{x\leq\sqrt{X}}\Lambda(x)k^{-1}f(x^2+Nk^{-2}Xt^2)\rho(k,x;d)\Big|dt.$$

The integrand is set up to apply the large sieve inequality from Section 3.3. To apply Lemma 23 to maximum effect, we can take advantage of the support of f to note that $k > t\sqrt{N}$ for the integrand to be nonzero.

$$\begin{aligned}\sum_{d\leq D}\Big|\sum_{t\sqrt{N}<k\leq K}\sum_{x\leq\sqrt{X}}\Lambda(x)k^{-1}f(x^2+Nk^{-2}Xt^2)\rho(k,x;d)\Big| &\ll D^{1/2}(D+K\sqrt{X})^{1/2}X^{1/4}t^{-1/2}(\log X)^2\\ &\ll DX^{5/4+\varepsilon}t^{-1/2}\delta^{-1/2}(\log X)^2.\end{aligned}$$

Integrating over t now, we find that

$$\begin{aligned}\sum_{d\leq D} d|r_d^*(f)| &\ll D^{3/2}X^{3/2+\varepsilon}\delta^{-1}(\log X)^2\\ \sum_{d\leq D}|r_d^*(f)| &\ll D^{1/2}X^{3/2+\varepsilon}\delta^{-1}(\log X)^2,\end{aligned}$$

where we removed the factor of D by summing over dyadic intervals. □

Corollary 28. *For any $\varepsilon > 0$, if f is smooth with parameter $\delta = X^{1-\varepsilon}$, then*

$$R_D(f) \ll X^{1-\varepsilon}$$

whenever $D \ll X^{1-3\varepsilon}$.

Combined with Lemma 25, we get the level of distribution required, proving Lemma 24.

3.5. **The Bilinear Forms Condition.** Controlling the bilinear forms condition has three main ingredients.

The first step is to massage the double sum by expanding a_{mn} as a sum over factorizations of m and n over Gaussian integers and thereby "unfold the multiplicity" present in the sum – otherwise, some terms may be much larger than others. Once this expansion is achieved we can apply Cauchy-Schwarz while losing only a log factor.

The next step is to apply two-dimensional Fourier analysis to estimate the various sums that result from Cauchy-Schwarz in terms of weighted sums of the Möbius function over congruence lattices in $\mathbb{Z}[i]$.

Finally, these sums are treated with standard techniques from the theory of zero-free regions of L-functions of quadratic fields, although this part is omitted altogether from Fouvry-Iwaniec.

The main difference in our problem will be to work in a general imaginary quadratic field $\mathbb{Q}[\sqrt{-D}]$ with ring of integers $\mathcal{O}$, and we will not have the luxury of unique factorization. Fortunately, for any fixed D the class group is finite and we can split the resulting sums along individual ideal classes; after this each individual sum is completely analogous to those in Fouvry-Iwaniec.

The last part of this calculation is extremely standard and Fouvry and Iwaniec already skipped a much of it – we will skip even more, only sketching the reductions and focusing on how to deal with the class group.

3.5.1. *Preliminaries.* In this section we replicate the calculations of Fouvry and Iwaniec to reduce $B(X; y, z)$ to a manageable form for Cauchy-Schwarz. Recall the bound we need to finish the proof of Theorem 1. With

$$B(X; y, z) = \sum_{bd \leq X, b > y} \mu(b) a_{bd} \Big(\sum_{c | d, c > z} \Lambda(c) \Big),$$

then we want

$$|B(X; y, z)| \ll_A X (\log X)^{-A}$$

for any $A > 0$. Write $\Delta = (\log X)^{-A}$. We first simplify the sum we want to handle.

$$|B(X; y, z)| \leq \log X \sum_{d > z} \Big| \sum_{y < b \leq X/d} \mu(b) a_{bd} \Big|.$$

It suffices to show cancellation in the latter sum when we break up the ranges of b and d into short intervals. Write

$$B(M, N) = \sum_{M < m \leq 2M} \Big| \sum_{N < n \leq N'} \mu(n) a_{mn} \Big|,$$

where $N' = e^{\Delta} N$. The original ranges on b and d can be split up into these intervals with negligible error, M varying over $2^j z$ and N varying over $e^{\Delta k} y$.

