ACTA ARITHMETICA
LXXIX.3 (1997)

Gaussian primes

by

Etienne Fouvry (Orsay) and Henryk Iwaniec (New Brunswick, N.J.)

Dedicated to Professor Cassels

1. Introduction and main theorem. Gaussian primes are the irreducible elements of the ring $\mathbb{Z}[i]$; they emerge from factorization of the rational primes. Any prime $p \equiv 1 \pmod 4$ is represented as the sum of two squares

$$p = l^2 + m^2. \tag{1.1}$$

Therefore such a p factors into two complex conjugate Gaussian primes, say

$$p = (l + im)(l - im) = \pi\overline{\pi}. \tag{1.2}$$

Moreover, $2 = (1 + i)(1 - i)$, but the primes which are $\equiv -1 \pmod 4$ do not factor in $\mathbb{Z}[i]$. The Gaussian primes $\pi = l + im$ can be viewed as two-dimensional lattice points, and this allows us to explore the distribution problems from various directions. For instance, unlike for the rational primes, one may seek an infinite string of Gaussian primes with absolutely bounded distances between consecutive points (can one walk to infinity stepping on Gaussian primes?).

Applying standard methods of prime number theory to L-functions with Grossencharacters E. Hecke [H] showed that Gaussian primes are equidistributed over arithmetic progressions within regular planar domains. Actually, Hecke applied his method to prime ideals in any number field. However, the Grossencharacters are not capable of controling the coordinates of $\pi = l + im$ as much as we would like, say to fix $l = 1$ producing primes of type $p = m^2 + 1$, or to put l in a thin set of arithmetic nature.

In this paper we apply ideas of sieve methods to prove that there are infinitely many primes of type $p = l^2 + m^2$ where l is a prime number. Actually, our main result is more general, and we also establish an asymptotic formula.

Research of the second author supported in part by the NSF grant DMS-9500797.

THEOREM 1. *Let λ_l be complex numbers with $|\lambda_l| \le 1$. Then*

$$\sum_{l^2+m^2\le x} \lambda_l \Lambda(l^2+m^2) = \sum_{l^2+m^2\le x} \lambda_l \psi(l) + O(x(\log x)^{-A}) \tag{1.3}$$

where Λ is the von Mangoldt function,

$$\psi(l) = \prod_{p\nmid l}\Big(1 - \frac{\chi(p)}{p-1}\Big), \tag{1.4}$$

χ is the non-trivial character to modulus 4, A is any positive number and the implied constant in the error term depends only on A.

In particular, by (1.3) one infers

$$\sum_{l^2+m^2\le x} \Lambda(l)\Lambda(l^2+m^2) = 2\prod_p \Big(1 - \frac{\chi(p)}{(p-1)(p-\chi(p))}\Big)x + O(x(\log x)^{-A}). \tag{1.5}$$

Theorem 1 admits various modifications. Employing the Hecke Grossen-characters in our proof one can derive the expected asymptotic formula when the points $z = l + im$ run over any regular planar domain and are restricted to any fixed primitive residue class. To put these results in perspective we shall write explicitly the representations of (1.1). These are determined by p up to order and sign by virtue of the unique factorization in $\mathbb{Z}[i]$. Since $p \equiv 1 \pmod 4$, we can require

$$l \equiv (-1)^{(p-1)/4} \pmod 4.$$

Such an l is unique, namely we have $l = \frac{1}{2}a_p$, where

$$a_p = -\sum_{x \,(\mathrm{mod}\, p)} \Big(\frac{x^3-x}{p}\Big).$$

This is (apart from the sign) the Jacobsthal sum [J]. For such choice of l we have m even (m is determined up to sign) and the factorization (1.2) has $\pi = l+im \equiv 1 \pmod{2(1+i)}$, thus π and $\overline{\pi}$ are primary and $a_p = \pi + \overline{\pi} = 2l$. These primary primes are building blocks for the Hasse–Weil L-function of the elliptic curve (see [IR], p. 307)

$$E: \quad y^2 = x^3 - x.$$

For any $p \equiv 1 \pmod 4$ the number of points on E over the field $\mathbb{F}_p$ is equal to $p - a_p$, where a_p is also the eigenvalue of the Hecke operator T_p on the modular form associated with E (a certain theta function). Therefore Theorem 1, with its restriction by $l+m \equiv 1 \pmod 4$ ensures that any ample set of integers $\equiv 2 \pmod 4$ must contain the Hecke eigenvalues a_p, with the correct asymptotic frequency.

Another interpretation of the results is offered through the solutions to the quadratic congruence

$$\nu^2 + 1 \equiv 0 \pmod{p}. \tag{1.6}$$

For $p \equiv 1 \pmod 4$ the solutions are given by $\nu \equiv \pm m/l \pmod p$; therefore (1.5) implies that infinitely often $\nu \pmod p$ can be seen as a fraction with prime denominator.

In our approach (based on sieve ideas) the special rational points ν/d with

$$\nu^2 + 1 \equiv 0 \pmod{d} \tag{1.7}$$

will play a prominent role. The key observation is that the points ν/d are very well-spaced modulo 1, considerably better than the set of all the rationals c/d with $(c, d) = 1$. Although spacing property is not as deep as the equidistribution (see [DFI]), nevertheless it yields a powerful large sieve type inequality (see Lemma 2).

Besides the main Theorem 1 we shall establish several easier results on norms of ideals in abelian fields in place of primes (see Section 5).

We conclude this introduction by mentioning only three somewhat related results from a vast literature on the subject. In 1968, G. J. Rieger [R] established that the number of integers $n \le x$ which can be represented as the sum of two squares $n = l^2 + m^2$, with l a prime (each n counted without the multiplicity of such representations) has order of magnitude $x/\log x$. More recently M. Coleman [C] showed there are infinitely many primes $p = l^2 + m^2$ with l a small positive integer, namely $l < p^{0.1631}$. This is an improvement of many earlier results of that kind which are obtained by employing the theory of Hecke L-functions. We also recommend the work of W. Duke [D], which gives a powerful treatment of a variety of related problems by means of Grossencharacters.

Using sieve methods J. Pomyka/la [P] has considered the equation $N\mathfrak{a} = l^2+m^2$ where $\mathfrak{a}$ runs over the integral ideals of a fixed cubic, normal field, and has shown there are infinitely many of these with l a small prime, namely $l < m^{7/20}$. We shall improve 7/20 to 9/40, which result follows from a more general Theorem 7.

Acknowledgements. We thank A. Schinzel for pointing out the paper [R], and J. Friedlander for helpful suggestions. Our work on this problem began during the visit of E. Fouvry to Rutgers University in February–March 1995; he is thankful for receiving a warm welcome.

2. A large sieve inequality for roots of quadratic congruences. The classical large sieve inequality gives an ℓ_2-estimate for a general trigono-

metric polynomial

$$S(x) = \sum_{n \le N} \alpha_n e(xn)$$

at well-spaced points $x \pmod 1$. Precisely, if $\|x_r - x_s\| \ge \delta$ for $r \ne s$, then

$$\sum_r |S(x_r)|^2 \le c(\delta^{-1} + N)\|\alpha\|^2$$

where

$$\|\alpha\|^2 = \sum |\alpha_n|^2,$$

and c is an absolute constant. The first result of this type was established by Davenport and Halberstam [DH], and the best possible constant $c = 1$ is due to Selberg [S] and Montgomery–Vaughan [MV]. We shall employ their result for the arithmetic points $\nu/d \pmod 1$ with ν varying over the roots of the congruence (1.7). These points can be expressed by the primitive representations of the modulus as the sum of two squares,

$$d = r^2 + s^2 \quad \text{with} \quad (r, s) = 1 \text{ and } -s < r \le s.$$

Each such representation corresponds to a unique root of (1.7) given by $\nu s \equiv r \pmod d$. Hence

$$\frac{\nu}{d} \equiv \frac{r}{sd} - \frac{\overline{r}}{s} \pmod 1 \quad \text{where } r\overline{r} \equiv 1 \pmod s.$$

Here the fraction $\overline{r}/s$ has much smaller denominator than ν/d, and the other term is negligible. Precisely, we have

$$\frac{|r|}{sd} < \frac{1}{2s^2}.$$

Hence we infer that the distinct points $\nu/d \pmod 1$ for which the corresponding r have a fixed sign and the moduli restricted to $8D < d \le 9D$ are well-spaced. Indeed, $2D^{1/2} < s < 3D^{1/2}$ so

$$\left\| \frac{\nu}{d} - \frac{\nu_1}{s_1} \right\| > \frac{1}{ss_1} - \max\left(\frac{1}{2s^2}, \frac{1}{2s_1^2}\right) > \frac{1}{4ss_1} > \frac{1}{36D}.$$

Therefore by the large sieve inequality we conclude the following:

Lemma 2. *For any complex numbers α_n we have*

$$\sum_{8D < d \le 9D} \sum_{\nu^2 + 1 \equiv 0 \,(\mathrm{mod}\, d)} \left| \sum_{n \le N} \alpha_n e\left(\frac{\nu n}{d}\right) \right|^2 \le 72(D + N)\|\alpha\|^2.$$

Applying Cauchy's inequality and counting lattice points inside a quarter of a disk we deduce from Lemma 2 that

$$\sum_{d \le D} \sum_{\nu^2 + 1 \equiv 0 \,(\mathrm{mod}\, d)} \left| \sum_{n \le N} \alpha_n e\left(\frac{\nu n}{d}\right) \right| \le 150 D^{1/2}(D + N)^{1/2}\|\alpha\|. \tag{2.1}$$

We shall use (2.1) for

$$\alpha_n = \sum_{kl=n} \alpha_{k,l}$$

with $n > 0$ where $\alpha_{k,l}$ are any complex numbers. We define $\widetilde{\alpha}_n = \alpha_n \sqrt{\tau(n)}$ where τ is the usual divisor function. Since

$$\|\alpha\|^2 \le \|\widetilde{\alpha}\|^2 = \sum_k \sum_l |\alpha_{k,l}|^2 \tau(kl)$$

we obtain

$$\sum_{d\le D} \sum_{\nu^2+1\equiv 0 \,(\mathrm{mod}\, d)} \Big| \sum_{\substack{0<k\le K\\ 0<l\le L}} \alpha_{k,l} e\Big(\frac{\nu kl}{d}\Big)\Big| \le 150 D^{1/2}(D+KL)^{1/2}\|\widetilde{\alpha}\|.$$

Next we introduce the condition $(d,l)=1$. This will cost us an additional factor $\log 3D$ in the upper bound. Indeed, relaxing the condition $(d,l)=1$ by Möbius inversion, we find that the restricted sum is bounded by

$$\sum_{b\le D} \varrho(b) \sum_{d\le D/b} \sum_{\nu^2+1\equiv 0 \,(\mathrm{mod}\, d)} \Big| \sum_{\substack{0<k\le K\\ 0<l\le L/b}} \alpha_{k,bl} e\Big(\frac{\nu kl}{d}\Big)\Big| \le 150\Big(\sum_{b\le D} \varrho(b) b^{-1}\Big) D^{1/2}(D+KL)^{1/2}\|\widetilde{\alpha}\|$$

where $\varrho(b)$ is the number of solutions to $\nu^2+1 \equiv 0 \pmod b$. We have

$$\varrho(b) \le \sum_{c|b} \chi(c),$$

and

$$\sum_{b\le D} \varrho(b) b^{-1} \le \sum_{d\le D} d^{-1} \sum_{c\le D/d} \chi(c) c^{-1} < \sum_{d\le D} d^{-1} < \log 3D;$$

therefore

$$\sum_{d\le D} \sum_{\nu^2+1\equiv 0 \,(\mathrm{mod}\, d)} \Big| \sum_{0<k\le K} \sum_{\substack{0<l\le L\\ (l,d)=1}} \alpha_{k,l} e\Big(\frac{\nu kl}{d}\Big)\Big| \le 150(\log 3D) D^{1/2}(D+KL)^{1/2}\|\widetilde{\alpha}\|. \tag{2.2}$$

Finally, we shall replace $e(\nu kl/d)$ in (2.2) by the arithmetic function

$$\varrho_{k,l}(d) = \sum_{\nu^2+l^2\equiv 0 \,(\mathrm{mod}\, d)} e(\nu k/d). \tag{2.3}$$

This function serves as a "harmonic" à la Weyl for the equidistribution of roots of the congruence $\nu^2 + l^2 \equiv 0 \pmod d$. From (2.2) we derive

Lemma 3. *For any complex numbers $\alpha_{k,l}$ we have*

$$\sum_{d\le D}\Big|\sum_{0<k\le K}\sum_{0<l\le L}\alpha_{k,l}\varrho_{k,l}(d)\Big| \le 150(\log 3D)^3 D^{1/2}(D+KL)^{1/2}\|\widetilde{\alpha}\|.$$

Proof. First we relate $\varrho_{k,l}(d)$ to $\varrho_{kl,1}(d)$. These are equal if $(d,l)=1$. In general we write $(d,l^2)=ab^2$ where a is squarefree so $d=ab^2d_1$, $l=abl_1$ and $(d_1,al_1)=1$. The congruence $\nu^2+l^2\equiv 0 \pmod d$ reduces to $\nu_1^2+l_1^2\equiv 0 \pmod{d_1}$ after the substitution $\nu=ab\nu_1$ and division by a^2b^2. Hence

$$\varrho_{k,l}(d)=\sum_{\substack{\nu_1 \,(\mathrm{mod}\, bd_1)\\ \nu_1^2+l_1^2\equiv 0\,(\mathrm{mod}\, d_1)}} e(\nu_1 k/(bd_1)).$$

This sum vanishes unless $k=bk_1$, in which case we obtain

$$\varrho_{k,l}(d)=b\sum_{\substack{\nu_1 \,(\mathrm{mod}\, d_1)\\ \nu_1^2+l_1^2\equiv 0\,(\mathrm{mod}\, d_1)}} e(\nu_1 k_1/d_1)=b\varrho_{k_1l_1,1}(d_1) \tag{2.4}$$

by changing ν_1 into $\nu_1 l_1$ modulo d_1 and dividing the new congruence by l_1^2.

By (2.4) it follows that the sum in Lemma 3 is majorized by

$$\sum_{ab^2d\le D} b \sum_{\nu^2+1\equiv 0\,(\mathrm{mod}\, d)}\Big|\sum_{0<k\le K/b}\sum_{\substack{0<l\le L/(ab)\\ (l,d)=1}}\alpha_{bk,abl}\,e\Big(\frac{\nu kl}{d}\Big)\Big|.$$

Hence we obtain the same bound as (2.2) but with the extra factor

$$\sum_{ab^2\le D} b(ab^2)^{-1} < (\log 3D)^2.$$

This completes the proof of Lemma 3.

Before concluding this section we offer a slight generalization of Lemma 3 with

$$\varrho_{k,l}(d;q,a)=\sum_{\substack{\nu^2+l^2\equiv 0\,(\mathrm{mod}\, d)\\ \nu^2+l^2\equiv a\,(\mathrm{mod}\, q)}} e(\nu k/(dq)) \tag{2.3$'$}$$

in place of $\varrho_{k,l}(d)$ where $(a,q)=1$. Here q must be relatively small since our estimate will be weakened by a factor q^3 (for the actual applications we have in mind even a fixed q would be useful). Precisely, one can derive from (2.1) in the same fashion as Lemma 3 the following:

Lemma 3$'$. *For any complex numbers $\alpha_{k,l}$ we have*

$$\sum_{d\le D}\Big|\sum_{\substack{0<k\le K\\ 0<l\le L}}\alpha_{k,l}\varrho_{k,l}(d;q,a)\Big| \le 150(q\log 3D)^3 D^{1/2}(D+KL)^{1/2}\|\widetilde{\alpha}\|.$$

P r o o f (sketch). Note that the sum (2.3′) is void unless $(d,q)=1$, in which case it factors into

$$\varrho_{k,l}(d;q,a)=\sum_{\nu^2+l^2\equiv 0 \,(\mathrm{mod}\, d)} e(\nu k\overline{q}/d)\sum_{\omega^2+l^2\equiv a \,(\mathrm{mod}\, q)} e(\omega k\overline{d}/q).$$

The second sum depends on the residue classes of k, l modulo q but not on k, l in any other way. Let k_0, l_0 denote these classes. The first sum is just $\varrho_{k\overline{q},l}(d)$. Assuming $(d,l)=1$ (as we can by applying (2.4)) we have

$$\varrho_{k\overline{q},l}(d)=\varrho_{kl\overline{q},1}(d)=\sum_{\nu^2+1\equiv 0 \,(\mathrm{mod}\, d)} e\Big(\frac{\nu kl\overline{q}}{d}\Big).$$

Now we remove the condition $(d,l)=1$ by Möbius inversion (the same device was used for (2.2)). After this we write $kl=nq+n_0$ where n_0 is the fixed residue class of kl modulo q (i.e. $n_0\equiv k_0l_0 \pmod q$) to get

$$\sum_{\nu^2+1\equiv 0 \,(\mathrm{mod}\, d)} e\Big(\frac{\nu n_0\overline{q}}{d}\Big)e\Big(\frac{\nu n}{d}\Big).$$

Here the second exponential is free of $\overline{q}$ (recall that $\overline{q}$ stands for the multiplicative inverse of q modulo d). Now (2.1) can be applied for

$$\alpha_n=\sum_{kl=nq+n_0}\alpha_{k,l}$$

giving the same results as before. The above operations are performed on each partial sum restricted by the residue classes k_0, l_0 and ω such that $\omega^2+l_0^2\equiv a \pmod q$. Since the number of such partial sums does not exceed q^3, we multiply by q^3 to get the bound for the whole original sum.

R e m a r k. One could establish stronger estimates with respect to q but the resulting refinement is not significant to produce new applications.

3. The remainder term. Given complex numbers λ_l with $l\ge 1$ we consider the sequence

$$a_n=\sum_{l^2+m^2=n}\lambda_l \tag{3.1}$$

with the intention of applying sieve methods (a combinatorial device of exclusion-inclusion). This will lead us to the problem of estimating sums of the type

$$A_d(x)=\sum_{\substack{n\le x\\ n\equiv 0 \,(\mathrm{mod}\, d)}} a_n \tag{3.2}$$

for various $d \geq 1$. We expect that the main term for $A_d(x)$ is

$$M_d(x) = \frac{1}{d} \sum_{l^2+m^2 \leq x} \lambda_l \varrho_l(d) \tag{3.3}$$

where $\varrho_l(d) = \varrho_{0,l}(d)$ denotes the number of roots of $\nu^2 + l^2 \equiv 0 \pmod{d}$ (see (2.3)). By (2.4) we have $\varrho_{0,l}(d) = b\varrho(d/(ab^2))$ where $ab^2 = (d, l^2)$ with a squarefree. Let $r = r(d)$ be the largest integer such that $r^2 \mid d$. We have $b = (r(d), l)$ and

$$\varrho_l(d) = (r(d), l)\varrho(d/(d, l^2)). \tag{3.4}$$

We define

$$R_d(x) = A_d(x) - M_d(x), \tag{3.5}$$

which we expect to be a small error term. Estimating trivially one gets

$$|R_d(x)| \leq 4x^{1/2}d^{-1} \sum_l |\lambda_l| \varrho_l(d) \tag{3.6}$$

if $d \leq x^{1/2}$. Only a slightly better bound would suffice for applications; however, we need the relevant improvements in a large range of d. We call

$$R(x, D) = \sum_{d \leq D} |R_d(x)| \tag{3.7}$$

the *remainder term*, and we prove the following:

LEMMA 4. *Let λ_l be any complex numbers for $1 \leq l \leq \sqrt{x}$. Then for $1 \leq D \leq x$,*

$$R(x, D) \ll \|\lambda\| D^{1/4} x^{1/2+\varepsilon} \tag{3.8}$$

with any $\varepsilon > 0$, the implied constant depending only on ε. Here

$$\|\lambda\| = \Big(\sum_l |\lambda_l|^2 \Big)^{1/2}.$$

Remarks. Estimating trivially one gets $R(x, D) \ll \|\lambda\|_1 x^{1/2+\varepsilon}$ with

$$\|\lambda\|_1 = \sum_l |\lambda_l|.$$

However, for applications we need $R(x, D) \ll \|\lambda\|_1 x^{1/2-\varepsilon}$; therefore our result (3.8) beats this for D as large as $D = \|\lambda\|_1^4 \|\lambda\|^{-4} x^{-8\varepsilon}$. If the sequence λ_l is not sparse in the sense that

$$\|\lambda\|_1 \gg \|\lambda\| x^{1/4-\varepsilon} \tag{3.9}$$

we have a satisfactory bound for the remainder term $R(x, D)$ with $D = x^{1-12\varepsilon}$.