$$|B(X; y, z)| \leq \log X \sum_{M > z, N > y} \sum_{\Delta X < MN < X} B(M, N) + O(\Delta X (\log X)^2),$$

where the few terms $\mu(n) a_{mn}$ not covered by these intervals are bounded trivially. Note that M and N range through at most $O(\log X)$ and $O(\Delta^{-1} \log X)$ values, respectively. It suffices to prove that for any M, N satisfying $M > z, N > y, \Delta X < MN < X$,

$$B(M, N) \ll \Delta^2 MN (\log X)^{-3}.$$

Next, we would like to reduce to the case $(m, n) = 1$. Write

$$B_d(M, N) = \sum_{M < m \leq 2M} \Big| \sum_{N < n \leq N', (m,n) = d} \mu(n) a_{mn} \Big|,$$

noting that

$$\begin{aligned} a_{mn} &\leq \log X \sum_{x^2 + Ny^2 = mn} 1 \\ &\leq \tau(mn) \log X, \end{aligned}$$

where the divisor function $\tau(n)$ is an upper bound bound on the number of $x + y\sqrt{-N}$ with norm n. Thus,

$$\begin{aligned} B_d(M, N) &\ll \log X \tau(d^2) \sum_{M/d < m \leq 2M/d} \tau(m) \sum_{N/d < n \leq e^{\Delta} N/d} \tau(n) \\ &\ll (\log X) \tau(d^2) (M \log M) (e^{\Delta} - 1) (N \log N) d^{-2} \\ &\ll \frac{\tau(d^2)}{d^2} \Delta MN (\log X)^3. \end{aligned}$$

Also, $B_d(M,N) \le B_1(dM, N/d)$, simply by moving the factors in n to m. Thus

$$\begin{aligned} B(M,N) &= \sum_{d\le\Delta^{-2}} B_d(M,N) + O\Big(\Delta MN(\log X)^3 \sum_{d>\Delta^{-2}} \frac{\tau(d^2)}{d^2}\Big) \\ &= \sum_{d\le\Delta^{-2}} B_d(M,N) + O\Big(\Delta^{3-\varepsilon} MN(\log X)^3\Big) \\ &= \sum_{d\le\Delta^{-2}} B_1(dM,N/d) + O(\Delta^2 MN(\log X)^{-3}), \end{aligned}$$

for Δ sufficiently large, and it suffices to show Δ^5 cancellation in $B_1(M,N)$ over a slightly larger range of M, N. We need to show that if $M > z, N > \Delta^2 y$, and $\Delta X < MN < X$, then

$$B_1(M,N) \ll \Delta^5 MN.$$

Next, we expand a_{mn} as a sum over factorizations of m and n as ideals over $\mathcal{O}_N$, the ring of integers of $\mathbb{Q}[\sqrt{-N}]$. We can write a_{mn} as

$$a_{mn} = \sum_{N(I)=m} \sum_{N(J)=n} \sum_{(z)=IJ} \Lambda(\mathrm{Re}(z)),$$

where for two ideals I, J whose product is principal, we pick every possible generator for the product IJ and sum Λ over their real parts (we extend Λ by zero over the nonpositive integers). Since m, n are coprime the factorization $(z) = IJ$ is unique for any given z whose norm is mn.

Now, we can rewrite $B_1(M,N)$ in terms of pairs of ideals in conjugate ideal classes. There are only a fixed finite number of choices of generator for a principal ideal, so let's write

$$\Lambda(I) = \sum_{(z)=I} \Lambda(\mathrm{Re}(z)).$$

With this notation, we can expand out $B_1(M,N)$ for some complex numbers $\beta(m)$ of norm 1:

$$B_1(M,N) = \sum_{M<N(I)\le 2M} \sum_{N<N(J)\le N', (N(I),N(J))=1} \beta(N(I))\mu(N(J))\Lambda(IJ).$$

Of course, there is a hidden restriction in the second sum, namely that JI must be principal. To remedy this situation, we fix representatives $I_1, I_2, \dots, I_h$ of the elements of the ideal class group of $\mathcal{O}_D$, and break up the sum into $h = h(D)$ pieces:

$$\begin{aligned} B_1(M,N) &= \sum_{i\le h} B_1^i(M,N) \\ B_1^i(M,N) &= \sum_{I\in[I_i]} \sum_{J\in[I_i]^{-1}, (N(I),N(J))=1} \beta^*(N(I))\mu^*(N(J))\Lambda(IJ). \end{aligned}$$

In preparation for future manipulations we folded the bounds on $N(I), N(J)$ into the support of μ, β – that is, β^* is β cut off outside of $(M, 2M]$ and μ^* is μ cut off outside $(N, N']$.