Lemma 4 will be derived from a stronger result for sums of type

$$A_d(f) = \sum_{n \equiv 0 \,(\mathrm{mod}\, d)} a_n f(n) \tag{3.10}$$

where f is a smooth function on $\mathbb{R}^+$ such that

$$f(u) = 0 \quad \text{if } u \geq x, \tag{3.11}$$

$$f^{(j)}(u) \ll \Delta^j \quad \text{if } 1 \leq u < x, \tag{3.12}$$

with some Δ subject to $x^{-1} \leq \Delta \leq 1$, the implied constant depending on j only. Splitting the summation into residue classes (mod d) and applying Poisson's formula we derive

$$\begin{aligned} A_d(f) &= \sum_l \lambda_l \sum_{\nu^2 + l^2 \equiv 0 \,(\mathrm{mod}\, d)} \sum_{m \equiv \nu \,(\mathrm{mod}\, d)} f(l^2 + m^2) \\ &= \frac{1}{d} \sum_k \sum_l \lambda_l \varrho_{k,l}(d) F_l(k/d) \end{aligned}$$

where

$$F_l(z) = \int_{-\infty}^{\infty} f(l^2 + t^2) e(-zt)\, dt.$$

The zero frequency ($k = 0$) yields

$$M_d(f) = \frac{1}{d} \sum_l \lambda_l \varrho_l(d) F_l(0), \tag{3.13}$$

which we regard as the main term for $A_d(f)$. Here

$$F_l(0) = \int_{-\infty}^{\infty} f(l^2 + t^2)\, dt. \tag{3.14}$$

Subtracting we define the error term as

$$R_d(f) = A_d(f) - M_d(f). \tag{3.15}$$

Lemma 5. *Let λ_l be any complex numbers for $1 \leq l \leq \sqrt{x}$ and f be a smooth function supported on $[0, x]$ whose derivatives satisfy* (3.12) *with $x^{-1} \leq \Delta \leq 1$. Then for $1 \leq D \leq x$,*

$$\sum_{d \leq D} |R_d(f)| \ll \|\lambda\| \Delta D^{1/2} x^{5/4+\varepsilon} \tag{3.16}$$

with any $\varepsilon > 0$, the implied constant depending only on ε.

Remarks. The sharpest result is obtained when we have the best possible smoothing. This allows $\Delta = x^{-1}$. In that case Lemma 5 becomes

$$\sum_{d \leq D} |R_d(f)| \ll \|\lambda\| D^{1/2} x^{1/4+\varepsilon} \tag{3.17}$$

while the trivial argument yields $O(\|\lambda\|_1 x^{1/2+\varepsilon})$. Applications require $O(\|\lambda\|_1 x^{1/2-\varepsilon})$; therefore our result (3.17) satisfies this bound for $D = \|\lambda\|_1^2 \|\lambda\|^{-2} x^{1/2-4\varepsilon}$. In particular, if λ_l is the characteristic sequence of a set $\mathcal{L} \subset [1, \sqrt{x}\,]$ we have a satisfactory estimate for the smoothed remainder term of level

$$D = |\mathcal{L}| x^{1/2-4\varepsilon} \tag{3.18}$$

where $|\mathcal{L}|$ denotes the number of elements of $\mathcal{L}$.

Proof of Lemma 5. The error terms have the Fourier expansion

$$R_d(f) = \frac{2}{d} \sum_{k=1}^{\infty} \sum_{l} \lambda_l \varrho_{k,l}(d) F_l(k/d).$$

We can truncate the series over k at some point because it converges rapidly. To determine this point we give an estimate for the Fourier transform $F_l(z)$. First by j-fold partial integration we write

$$F_l(z) = (2\pi i z)^{-j} \int_{-\sqrt{x}}^{\sqrt{x}} e(-zt)(\partial^j/\partial t^j) f(l^2 + t^2)\, dt,$$

then we estimate the partial derivative

$$(\partial^j/\partial t^j) f(l^2 + t^2) = \sum_{0 \le 2i \le j} c_{ij} t^{j-2i} f^{(j-i)}(l^2 + t^2) \ll (\Delta\sqrt{x})^j$$

and we get

$$F_l(z) \ll \sqrt{x}(\Delta\sqrt{x}/z)^j$$

for all $z > 0$ with any $j \ge 0$. Since $1 \le d \le D$ the above estimate yields

$$F_l(k/d) \ll k^{-2} D^{-1} \quad \text{if } k \ge K = \Delta D x^{1/2+\varepsilon}$$

by choosing $j = j(\varepsilon)$ sufficiently large. Hence the tail of the Fourier series for $R_d(f)$ over $k \ge K$ is negligible; it contributes $O(\varrho(d) d^{-1} \|\lambda\|_1)$. The remaining double sum over k, l is an imitation of that in Lemma 3 with the coefficients $\alpha_{k,l} = \lambda_l F_l(k/d)$. However, these must not depend on d. To separate d from k, l in $F_l(k/d)$ we change the variable of integration,

$$F_l(k/d) = 2\sqrt{x} k^{-1} \int_0^{\infty} f(l^2 + x v^2 k^{-2}) \cos(2\pi v \sqrt{x}/d)\, dv.$$

Note that $k > v$ and $l < \sqrt{x}$ or else the integrand vanishes. Hence we derive

$$d|R_d(f)| \le 4\sqrt{x} \int_0^K \Big| \sum_{\substack{v<k<K \\ 0<l<\sqrt{x}}} \lambda_l k^{-1} f(l^2 + x v^2 k^{-2}) \varrho_{k,l}(d) \Big|\, dv + O(\varrho(d)\|\lambda\|_1).$$

Now Lemma 3 can be applied with $\alpha_{k,l} = \lambda_l k^{-1} f(l^2 + xv^2k^{-2})$. We have

$$\|\widetilde{\alpha}\|^2 \ll \|\widetilde{\lambda}\|^2 \sum_{k>v} \tau(k)k^{-2} \ll \|\widetilde{\lambda}\|^2 \frac{\log(v+2)}{v+1}$$

where

$$\|\widetilde{\lambda}\|^2 = \sum_l |\lambda_l|^2 \tau(l),$$

and

$$\int_0^K \left(\frac{\log(v+2)}{v+1}\right)^{1/2} dv \ll (K\log(K+1))^{1/2}.$$

Therefore Lemma 3 yields

$$\sum_{d\le D} d|R_d(f)| \ll \|\widetilde{\lambda}\|(D + K\sqrt{x})^{1/2}(DKx)^{1/2}(\log x)^4.$$

This implies the inequality of Lemma 5.

We derive Lemma 4 from Lemma 5 by comparing $R_d(x)$ with $R_d(f)$ for f such that

$$\begin{aligned} f(u) &= 1 \quad \text{if } 0 < u \le x - y, \\ f^{(j)}(u) &\ll y^{-j} \quad \text{if } x - y < u < x, \\ f(u) &= 0 \quad \text{if } u \ge x, \end{aligned}$$

and then we choose the y which minimizes the resulting bound. All terms of $A_d(x)$ agree with those of $A_d(f)$ except for $x - y < n \le x$. In this short segment we estimate trivially as follows:

$$\begin{aligned} \sum_d |A_d(x) - A_d(f)| &\le \sum_{x-y<n\le x} |a_n|\tau(n) \ll x^\varepsilon \sum_{x-y<l^2+m^2\le x} |\lambda_l| \\ &\ll yx^\varepsilon \sum_{l\le\sqrt{x}} |\lambda_l|(x+y-l^2)^{-1/2} \\ &\ll \|\lambda\| yx^\varepsilon \Big(\sum_{l\le\sqrt{x}} (x+y-l^2)^{-1}\Big)^{1/2} \\ &\ll \|\lambda\|(y^{1/2} + yx^{-1/4})x^\varepsilon. \end{aligned}$$

With the main terms $M_d(x)$, $M_d(f)$ we argue similarly and obtain the same estimate. Combining both with the estimate of Lemma 5 (f satisfies (3.12) with $\Delta = y^{-1}$) we deduce

$$\sum_{d\le D} |R_d(x)| \ll \|\lambda\|(y^{-1}D^{1/2}x^{5/4} + y^{1/2} + yx^{-1/4})x^\varepsilon.$$

Finally, choosing $y = D^{1/4}x^{3/4}$ we complete the proof of Lemma 4.

R e m a r k. If λ is supported on a thin set the above argument is wasteful but it can be improved.

As we did at the end of the previous section, we can generalize these results to cover any fixed residue class $a \pmod q$ with $(a,q)=1$. First we consider smoothed sums

$$A_d(f;q,a) = \sum_{\substack{n\equiv 0 \,(\mathrm{mod}\, d)\\ n\equiv a \,(\mathrm{mod}\, q)}} a_n f(n) \tag{3.10$'$}$$

for which the main term is

$$M_d(f;q,a) = \frac{1}{dq}\sum_l \lambda_l \varrho_l(d;q,a) F_l(0). \tag{3.13$'$}$$

Here $\varrho_l(d;q,a) = \varrho_{0,l}(d;q,a)$ is the number of solutions to the system of congruences

$$\nu^2 + l^2 \equiv 0 \pmod d, \qquad \nu^2 + l^2 \equiv a \pmod q$$

(see (2.3$'$)). Since $(d,q)=1$, or else both $A_d(f;q,a)$ and $M_d(f;q,a)$ would vanish, we have

$$\varrho_l(d;q,a) = \varrho_l(d)\varrho_l(q,a) \tag{3.19}$$

where the second factor is the number of solutions to the second congruence above (it does not depend on d). Recall that the first factor can be expressed by the simpler function $\varrho(d)$ (see (3.4)). The error term

$$R_d(f;q,a) = A_d(f;q,a) - M_d(f;q,a) \tag{3.15$'$}$$

has the Fourier expansion

$$R_d(f;q,a) = \frac{2}{dq}\sum_{k=1}^{\infty}\sum_l \lambda_l \varrho_{k,l}(d;q,a) F_l(k/(dq))$$

by Poisson summation as in the proof of Lemma 5. The remaining arguments are identical as before except that we use Lemma 3$'$ rather than Lemma 3 at the very end. We obtain

Lemma 5$'$. *Let λ_l and f be as in Lemma 5. Let $q \geq 1$ and $(a,q)=1$. Then for $1 \leq D \leq x$,*

$$\sum_{d\leq D} |R_d(f;q,a)| \ll \|\lambda\| \Delta D^{1/2} x^{5/4+\varepsilon} \tag{3.16$'$}$$

with any $\varepsilon > 0$, the implied constant depending on ε and q only.

We shall demonstrate the strength of our estimates for the remainder terms with selected applications in Sections 5 and 6 before employing these for the proof of the main Theorem 1.

4. Digressions on convolution sequences. In number theory we often ask if one set of integers meets another, i.e. we want to know if the equation $m = n$ has solutions in m, n from the prescribed sets. When counting these solutions one may as well evaluate the sum

$$S = \sum_n a_n b_n f(n)$$

where a_n, b_n (the multiplicities) are arithmetic functions supported on the prescribed sets in question, and f is a suitably chosen smooth test function. We can write

$$S = \sum_{m=n} \sum a_m b_n g(m) h(n)$$

with $gh = f$ and try to relax the equation $m = n$ by means of some kind of harmonics. For instance, we can detect this equation with the integral

$$\int_0^1 e(\alpha(m-n))\, d\alpha = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{otherwise,} \end{cases}$$

getting

$$S = \int_0^1 \Big(\sum_m a_m g(m) e(\alpha m)\Big) \Big(\sum_n b_n h(n) e(-\alpha n)\Big) d\alpha.$$

In this way the desired twisting of a_n with b_n is diverted to twisting either one with additive characters, and the latter problem can be considerably more approachable, especially so if both sequences $a = (a_m)$, $b = (b_n)$ have intrinsic additive properties. This idea lies at the foundation of the circle method; it led (through sophisticated refinements) to solutions of very attractive problems in additive number theory.

However, for this paper we have in mind sequences $b = (b_n)$ which conceal some multiplicative properties. Analytic number theory supplies a variety of adequate harmonics, either classical (the Dirichlet characters) or modern ones (the Fourier coefficients of automorphic forms). One can also do quite well with elementary considerations if the sequence $b = (b_n)$ is of convolution type, say for example

$$b_n = \sum_{d|n} \gamma_d. \tag{4.1}$$

Unfolding the convolution and interchanging the order of summation we get

$$S = \sum_d \gamma_d A_d(f)$$

where

$$A_d(f) = \sum_{n \equiv 0 \,(\mathrm{mod}\, d)} a_n f(n).$$

We assume we have some knowledge of $A_d(f)$, namely that it approximates very well a simpler sum $M_d(f)$. Thus we expect that the error term $R_d(f) = A_d(f) - M_d(f)$ is small. Now, replacing $A_d(f)$ by the expected main term $M_d(f)$ we get

$$S = \sum_d \gamma_d M_d(f) + R$$

where R is the remainder

$$R = \sum_d \gamma_d R_d(f).$$

This scheme is interesting if we can control the support of $\gamma = (\gamma_d)$. Suppose $\gamma_d = 0$ if $d > D$ and $|\gamma_d| \le 1$ for $1 \le d \le D$; then $|R|$ is bounded by

$$R(f, D) = \sum_{d \le D} |R_d(f)|,$$

which we call the remainder term *of level* D. The latter can be estimated successfully provided D is not too large. On the other hand, the main term

$$T = \sum_d \gamma_d M_d(f)$$

can be arranged as a sum of multiplicative functions so its evaluation can be performed routinely by means of associated zeta-functions (see, for example, the proof of Theorem 6).

5. Representations by the norm of ideals. In practice the exact convolution shape (4.1) rarely occurs so it is necessary to employ a bit of imagination to furnish (4.1) from a given sequence. Consider the function

$$b(n) = \sum_{\substack{\mathfrak{a} \subset K \\ N\mathfrak{a} = n}} 1 \tag{5.1}$$

which is the number of integral ideals $\mathfrak{a}$ in a field K of norm $N\mathfrak{a} = n$. The generating Dirichlet series of these numbers is the Dedekind zeta-function

$$\zeta_K(s) = \sum_{n=1}^{\infty} b(n) n^{-s} = \sum_{\mathfrak{a}} (N\mathfrak{a})^{-s}.$$

Suppose $K/\mathbb{Q}$ is abelian of degree $g \ge 2$ and discriminant $\pm q$. Then $\zeta_K(s)$ factors into Dirichlet L-functions

$$\zeta_K(s) = \prod_{1 \le j \le g} L(s, \chi_j) \tag{5.2}$$

where χ_j are distinct primitive characters of conductors q_j such that $q_1 \dots q_g$

$= q$ and exactly one is trivial ($\zeta_K(s)$ has a simple pole at $s = 1$). Hence

$$b(n) = \sum_{n_1 \dots n_g = n} \chi_1(n_1) \dots \chi_g(n_g). \tag{5.3}$$

We shall establish an asymptotic formula for the sum

$$S = \sum_{(n,q)=1} a_n b(n) f(n) \tag{5.4}$$

where the a_n are given by (3.1) and $f(u)$ is a smooth function supported on $x \le u \le 2x$ with derivatives $f^{(j)}(u) \ll x^{-j}$. For convenience we break S into sums over reduced residue classes

$$S(a) = \sum_{n \equiv a \,(\mathrm{mod}\, q)} a_n b(n) f(n). \tag{5.5}$$

Note that the summation is void unless $a \equiv N\mathfrak{a} \pmod q$ for some $\mathfrak{a} \subset K$. Such residue classes form a group, say $\mathcal{H}$, which has index g in $(\mathbb{Z}/q\mathbb{Z})^*$.

We split the formula (5.3) for $b(n)$ by applying a smooth partition of unity to each of the variables $n_1, \dots, n_g$. Let φ, ψ be smooth functions on $\mathbb{R}^+$ such that

$$0 \le \varphi, \psi \le 1, \quad \varphi + \psi = 1,$$
$$\varphi(u) = 0 \text{ if } u \ge 2z \quad \text{and} \quad \psi(u) = 0 \text{ if } u \le z,$$

for some z. By successive application of $\varphi + \psi = 1$ we arrive at

$$b(n) = \sum_{1 \le j \le g} b_j(n) + b'(n) \tag{5.6}$$

where

$$b_j(n) = \sum_{n_1 \dots n_g = n} \chi_1(n_1) \dots \chi_g(n_g) \varphi(n_1) \dots \varphi(n_{j-1}) \psi(n_j)$$

and

$$b'(n) = \sum_{n_1 \dots n_g = n} \chi_1(n_1) \dots \chi_g(n_g) \varphi(n_1) \dots \varphi(n_g).$$

Note that $b'(n) = 0$ if $n \ge (2z)^g$ so we choose

$$(2z)^g = x \tag{5.7}$$

to kill the term $b'(n)$ in (5.6). Considering (5.6) we split

$$S(a) = \sum_{1 \le j \le g} S_j(a) \tag{5.8}$$

where

$$S_j(a) = \sum_{n \equiv a \,(\mathrm{mod}\, q)} a_n b_j(n) f(n). \tag{5.9}$$

Next we write $b_j(n)$ in a convolution form. To accomplish this we single out the variable n_j, for which we replace $\chi_j(n_j)$ by

$$\chi_j(n_j) = \chi_j(a) \prod_{i \neq j} \overline{\chi}_j(n_i).$$

This yields

$$b_j(n) = \sum_{d|n} \gamma_j(d) \psi(n/d)$$

with

$$\gamma_j(d) = \chi_j(a) \sum_{n_1 \dots \hat{n}_j \dots n_g = d} \Big(\prod_{i \neq j} \chi_i \overline{\chi}_j(n_i) \Big) \Big(\prod_{i<j} \varphi(n_i) \Big) \tag{5.10}$$

(the hat over a variable indicates the variable is deleted). Observe that $\gamma_j(d) \ll d^{\varepsilon}$. Hence by changing the order of summation

$$S_j(a) = \sum_{\substack{(d,q)=1 \\ d<D}} \gamma_j(d) \sum_{\substack{n \equiv 0 \,(\mathrm{mod}\, d) \\ n \equiv a \,(\mathrm{mod}\, q)}} a_n \psi(n/d) f(n)$$

where

$$D = xz^{-1}. \tag{5.11}$$

Here the condition $d < D$ is redundant since $\psi(n/d) f(n) = 0$ if $d \geq D$; nevertheless we shall display this condition when estimating the remainder term. According to (3.13′) the main term for the inner sum in $S_j(a)$ is

$$M_d(f; q, a) = \frac{1}{dq} \sum_l \lambda_l \varrho_l(d; q, a) \int_{-\infty}^{\infty} \psi\Big(\frac{l^2 + t^2}{d} \Big) f(l^2 + t^2)\, dt.$$

We estimate the resulting error terms by an appeal to (3.16′) (with $\Delta = x^{-1}$) and obtain

$$S_j(a) = \sum_{(d,q)=1} \gamma_j(d) M_d(f; q, a) + O(\|\lambda\| D^{1/2} x^{1/4+\varepsilon}).$$

One should note carefully that we have applied Lemma 5′ for the test function $\psi(n/d) f(n)$, which strictly speaking is not admissible because it depends on d (though mildly). This problem can be easily resolved by any standard method of separation of variables (choose ψ to be of a convolution type and change variables).