For a fixed ideal class $[I_i]$, all (fractional) ideals in the class can be written in the form $(w)I_i$, where w ranges through $K^\times = \mathbb{Q}[\sqrt{-D}]^\times$. The choice of w is unique up to the number of units u_K in K, of which there will be either 4 (if $D = 1$), 6 (if $D = 3$), or 2. We only care about *integral* ideals appearing in this class, but this is not the same as the set of $(w)I_i$ where

w itself is in $\mathcal{O}$. In fact, the set will be of all $w \in I_i^{-1}$, the inverse ideal of I_i. Thus, we can rewrite, where Λ is now the regular von Mangoldt function (extended by zero to negatives):

$$B_1^i(M,N) = \frac{1}{u_K} \sum_{w \in I_i^{-1}} \sum_{z \in I_i, (N(wI_i), N(zI_i^{-1}))=1} \beta^*(N(wI_i))\mu^*(N(zI_i^{-1}))\Lambda(\mathrm{Re}(wz)).$$

Note that $I_i I_i^{-1} = (1)$ so $[I_i]^{-1}$ is the same ideal class as $[I_i^{-1}]$ and we may assume we picked our class representatives in such a way that the representative of $[I]^{-1}$ is the ideal inverse I^{-1}. Also, the term in front comes from the fact that any given generator of (wz) is represented in the sum u_K times.

We are led to bound general double sums of the form

$$B_I(M,N) = \sum_{w \in I^{-1}} \sum_{z \in I, (N(wI), N(zI^{-1}))=1} \beta^*(N(wI))\mu^*(N(zI^{-1}))\Lambda(\mathrm{Re}(wz))$$

where μ^* is Möbius cut off to the interval $(N, N']$ and β^* is a complex-valued function supported on $(M, 2M]$ with values of norm 1.

The sole purpose of introducing the condition $(m,n) = 1$ was to separate w and z; this being done, we would like to remove it again. We use the identity

$$\sum_{r|(m,n)} \mu(r) = \begin{cases} 1 & (m,n) = 1 \\ 0 & \text{otherwise} \end{cases},$$

which gives, for a fixed w,

$$\sum_{z \in I, (N(wI), N(zI^{-1}))=1} \mu^*(N(zI^{-1}))\Lambda(\mathrm{Re}(wz)) = \frac{1}{u_K} \sum_{r | N(wI)} \mu(r) \sum_{N(\zeta)=r} \sum_{z \in I} \mu^*(N(\zeta z I^{-1}))\Lambda(\mathrm{Re}(w\zeta z)),$$

We insert this into $B_I(M,N)$ and expand as a sum over r, bounding the number of choices of $\zeta \in \mathcal{O}$ with norm r trivially by $\tau(r)$, and folding ζ into w:

$$B_I(M,N) \ll \sum_r \tau(r) \sum_{w \in I^{-1}, r^2 | N(wI)} \Big| \sum_{z \in I} \mu^*(rN(zI^{-1}))\Lambda(\mathrm{Re}(wz)) \Big|.$$

Next we cut off the tail above $r > \Delta^{-5}$, which contribute at most

$$\Delta MN \sum_{r > \Delta^{-5}} \tau(r)^2 r^{-2} \ll \Delta^{6-\varepsilon} MN (\log X)^2,$$

smaller than the bound on the whole of B_I that we need. Put

$$C_I(M,N;r) = \sum_{w \in I^{-1}} \sum_{z \in I} \beta^*(N(wI)/r^2)\mu^*(rN(zI^{-1}))\Lambda(\mathrm{Re}(wz)),$$

where we replaced the condition $r^2 | N(wI)$ by the weaker condition of $N(wI)/r^2$ lying in the support of β^*. It will suffice to show that for each $r \leq \Delta^{-5}$,

$$C_I(M,N;r) \ll \Delta^{11} MN,$$

for every $M \geq z, N \geq \Delta^5 y$, and $\Delta X < MN < X$.

Finally, it is not difficult to reduce bounding $C_I^*(M,N;r)$, where we restrict the variable w to be primitive in I, that is to say $w \notin cI$ for any rational integer $c > 1$.

$$C_I(M,N;r) = \sum_{c \geq 1} C_{cI}^*(c^{-2}M, N; r),$$

and for large $c \gg \Delta^{-K}$ the terms in the sum are negligible.

At this point we can forget about the exact power of Δ needed. It suffices to show the following.

Lemma 29. *Let $A > 0$ be a positive constant. Then, for all $M \geq z, N \geq y(\log X)^{-A}$ and $X(\log X)^{-A} < MN < X$, and $r \leq (\log X)^A$,*

$$C_I^*(M, N; r) \ll MN(\log X)^{-A},$$

where the implicit constant depends only on A.