It remains to compute the leading term. In particular, we wish to relax its dependence on the partition of unity. First we simplify the sum over moduli. Factoring $\varrho_l(d; q, a)$ as in (3.19) we arrange

$$S_j(a) = q^{-1} \sum_l \lambda_l \varrho_l(q, a) \int_{-\infty}^{\infty} S_j(l, t) f(l^2 + t^2)\, dt + O(\|\lambda\| D^{1/2} x^{1/4+\varepsilon})$$

where

$$S_j(l,t) = \sum_{(d,q)=1} d^{-1}\gamma_j(d)\varrho_l(d)\psi\Big(\frac{l^2+t^2}{d}\Big).$$

Recall that $\varrho_l(d)$ is a multiplicative function; it is equal to $\varrho(d)$ if $(d,l)=1$, and in general is given by (3.4). But $\gamma_j(d)$ is not multiplicative because of the truncation factor $\prod_{i<j}\varphi(n_i)$ in (5.10). This obstruction can be removed at a small cost. We write

$$\prod_{i<j}\varphi(n_i) = 1 - \sum_{1\le i<j}\Big(\prod_{1\le k<i}\varphi(n_k)\Big)\psi(n_i)$$

and accordingly

$$\gamma_j(d) = \gamma_{jj}(d) - \sum_{1\le i<j}\gamma_{ij}(d),$$
$$S_j(l,t) = S_{jj}(l,t) - \sum_{1\le i<j} S_{ij}(l,t).$$

For each $1\le i<j$ the sum S_{ij} is quite small. We shall get a good bound by exploiting cancellation in the sum over n_i

$$\sum_{(n_i,q)=1} n_i^{-1}\chi_i\overline{\chi}_j(n_i)\varrho_l(d_in_i)\psi(n_i)\psi\Big(\frac{l^2+t^2}{d_in_i}\Big) \ll z^{-1/2}x^{\varepsilon} \tag{5.12}$$

where $d_i = n_1\dots\widehat{n}_i\dots\widehat{n}_j\dots n_g$. The zeta-function for this sum is $L(s,\chi_i\overline{\chi}_j)L(s,\chi_i\overline{\chi}_j\chi)$ up to an Euler product which converges absolutely in $\operatorname{Re} s > 1/2$. Assuming $2\nmid q$ both characters $\chi_i\overline{\chi}_j$ and $\chi_i\overline{\chi}_j\chi$ are non-trivial (recall that χ is the non-trivial character modulo 4) so the zeta-function is holomorphic in $\operatorname{Re} s > 1/2$, whence (5.12) follows. By (5.12) we get

$$S_{ij}(l,t) \ll z^{-1/2}x^{\varepsilon} \quad \text{if } 1\le i<j,$$

so

$$S_j(l,t) = S_{jj}(l,t) + O(z^{-1/2}x^{\varepsilon}).$$

Next we compute

$$\gamma_{jj}(d) = \chi_j(a)\sum_{n_1\dots\widehat{n}_j\dots n_g=d}\Big(\prod_{i\ne j}\chi_i\overline{\chi}_j(n_i)\Big)$$
$$= \chi_j(a)\overline{\chi}_j(d)\sum_{n_1\dots\widehat{n}_j\dots n_g=d}\chi_1(n_1)\dots\widehat{\chi}_j(n_j)\dots\chi_g(n_g),$$
$$\sum_{j=1}^{\infty}\gamma_{jj}(d)\chi_j(d)d^{-s} = \chi_j(a)\prod_{i\ne j}L(s,\chi_i) = \chi_j(a)\zeta_K(s)/L(s,\chi_j);$$

hence

$$\gamma_{jj}(d) = \chi_j(a) \sum_{m|d} \mu(d/m) \overline{\chi}_j(m) b(m).$$

Since for $(h, q) = 1$,

$$\sum_{1 \le j \le g} \chi_j(h) = \begin{cases} g & \text{if } h \pmod q \in \mathcal{H}, \\ 0 & \text{otherwise,} \end{cases}$$

we obtain

$$c(d) = \frac{1}{g} \sum_{1 \le j \le g} \gamma_{jj}(d) = \sum_{m|d} \mu(d/m) b(m) \tag{5.13}$$

provided $a \pmod q \in \mathcal{H}$ (otherwise we get nothing), and

$$S(l, t) = \sum_{1 \le j \le g} S_{jj}(l, t) = g \sum_{(d,q)=1} d^{-1} c(d) \varrho_l(d) \psi\Big(\frac{l^2 + t^2}{d}\Big).$$

Here the truncation factor $\psi\big(\frac{l^2+t^2}{d}\big)$ can be removed at a low cost. To this end we employ the associated zeta-function

$$Z_l(s) = \sum_{(d,q)=1} c(d) \varrho_l(d) d^{-s}. \tag{5.14}$$

It factors into

$$Z_l(s) = P_l(s) \zeta_K(s) L_K(s, \chi \circ N) / \zeta(s) L(s, \chi)$$

where

$$L_K(s, \chi \circ N) = \sum_{\mathfrak{a} \subset K} \chi(N\mathfrak{a}) (N\mathfrak{a})^{-s}$$

and $P_l(s)$ is an Euler product which converges absolutely in $\operatorname{Re} s > 1/2$; therefore $Z_l(s)$ is holomorphic in $\operatorname{Re} s > 1/2$ and has moderate growth. This shows that

$$S(l, t) = g Z_l(1) + O(x^{\varepsilon} (z/x)^{1/2})$$

where the error term represents the estimate for the complementary sum with ψ replaced by φ (the complementary sum ranges over $d > x/(2z)$). Check that

$$Z_l(1) = P_l(1) \frac{L_K(1, \chi \circ N)}{L(1, \chi)} \prod_{\chi_i \ne 1} L(1, \chi_i) > 0.$$

Gathering the above results we obtain

$$\begin{aligned} S(a) = g q^{-1} \sum_l \lambda_l \varrho_l(q, a) Z_l(1) F_l(0) \\ + O((\|\lambda\|_1 + \|\lambda\| x^{1/4}) x^{1/2 - 1/(2g) + \varepsilon}) \end{aligned} \tag{5.15}$$

provided $a \pmod q \in \mathcal{H}$ or else $S(a)$ vanishes. Summing over a we get

$$\omega_l(q) = \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \varrho_l(q,a) = \sum_{\omega^2+l^2 \,(\mathrm{mod}\, q)\in\mathcal{H}} 1. \tag{5.16}$$

Finally, we conclude:

THEOREM 6. *Let $K/\mathbb{Q}$ be an abelian extension of degree $g \geq 2$ and discriminant $\pm q$ which is odd. Let $b(n)$ denote the number of integral ideals in K of norm n. Then for any complex numbers λ_l and a smooth test function f supported on $[x, 2x]$ with derivatives $f^{(j)} \ll x^{-j}$ we have*

$$\begin{aligned}\sum_{(l^2+m^2,q)=1}\sum \lambda_l b(l^2+m^2) f(l^2+m^2) \\ = \sum_l \lambda_l p_l(K) \int_{-\infty}^{\infty} f(l^2+t^2)\,dt + O((\|\lambda\|_1 + \|\lambda\| x^{1/4}) x^{1/2-1/(2g)+\varepsilon})\end{aligned} \tag{5.17}$$

where $p_l(K) = gq^{-1}\omega_l(q) Z_l(1) > 0$ and with $\omega_l(q)$ defined by (5.16). *The implied constant depends only on ε and q.*

Remarks about the proof. Our introduction of the partition of unity was not only a technical device which controls the range of the moduli in the remainder term, but above all it was necessary to go through this careful argument to get the correct main term. Indeed, if we treated $b(n)$ straightforwardly as a convolution by writing $n = n_1 \dots n_g = n_1 d$, say, with the trivial character χ_1 attached to n_1 and the non-trivial characters $\chi_2, \dots, \chi_g$ attached to d, we would easily get an asymptotic result by ignoring the remainder term; however, this could be a wrong result! In particular, applying this simple-minded approach we would not be able to capture the arithmetical conditions stemming from $\omega_l(q) > 0$, i.e. that the congruence

$$\omega^2 + l^2 \equiv N\mathfrak{a} \pmod q \tag{5.18}$$

has a solution $\omega \pmod q$ for some ideal $\mathfrak{a} \subset K$ with $(\mathfrak{a}, q) = 1$.

We apply Theorem 6 for the characteristic function of a set $\mathcal{L} \subset [1, \sqrt{x}\,]$ contained in arithmetic progressions $l \pmod q$ for which $\omega_l(q) > 0$. For such a set the main term in (5.17) is $\gg |\mathcal{L}| x^{1/2}$ whereas the remainder term is $\ll |\mathcal{L}|^{1/2} x^{3/4-1/(2g)+\varepsilon}$ so the asymptotic formula is meaningful if

$$|\mathcal{L}| \gg x^{1/2-1/g+\varepsilon}. \tag{5.19}$$

Therefore for any set $\mathcal{L}$ satisfying the local conditions (5.18) and of cardinality (5.19) with $x > x_0(\varepsilon, q)$ there are integral ideals $\mathfrak{a} \subset K$ of norm $x < N\mathfrak{a} < 2x$ such that

$$N\mathfrak{a} = m^2 + l^2 \quad \text{with } l \in \mathcal{L}, \tag{5.20}$$

and we have an asymptotic formula for the number of the above representations.

Next we shall swap the asymptotic (5.17) for a lower bound (of correct order of magnitude) by applying a sieve method of Brun type to produce results which are valid for sets $\mathcal{L}$ thinner than those satisfying (5.19). Let H be the collection of arithmetic progressions modulo q which are represented by norms of ideals prime to q,

$$H = \{h \in \mathbb{Z} : h \pmod q \in \mathcal{H}\}. \tag{5.21}$$

The sieve method works nicely with the function

$$b^*(n) = \begin{cases} 1 & \text{if } p \mid n \Rightarrow p \in H, \\ 0 & \text{otherwise,} \end{cases} \tag{5.22}$$

in place of $b(n)$. Recall that the primes $p \in H$ are unramified, and they split completely in $K/\mathbb{Q}$ so they are norms of prime ideals of degree 1. Therefore $b^*(n)$ is supported on norms of integral ideals. We wish to estimate

$$S^* = \sum_n a_n b^*(n) f(n)$$

(this sum takes numbers a_n for $n = N\mathfrak{a}$ without multiplicity). To this end we consider the sifting sum

$$S^*(z) = \sum_{\substack{n \in H \\ (n, P(z))=1}} a_n f(n)$$

where $P(z)$ is the product of primes $p < z$, $p \notin H$. We have

$$S^*(z) = S^* \quad \text{if } z \geq \sqrt{2x}.$$

Indeed, a number n accounted for in $S^*(z)$ has at most one bad prime factor, i.e. outside H, because $n < 2x \leq z^2$. On the other hand, the total number of bad prime factors of n (counted with multiplicity) must be even because $n \in H$; thus n has none.

The sieve method (in the context of $S^*(z)$) replaces $b^*(n)$ by a lower bound

$$b^*(n) \geq \sum_{d \mid n} \gamma_d^- \tag{5.23}$$

with certain numbers γ_d^- for $d \mid P(z)$ satisfying $|\gamma_d^-| \leq 1$ if $1 \leq d \leq D$ and $\gamma_d^- = 0$ if $d > D$ where D can be chosen at will. If D is not too small, precisely if

$$D \geq z^{\beta+\varepsilon} \tag{5.24}$$

for a certain positive number β called the *sieving limit* (β depends on the dimension of the sieve which in our case is $\kappa = 1 - g^{-1}$), and simultaneously D

is not too large for the successful estimation of the remainder term, namely

$$D \le |\mathcal{L}| x^{1/2-4\varepsilon} \tag{5.25}$$

(see (3.18)), then the sieve theory yields (see [I])

$$S^*(z) \gg x^{1/2} \sum_{l\in\mathcal{L}} \omega_l(q) V_l(z)$$

with

$$V_l(z) = \prod_{p|P(z)} (1 - \varrho_l(p)p^{-1}) \gg (\log z)^{-\kappa}.$$

We have tacitly assumed that $\mathcal{L}$ is contained in admissible residue classes modulo q, i.e. in the set

$$\mathcal{L}(K) = \{l \in \mathbb{Z} : \omega_l(q) > 0\}, \tag{5.26}$$

and the test function f is such that

$$F_l(0) = \int f(l^2 + t^2)\, dt \gg x^{1/2}$$

for any $l \le \sqrt{x}$. Suppose $\mathcal{L}$ is contained in

$$\mathcal{L}(K, \sqrt{x}) = \{1 \le l \le \sqrt{x} : \omega_l(q) > 0\}$$

and that $|\mathcal{L}| > x^{(\beta-1)/2+5\varepsilon}$ so there is room between (5.24) and (5.25) for the choice $z = \sqrt{2x}$ and $D = z^{\beta+\varepsilon}$ giving

$$S^*(\sqrt{2x}) \gg |\mathcal{L}| x^{1/2} (\log x)^{-\kappa}.$$

This establishes

THEOREM 7. *Let $\mathcal{L}$ be a subset of $\mathcal{L}(K, \sqrt{x})$ such that*

$$|\mathcal{L}| > x^{(\beta-1)/2+\varepsilon} \tag{5.27}$$

where β is the limit for the sieve of dimension $\kappa = 1 - g^{-1}$. Then

$$\sum_{l\in\mathcal{L}} \sum_{m\le\sqrt{x}} b^*(l^2 + m^2) > \eta |\mathcal{L}| x^{1/2} (\log x)^{-\kappa} \tag{5.28}$$

with some positive constant $\eta = \eta(\varepsilon, K)$ provided x is sufficiently large in terms of ε and the field K.

As an example consider a cubic normal extension $K/\mathbb{Q}$ of odd discriminant $\pm q$. In this case $g = 3$, $\kappa = 2/3$ and $\beta = 1.2242\ldots$ (see [I]), so Theorem 7 implies that any set $\mathcal{L} \subset \mathcal{L}(K, \sqrt{x}\,)$ with $|\mathcal{L}| > x^{0.1122}$ contains elements l such that $l^2 + m^2$ is the norm of an integral ideal $\mathfrak{a} \subset K$ prime to q. In particular, we can solve the equation $l^2 + m^2 = N\mathfrak{a}$ with $\mathfrak{a} \subset K$ and l a prime number, $l < m^{9/40}$, to which we referred in the introduction.

Another possibility is to take for $\mathcal{L}$ the set of biquadrates (note that the congruence $m^2 + n^8 \equiv 1 \pmod q$ has solutions). By this choice one concludes

COROLLARY. *Let $K/\mathbb{Q}$ be a cubic normal extension of odd discriminant. Then the number of solutions to*

$$N\mathfrak{a} = m^2 + n^8 \le x \tag{5.29}$$

in integral ideals $\mathfrak{a} \subset K$ and rational integers m, n has the order of magnitude $x^{5/8}(\log x)^{-2/3}$.

One can obtain comparable results for abelian fields of any degree $g \ge 2$. Here is a selection of values of the sieving limit β_κ for dimension $\kappa = 1 - g^{-1}$ (see Table 2 of [I]): $\beta_{1/2} = 1$, $\beta_{2/3} = 1.2242\ldots$, $\beta_{3/4} = 1.3981\ldots$, $\beta_{4/5} = 1.5107\ldots$, $\beta_{5/6} = 1.5884\ldots$ As κ approaches 1 for increasing degree g the sieving limit β_κ tends to 2, and the condition (5.27) requires $\mathcal{L}$ to be a set of almost full size in the logarithmic scale.

6. An application of Bombieri's sieve. Our final destination is the sum

$$P(x) = \sum_{n \le x} a_n \Lambda(n) \tag{6.1}$$

with a_n given by (3.1). Nevertheless it will be instructive to consider prior to $P(x)$ the allied sum

$$P_k(x) = \sum_{n \le x} a_n \Lambda_k(n) \tag{6.2}$$

where Λ_k is the von Mangoldt function of order k defined by

$$\Lambda_k(n) = \sum_{d \mid n} \mu(d) \left(\log \frac{n}{d}\right)^k \tag{6.3}$$

or by the recurrence formula $\Lambda_{k+1} = \Lambda_k * \Lambda + \Lambda_k \cdot L$ where L denotes the logarithm function, $L(n) = \log n$. Hence $0 \le \Lambda_k \le L^k$ and Λ_k is supported on positive integers having at most k distinct prime factors.

Since Λ_k is given by the convolution formula (6.3) one might follow the procedure described in Section 4 for $\gamma_d = \mu(d)$ (the smooth function $\left(\log \frac{n}{d}\right)^k$ can be incorporated in the procedure by partial summation). To succeed one must first reduce the support of γ_d to the level required by Lemma 4 since the error terms $R_d(x)$ are out of control for large moduli. Even if (3.9) holds, Lemma 4 does not cover the range $x^{1-\varepsilon} < d < x$.

E. Bombieri [B1, B2] has shown how to proceed in the upper range $x^{1-\varepsilon} < d < x$ provided $k > 1$. He observed that $\left(\log \frac{n}{d}\right)^k$ is relatively small in this critical range (it is still small if $k = 1$ but not enough), and he applied Selberg's sieve to take advantage of this observation. Of course, the complete argument is quite sophisticated; it requires the a_n to be real, non-negative numbers together with a few minor conditions. By virtue of

Lemma 4 we can apply Bombieri's sieve (see a new version in [FI]) to the sequence (3.1) getting

THEOREM 8. *Let λ_l be real numbers such that $0 \le \lambda_l \le 1$ and*

$$\sum_{l \le y} \lambda_l \gg y^{1-\varepsilon} \tag{6.4}$$

for any $y \ge 1$ and $\varepsilon > 0$, the implied constant depending only on ε. Then for $k \ge 2$,

$$\sum_{l^2+m^2 \le x} \lambda_l \Lambda_k(l^2+m^2) \sim k(\log x)^{k-1} \sum_{l^2+m^2 \le x} \lambda_l \psi(l) \tag{6.5}$$

as $x \to \infty$. Here $\psi(l)$ is the same as in Theorem 1.

Remark. Theorem 8 does not follow from Theorem 1 by induction on k because the latter requires a somewhat stronger condition than (6.4) (in order to neglect the error term in (1.3)).

7. Sums over primes. It was hoped at the time of its creation that the linear sieve (i.e. of dimension $\kappa = 1$) would be a tool for treating sums over primes or the allied sum

$$P(x) = \sum_{n \le x} a_n \Lambda(n), \tag{7.1}$$

but it failed for a serious reason, which is known as the parity problem. Bombieri's results [B1, B2] offer a great deal of insight into this intricate matter. The parity problem of sieve theory implies in general that any reasonable approximation to

$$A_d(x) = \sum_{\substack{n \le x \\ n \equiv 0 \,(\mathrm{mod}\, d)}} a_n \tag{7.2}$$

for all $d < x^{1-\varepsilon}$ is not sufficient to produce an asymptotic formula for $P(x)$, nor even a lower bound of the right order of magnitude. In recent work [DFI] the parity problem was resolved for the sequence

$$a_n = \sum_{\nu^2+1 \equiv 0 \,(\mathrm{mod}\, n)} e\left(\frac{\nu k}{n}\right) \tag{7.3}$$

by a subtle application of the exclusion-inclusion argument (modelled on an old idea of I. M. Vinogradov) and by adding new information to sieve theory through estimates for very special bilinear forms. Both arguments of [DFI] are tight. To the contrary in the case of our sequence (3.1) we shall enjoy a great flexibility for building bilinear forms due to the robust Lemma 4.

In this section we treat $P(x)$ for a general sequence of complex numbers a_n by an appeal to the popular identity of R. C. Vaughan [V]. Choose $y \ge 1$

and $z \geq 1$. For any $n > z$ we have

$$\Lambda(n) = \sum_{\substack{b|n \\ b\leq y}} \mu(b) \log \frac{n}{b} - \sum_{\substack{bc|n \\ b\leq y,\, c\leq z}} \mu(b)\Lambda(c) + \sum_{\substack{bc|n \\ b>y,\, c>z}} \mu(b)\Lambda(c) \tag{7.4}$$

and if $n \leq z$, the right-hand side vanishes. Suppose $x > yz$. According to Vaughan's identity, $P(x)$ splits into

$$P(x) = A(x; y, z) + B(x; y, z) + P(z) \tag{7.5}$$

where

$$A(x; y, z) = \sum_{b\leq y} \mu(b)\Big\{ A'_b(x) - A_b(x) \log b - \sum_{c\leq z} \Lambda(c) A_{bc}(x) \Big\} \tag{7.6}$$

and

$$B(x; y, z) = \sum_{\substack{bd\leq x \\ b>y}} \mu(b)\Big(\sum_{\substack{c|d \\ c>z}} \Lambda(c) \Big) a_{bd}. \tag{7.7}$$

Note that

$$|B(x; y, z)| \leq \sum_{z<d<x/y} (\log d) \Big| \sum_{y<b\leq x/d} \mu(b) a_{bd} \Big|. \tag{7.8}$$