3.5.2. *Cauchy-Schwarz.* We will perform slight generalizations of the argument of Fouvry and Iwaniec to lattices beyond $\mathbb{Z}[i]$ in the complex plane. Write $w^* \in I^{-1}$ to mean w is primitive in the fractional ideal I^{-1}. We have

$$\begin{aligned} C_I^*(\alpha, \beta, \lambda) &= \sum_{z \in I} \sum_{w^* \in I^{-1}} \alpha(z)\beta(w)\lambda(\mathrm{Re}(zw)) \\ &= \sum_l \lambda(l) \sum_{z \in I} \sum_{\mathrm{Re}(zw)=l} \alpha(z)\beta(w) \\ &\leq \sum_l |\lambda(l)| \sum_{w^* \in I^{-1}} |\beta(w)| \Big| \sum_{\mathrm{Re}(zw)=l} \alpha(z) \Big| \\ &\leq \|\lambda\| \cdot \|\beta\| \Big(\sum_{w^* \in I^{-1}} g(w) \sum_l \Big| \sum_{\mathrm{Re}(zw)=l} \alpha(z) \Big|^2 \Big)^{1/2}. \end{aligned}$$

The last step is Cauchy-Schwarz, where g is the indicator function of the support of β, and we cut off λ to be finitely supported so that its norm is finite. For the purposes of later Fourier analysis it will be better to replace g with a smooth approximation; as long as it is at least 1 on the support of β the inequality still holds.

All cancellation we need will come from average cancellation of the inner sum, which in our case $\alpha(z) = \mu^*(N(zI^{-1}))$, over the set of $z \in I$ for which $\mathrm{Re}(zw) = l$. Write

$$D_I(\alpha) = \sum_{w^* \in I^{-1}} g(w) \sum_l \Big| \sum_{z \in I, \mathrm{Re}(zw)=l} \alpha(z) \Big|^2,$$

so that

$$D_I(\alpha) = \sum_{w^* \in I^{-1}} g(w) \sum_{z \in I, \mathrm{Re}(zw)=0} \Big(\sum_{z_1 - z_2 = z} \alpha(z_1)\bar{\alpha}(z_2) \Big),$$

when we expand the square. This is because two values of z satisfy $\mathrm{Re}(z_1 w) = \mathrm{Re}(z_2 w)$ iff $\mathrm{Re}((z_1 - z_2)w) = 0$.

At this point, the primitivity for w allows us to explicitly determine the values of z which satisfy $\mathrm{Re}(zw) = 0$. In general, this is equivalent to having $z = c\overline{w}\sqrt{-N}$ for some rational c. However, since $I^{-1} = \bar{I}/N(I)$, we see that $c\bar{w}\sqrt{-N} \in I$ iff $-cwN(I)^{-1}\sqrt{-N} \in I^{-1}$. Since w is assumed to be primitive in I^{-1}, the constraint that $z \in I$ resolves to $c \in c_I(w)\mathbb{Q}$ for the rational constant $c_I(w) = N(I)\mathrm{gcd}(\mathrm{Re}(w), N)^{-1}$. Rearranging the sum above, we get

$$\begin{aligned} D_I(\alpha) &= \sum_n \sum_{w^* \in I^{-1}} g(w) \Big(\sum_{z_1 - z_2 = n c_I(w) \overline{w}\sqrt{-N}} \alpha(z_1)\bar{\alpha}(z_2) \Big) \\ &= D_I^0(\alpha) + 2D_I^+(\alpha). \end{aligned}$$

Here $D_I^0(\alpha)$ is the term $n = 0$ and $D_I^+(\alpha)$ is the sum over $n > 0$. We have

$$D_I^0(\alpha) = \sum_{w \in I^{-1}} g(w)\|\alpha\|^2 \ll N^2\|\alpha\|^2,$$

since we pick g to be supported on an annulus of area $O(N^2)$ and bounded by a constant. It remains to bound D_I^+.

3.5.3. *Controlling D_I^+.* Because the original sum is over primitive w we can rewrite D_I^+as

$$D_I^+(\alpha) = \sum_{w^* \in I^{-1}} g(w) \sum_{n \geq 1} (\alpha * \alpha)_I(nw),$$

where the convolution is defined

$$(\alpha * \alpha)_I(w) = \sum_{z_1 - z_2 = c_I(w)\bar{w}\sqrt{-N}} \alpha(z_1)\bar{\alpha}(z_2)$$

over $z_1, z_2 \in I$. Next, we unfold the primitivity constraint on w via Möbius inversion:

$$D_I^+(\alpha) = \sum \sum_{b,c>0} \mu(b) D_I(\alpha; b, c),$$

where

$$D_I(\alpha; b, c) = \sum_{w \in bcI^{-1}} g(w/c)(\alpha * \alpha)_I(w).$$

Note that $N(w) \ll N$ since $\alpha = \mu^* \circ N$ and μ^* is supported on $[N, N')$, and also $N(w) = \Theta(cM)$ since g is supported on essentially the same annulus that β^* is. Incidentally this shows that the sum is nonempty only when $c \ll NM^{-1}$; write C for the largest such value of c.