Moreover, $A'_b(x)$ in (7.6) denotes the sum (7.2) derived from the sequence $a'_n = a_n \log n$. It can be expressed as

$$A'_b(x) = A_b(x) \log x - \int_1^x A_b(t)\, \frac{dt}{t}. \tag{7.9}$$

Now suppose $A_d(x)$ is well approximated by a sum of type

$$M_d(x) = \frac{1}{d} \sum_{n\leq x} a_n(d) \tag{7.10}$$

where the complex numbers $a_n(d)$ are somewhat simpler than the original a_n. Naturally one may assume that $a_n(1) = a_n$ but it is not necessary to do so. Define the error term

$$R_d(x) = A_d(x) - M_d(x) \tag{7.11}$$

and the remainder

$$R(x, D) = \sum_{d\leq D} |R_d(x)|. \tag{7.12}$$

Replacing $A_d(x)$ by $M_d(x) + R_d(x)$ in (7.6) and (7.9) we write

$$A(x; y, z) = M(x; y, z) + R(x; y, z) \tag{7.13}$$

where

$$M(x;y,z)=\sum_{n\le x}\sum_{b\le y}\frac{\mu(b)}{b}\Big\{a_n(b)\log\frac{n}{b}-\sum_{c\le z}\frac{\Lambda(c)}{c}a_n(bc)\Big\}$$

and

$$R(x;y,z)=\sum_{b\le y}\mu(b)\Big\{R_b(x)\log\frac{x}{b}-\int_1^x R_b(t)\frac{dt}{t}-\sum_{c\le z}\Lambda(c)R_{bc}(x)\Big\}.$$

Note that

$$|R(x;y,z)|\le R(x,yz)\log x+\int_1^x R(t,y)\,\frac{dt}{t}. \tag{7.14}$$

To proceed further with $M(x;y,z)$ we assume that every $a_n(d)$ in the main term (7.10) is a linear combination of nice multiplicative functions in d, say

$$a_n(d)=\sum_l \lambda_l(n)\varrho_l(d) \tag{7.15}$$

with $\lambda_l(n)=0$ for almost all l. Inserting these we obtain

$$M(x;y,z)=\sum_{n\le x}\sum_l \lambda_l(n)\sigma_l(n;y,z)$$

where

$$\sigma_l(n;y,z)=\sum_{b\le y}\frac{\mu(b)}{b}\Big\{\varrho_l(b)\log\frac{n}{b}-\sum_{c\le z}\frac{\Lambda(c)}{c}\varrho_l(bc)\Big\}.$$

Furthermore, we assume that each of the multiplicative functions $\varrho_l(d)$ satisfies the condition

$$\Big|\sum_{b\le y}\frac{\mu(b)}{b}\varrho_l(bc)\Big|\le(c,l)\tau(c)\Delta_l(y) \tag{7.16}$$

for all $y>1$ where $\Delta_l(y)(\log y)^2$ is decreasing. This condition implies that we can extend $\sigma_l(n;y,z)$ to an infinite series with respect to b, and its tail

$$\delta_l(n;y,z)=\sum_{b>y}\frac{\mu(b)}{b}\Big\{\varrho_l(b)\log\frac{n}{b}-\sum_{c\le z}\frac{\Lambda(c)}{c}\varrho_l(bc)\Big\} \tag{7.17}$$

is bounded by

$$\delta_l(n;y,z)\ll\Delta_l(y)\log(2lnz). \tag{7.18}$$

The complete series $\sigma_l(n;y,z)+\delta_l(n;y,z)=\psi(l)$, say, reduces to

$$\psi(l)=-\sum_b\frac{\mu(b)}{b}\varrho_l(b)\log b=\prod_p\Big(1-\frac{\varrho_l(p)}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-1}. \tag{7.19}$$

Hence we obtain

$$M(x;y,z)=\sum_{n\le x}\sum_{l}\lambda_l(n)(\psi(l)+\delta_l(n;y,z)). \tag{7.20}$$

Collecting (7.5), (7.13) and (7.20) we conclude:

PROPOSITION 9. *Suppose every function $\varrho_l(d)$ from the main terms satisfies* (7.16) *with some $\Delta_l(y)$ such that $\Delta_l(y)(\log y)^2$ is decreasing. Then for $y,z\ge 1$ and $x>yz$ we have the identity*

$$\begin{aligned}P(x)=&\sum_{n\le x}\sum_{l}\lambda_l(n)\{\psi(l)+\delta_l(n;y,z)\}\\&+B(x;y,z)+R(x;y,z)+P(z).\end{aligned} \tag{7.21}$$

Recall that $\delta_l(n;y,z)$ satisfies (7.18), $B(x;y,z)$ satisfies (7.8) and $R(x;y,z)$ satisfies (7.14).

Finally, we specialize Proposition 9 to the sequence a_n given by (3.1). We take

$$\lambda_l(n)=\sum_{l^2+m^2=n}\lambda_l.$$

We have $\varrho_l(d)=(r(d),l)\varrho(d/(d,l^2))$ (see (3.4)) so the condition (7.16) holds true with $\Delta_l(y)=c_A\tau(l)(\log y)^{-A}$ for any $A\ge 2$. This gives us the estimate (7.18) for $\delta_l(n;y,z)$. We also have the trivial bounds $P(z)\ll z$ and $R(t,y)\ll t^{1+\varepsilon}$. Combining the latter with Lemma 4 by (7.14) we obtain

$$R(x;y,z)\ll x^{1-\varepsilon/5}\quad\text{ if } yz\le x^{1-\varepsilon}.$$

Hence we conclude:

COROLLARY 10. *Let a_n be given by* (3.1) *with $|\lambda_l|\le 1$. Suppose $0<\varepsilon\le 1/3$, $x>1$, $y\ge x^{\varepsilon}$, $z\ge x^{\varepsilon}$ and $yz\le x^{1-\varepsilon}$. Then*

$$\sum_{n\le x}a_n\Lambda(n)=\sum_{l^2+m^2\le x}\lambda_l\psi(l)+B(x;y,z)+O(x(\log x)^{-A}) \tag{7.22}$$

with any $A\ge 2$, the implied constant depending only on ε and A.

8. Digressions on bilinear forms. Auxiliary transformations. The error term in (7.22) is admissible for (1.3), and the leading terms coincide. Therefore we are left with the bilinear form

$$B(x;y,z)=\sum_{z<d<x/y}\Big(\sum_{c\mid d,\,c>z}\Lambda(c)\Big)\sum_{y<b\le x/d}\mu(b)a_{bd}. \tag{8.1}$$

Its very presence in the formula for the sum over primes is indispensable in view of the parity problem of sieve theory. Of course $B(x;y,z)$ must contribute only to the error term, but proving this is the crux of the present paper.

Perhaps some of the forthcoming transformations will not be familiar to everybody so we dwell on expressing the key issues in a general context before focusing on $B(x;y,z)$. Given a matrix $A=(a_{mn})$ of complex numbers having some arithmetical nature we wish to estimate the bilinear form

$$uAv^t=\sum_m\sum_n u_m v_n a_{mn}$$

for two sequences $u=(u_m)$, $v=(v_n)$ one of which is fairly arbitrary and the other, say v, varies in a tractable manner (like a Dirichlet character, the Möbius function or a constant for example). However, the variation of a_{mn} with respect to n might be out of control. Therefore we cannot hope to execute either of the two summations directly. A standard procedure for estimating the bilinear form uAv^t is by applying Cauchy's inequality and by enlarging the outer summation so much (though not excessively) as to fill up gaps and straighten irregularities. We obtain

$$|uAv^t|\le\sum_m|u_m|\Big|\sum_n v_n a_{mn}\Big|\le\|u\|\Big(\sum_m g(m)\Big|\sum_n v_n a_{mn}\Big|^2\Big)^{1/2}$$

where $g(m)$ is a nice non-negative function with $g(m)\ge 1$ whenever $u_m\ne 0$. Two goals are achieved at once. The first is a kind of completeness in m (think spectrally); the second is a decrease in complexity of the original vector $u=(u_m)$ (compare the divisor function versus a smooth function). Reversing the order of summation we arrive at the sum

$$A(n_1,n_2)=\sum_m g(m)a_{mn_1}\overline{a}_{mn_2}.$$

This can be evaluated asymptotically with considerable uniformity in n_1, n_2. If the main term exists, say $M(n_1,n_2)$, it usually behaves nicely so the further summation

$$\sum_{n_1}\sum_{n_2} v_{n_1}\overline{v}_{n_2}M(n_1,n_2)$$

can be executed precisely, and it reduces the contribution of the main terms because the variations in signs of v_{n_1}, $\overline{v}_{n_2}$, $M(n_1,n_2)$ do not conspire (in the true setting of the method anyway).

Now we are ready to come to the point. There are situations in which the straightforward application of Cauchy's inequality is not a clever first move. Suppose the arithmetic entries a_{mn} have a hidden multiplicity with respect to n; this multiplicity carries over to the sum $A(n_1,n_2)$ making it hard to evaluate with decent uniformity in n_1,n_2. For example imagine that mn occurs with multiplicity equal to the number of representations as the sum of two squares (possibly restricted by suitable side conditions); then the summation in $A(n_1,n_2)$ amounts to counting lattice points on a 4-dimensional hyperboloid (subject to the relevant side conditions). Without

the involved multiplicity we would count integers in a segment, a much easier job indeed! In view of the above scenario one should not rush to the Cauchy inequality. Try first to unfold the suspected multiplicity using some sort of parametrization, and then pull out all but one of the involved parameter to the outer summation. Such performance will be much better not only because the multiplicities do not multiply but they can be wiped out entirely after Cauchy's inequality is applied. One can find such an arrangement in Section 5 of [DFI]. Here we arrange the bilinear form $B(x;y,z)$ in a similar manner.