It is fruitful to cut off the largest values of b as well as the values of c significantly smaller than C; note that the trivial bound is

$$D_I(\alpha; b, c) \ll \|\alpha\|^2 M^2 b^{-2},$$

since there are $O(M^2b^{-2})$ lattice points in bcI^{-1} with norm $N(w) = \Theta(cM)$. Then, for a parameter $1 \leq \Lambda \leq C$ that we will pick later, whenever $b \geq \Lambda$ or $c \leq C\Lambda^{-1}$, the contribution to $D_I^+(\alpha)$ is $O(\|\alpha\|^2 MN\Lambda^{-1})$, so

$$D_I^+(\alpha) = \sum_{b \leq \Lambda} \mu(b) \sum_{C\Lambda^{-1} < c < C} D_I(\alpha; b, c) + O(\|\alpha\|^2 MN\Lambda^{-1}).$$

3.5.4. *Controlling $D_I(\alpha; b, c)$.* Fouvry and Iwaniec deal with the convolution in D_I using Fourier inversion. Here because of the change in lattice, w runs through an congruence condition in $I^{-1} \in \mathbb{Q}[\sqrt{-N}]$ instead of $\mathbb{Z}[i]$. The goal of this section will be to reduce bounding $D_I(\alpha; b, c)$ to that of

$$D_I(\alpha; d) = 2\pi N^{-2} \sum_{z_1 - z_2 \in dI} \alpha(z_1)\bar{\alpha}(z_2) \exp(-2\pi|z_1 - z_2|N^{-1}).$$

Finally, getting nontrivial cancellation in this latter sum is standard for $\alpha(z) = \mu(N(zI^{-1}))$, reducing essentially to a Siegel-Walfisz type theorem involving zero-free regions of L-functions of certain Hecke Grossencharacters related to imaginary quadratic number fields. Such a result was already standard in the time of the Friedlander-Iwaniec theorem.

4. Acknowledgments

This thesis would not exist without the guidance of my thesis advisor, Arul Shankar. It would be riddled with mathematical and typographical errors if not for the efforts of my meticulous thesis reader, Noam Elkies. I would also like to thank Terry Tao for a helpful discussion on MathOverflow and Zeb Brady for insightful conversations about the parity problem.

References

[1] Balady, Steve. "Annotation: A discussion of the fundamental ideas behind Selberg's "Elementary proof of the prime-number theorem."" (2006).
[2] Duke, W., J. B. Friedlander, and H. Iwaniec. "Equidistribution of roots of a quadratic congruence to prime moduli." Annals of Mathematics (1995): 423-441.
[3] Ford, Kevin. "On Bombieri's asymptotic sieve." Transactions of the American Mathematical Society 357.4 (2005): 1663-1674.
[4] Ford, Kevin. "Vinogradov's integral and bounds for the Riemann zeta function." Proceedings of the London Mathematical Society 85, no. 03 (2002): 565-633.
[5] Fouvry, Etienne, and Henryk Iwaniec. "Gaussian primes." Acta Arithmetica 79 (1997): 249-287.
[6] Friedlander, John, and Henryk Iwaniec. *Opera de cribro*. Vol. 57. American Mathematical Soc., 2010.
[7] Friedlander, John, and Henryk Iwaniec. "Asymptotic sieve for primes." Annals of Mathematics 148 (1998): 1041-1065.
[8] Friedlander, John, and Henryk Iwaniec. "The polynomial $x^2 + y^4$ captures its primes." Annals of Mathematics 148 (1998): 945-1040.
[9] Harman, Glyn. *Prime-detecting sieves*. No. 33. Princeton University Press, 2007.
[10] Heath-Brown, D. Roger. "Primes represented by $x^3 + 2y^3$." Acta Mathematica 186, no. 2 (2001): 1-84.
[11] Montgomery, Hugh Lowell, and Robert Charles Vaughan. "The large sieve." Mathematika 20, no. 02 (1973): 119-134.
[12] Selberg, Atle. *Collected Papers, volume II*. Springer-Verlag, 1991.
[13] van Lint, J., and H. Richert. "On primes in arithmetic progressions." Acta Arithmetica 11.2 (1965): 209-216.
[14] Vaughan, R. C. "Mean value theorems in prime number theory." Journal of the London Mathematical Society 2, no. 2 (1975): 153-162.