Our target is the estimate

$$B(x;y,z) \ll \Delta x(\log x)^5 \tag{8.2}$$

where $\Delta = (\log x)^{-A}$ for any fixed $A \geq 5$. By (8.1) we have

$$|B(x;y,z)| \leq (\log x) \sum_{d>z} \Big| \sum_{y<b\leq x/d} \mu(b) a_{bd} \Big|. \tag{8.3}$$

In order to control the size and to separate the variables b, d (i.e. to relax the condition $bd \leq x$) we are going to break the sum into short sums of the type

$$\mathcal{B}(M,N) = \sum_{M<m\leq 2M} \Big| \sum_{N<n\leq N'} \mu(n) a_{mn} \Big| \tag{8.4}$$

where $N' = e^{\Delta} N$. Using these sums for $M = 2^j z$ and $N = e^{\Delta k} y$ we get

$$|\mathcal{B}(x;y,z)| \leq (\log x) \sum_{\substack{\Delta x < MN < x \\ M\geq z,\, N\geq y}} \sum \mathcal{B}(M,N) + O(\Delta x(\log x)^2) \tag{8.5}$$

where the error term $O(\Delta x(\log x)^2)$ represents a trivial bound for the contribution of $\mu(b)a_{bd}$ with $bd \leq 2\Delta x$ or $e^{-2\Delta} x < bd \leq x$, which terms are not covered exactly. There are fewer than $2\Delta^{-1}(\log x)^2$ short sums $\mathcal{B}(M,N)$ in (8.5) so we need to show that each of these satisfies

$$\mathcal{B}(M,N) \ll \Delta^2 x(\log x)^2. \tag{8.6}$$

Note the trivial bound

$$\mathcal{B}(M,N) \leq \sum_{M<m\leq 2M} \varrho(m) \sum_{N<n\leq N'} \varrho(n) \ll \Delta MN.$$

Let $\mathcal{B}_d(M,N)$ denote the sum (8.4) with the variables restricted by $(m,n) = d$. We have

$$\mathcal{B}(M,N) \leq \sum_{d<\Delta^{-1}} \mathcal{B}_d(M,N) + O(\Delta^2 x)$$

where the error term $O(\Delta^2 x)$ represents a trivial bound for the contribution

of $\mu(n)a_{mn}$ with $(m,n) \geq \Delta^{-1}$. Note that

$$\mathcal{B}_d(M,N) \leq \mathcal{B}_1(dM, N/d)$$

(to see this, transfer the common factor $d = (m,n)$ from n to m and ignore the property that the new m is divisible by d^2). Hence we need to show that

$$\mathcal{B}_1(M,N) \ll \Delta^3 x(\log x)^2 \tag{8.7}$$

for any M, N with $M \geq z$, $N \geq \Delta y$ and $\Delta x < MN < x$.

If $(m,n) = 1$ every representation of mn as the sum of two squares is obtained exactly four times from the representations of m and n (the multiplicity four is the number of units in $\mathbb{Z}[i]$). Thus, using Gaussian integers (think in terms of ideals) we write each a_{mn} in $\mathcal{B}_1(M,N)$ as

$$a_{mn} = \tfrac{1}{4} \sum_{|w|^2=m} \sum_{|z|^2=n} \lambda_l \tag{8.8}$$

where $z, w \in \mathbb{Z}[i]$ and $l = \operatorname{Re} z\overline{w}$ (here z has nothing in common with the real parameter z which was used throughout the former section). For notational simplicity we write

$$\lambda_l = \lambda(l), \tag{8.9}$$

$$\operatorname{Re} z\overline{w} = z \cdot w. \tag{8.10}$$

We shall often regard z, w as vectors in $\mathbb{R}^2$ so $z \cdot w$ denotes the inner product. Hence

$$4\mathcal{B}_1(M,N) \leq \sum_{M<|w|^2\leq 2M} \Big| \sum_{\substack{N<|z|^2\leq N' \\ (|z|^2,|w|^2)=1}} \mu(|z|^2)\lambda(z \cdot w) \Big|. \tag{8.11}$$

Next we relax the condition $(|z|^2, |w|^2) = 1$ by the Möbius formula

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

If n is squarefree with $r \,|\, n$, then all solutions to the equation $|z|^2 = n$ are accounted for exactly four times when counting the solutions to the system of equations

$$|z|^2 = n, \quad |\zeta|^2 = r \quad \text{with } \zeta \,|\, z,$$

and, of course, this system has no solutions if $r \nmid n$. Therefore the innermost sum in (8.11) is equal to

$$\tfrac{1}{4} \sum_{r||w|^2} \mu(r) \sum_{|\zeta|^2=r} \sum_{N<r|z|^2\leq N'} \mu(r|z|^2)\lambda(\zeta z \cdot w)$$

by changing z to ζz. Inserting this into (8.11) and changing ζw to w we

arrive at

$$4\mathcal{B}_1(M,N) \le \sum_r \varrho(r) \sum_{\substack{rM<|w|^2\le 2rM\\ r^2\mid\mid |w|^2}} \Big| \sum_{N<r|z|^2\le N'} \mu(r|z|^2)\lambda(z\cdot w)\Big|$$

(we have estimated the number of ζ's by $4\varrho(r)$ simply ignoring the condition $\zeta \mid w$). Estimating trivially we find that the terms with $r \ge \Delta^{-2}$ contribute

$$\ll \Delta MN \sum_{r>\Delta^{-2}} \varrho(r)^2 r^{-2} \ll \Delta^3 x(\log x)^2.$$

In the remaining terms we ignore the condition $r^2 \mid |w|^2$ and obtain

$$4\mathcal{B}_1(M,N) \le \sum_{r<\Delta^{-2}} \varrho(r) \sum_{rM<|w|^2\le 2rM} \Big| \sum_{N<r|z|^2\le N'} \mu(r|z|^2)\lambda(z\cdot w)\Big| + O(\Delta^3 x(\log x))^2. \tag{8.12}$$

Put

$$\mathcal{C}_r(M,N) = \sum_{M<|w|^2\le 2M} \Big| \sum_{N<|z|^2\le N'} \mu(r|z|^2)\lambda(z\cdot w)\Big|. \tag{8.13}$$

We need to show that

$$\mathcal{C}_r(M,N) \ll \Delta^5 x(\log x)^2 \tag{8.14}$$

for every r, M, N with $r<\Delta^{-2}$, $M\ge z$, $N\ge \Delta^3 y$ and $\Delta x < MN < x$.

A Gaussian integer $w=u+iv$ is said to be *primitive* if $(u,v)=1$. We put

$$\mathcal{C}_{cr}(M,N) = \sum_{M<|w|^2\le 2M}^{*} \Big| \sum_{N<|z|^2\le N'} \mu(r|z|^2)\lambda(cz\cdot w)\Big| \tag{8.15}$$

where $*$ restricts the summation to the primitive integers. Thus we have

$$\mathcal{C}_r(M,N) = \sum_{c\ge 1} \mathcal{C}_{cr}(c^{-2}M,N) = \sum_{1\le c\le \Delta^{-4}} \mathcal{C}_{cr}(c^{-2}M,N) + O(\Delta^5 x)$$

by the trivial bound $\mathcal{C}_{cr}(M,N) \ll \Delta MN$. Therefore it suffices to show that

$$\mathcal{C}_{cr}(M,N) \ll \Delta^5 MN \tag{8.16}$$

for every c, r, M, N with $c<\Delta^{-4}$, $r<\Delta^{-2}$, $M\ge \Delta^4 z$, $N>\Delta^3 y$ and $\Delta^5 x < MN < x$.

In the next section we shall establish estimates for bilinear forms slightly more general than (8.15). In particular, by (9.29) we obtain

$$\mathcal{C}_{cr}(M,N) \ll MN(\log N)^{-j} \tag{8.17}$$

subject to $N^\varepsilon < M < N^{1-\varepsilon}$ and $N<N'\le 2N$ uniformly in $c, r \le N$ for any ε, $j>0$, the implied constant depending only on ε, j. These conditions are

satisfied if we choose $y = x^\theta$ and $z = x^\vartheta$ with $1/2 < \theta < 1$ and $0 < \vartheta < 1-\theta$. By the estimate (8.17) we complete the proof of the main Theorem 1.

9. Bilinear forms over Gaussian integers. The following notation is in force throughout:

- z, w run over Gaussian integers,
- $z \cdot w = \operatorname{Re} z\overline{w}$ is the scalar product,
- $w = u + iv$ is primitive if $(u, v) = 1$,
- $w^* = w/(u, v)$ is the primitive kernel of w,
- $\sum^*$ denotes summation restricted to primitive Gaussian integers.

Let α, β be complex-valued functions on $\mathbb{Z}[i]$ such that

- $\alpha(z)$ is supported in the disc $|z| \le A$,
- $\beta(w)$ is supported in the annulus $B \le |w| \le 2B$.

We shall impose additional conditions on α in due course; for the time being we only assume that $A \ge B$.

For any complex valued function λ on $\mathbb{Z}$ we wish to estimate the bilinear form

$$\mathcal{C}(\alpha, \beta; \lambda) = \sum_z \sum_w{}^* \alpha(z)\beta(w)\lambda(z \cdot w). \tag{9.1}$$

Without loss of generality we assume that $\lambda(l)$ is supported in the interval $|l| \le 2AB$ so that

$$\|\lambda\|^2 = \sum_{l \in \mathbb{Z}} |\lambda(l)|^2 < \infty. \tag{9.2}$$

Naturally we set

$$\|\alpha\|^2 = \sum_{z \in \mathbb{Z}[i]} |\alpha(z)|^2, \qquad \|\beta\|^2 = \sum_{w \in \mathbb{Z}[i]} |\beta(w)|^2.$$

We begin by applying Cauchy's inequality as follows:

$$|\mathcal{C}(\alpha, \beta; \lambda)| \le \sum_l |\lambda(l)| \sum_w{}^* |\beta(w)| \Big| \sum_{z \cdot w = l} \alpha(z) \Big| \le \|\lambda\| \cdot \|\beta\| \mathcal{D}(\alpha)^{1/2} \tag{9.3}$$

where

$$\mathcal{D}(\alpha) = \sum_w{}^* g(w) \sum_l \Big| \sum_{z \cdot w = l} \alpha(z) \Big|^2. \tag{9.4}$$

Here $g(w)$ can be any non-negative function with $g(w) \ge 1$ if $B \le |w| \le 2B$. We do not need to be specific at this point; nevertheless it will be convenient to assume that $g(w)$ is radially smooth and compactly supported, say $g(w) =$

$G(|w|^2)$ where

$$0 \le G(t) \le 1, \quad G(t) = 1 \quad \text{if } B^2 \le t \le 4B^2,$$
$$\operatorname{supp} G \subset [B^2/4, 9B^2], \quad G^{(j)} \ll B^{-2j}.$$

Since l runs over all integers (without any restriction), after squaring out we obtain

$$\mathcal{D}(\alpha) = \sum_w{}^* g(w) \sum_{z \cdot w = 0} (\alpha * \alpha)(z) \tag{9.5}$$

where $\alpha * \alpha$ stands for the convolution

$$(\alpha * \alpha)(z) = \sum_{z_1 - z_2 = z} \alpha(z_1)\overline{\alpha}(z_2). \tag{9.6}$$

Note that

$$(\alpha * \alpha)(0) = \|\alpha\|^2.$$

The orthogonality relation $z \cdot w = 0$ for a primitive w in (9.5) is equivalent to the statement that z is a rational integer multiple of iw, i.e. $z = icw$ for some $c \in \mathbb{Z}$. Indeed, if $z = x + iy$ and $w = u + iv$ with $(u, v) = 1$, then $z \cdot w = ux + vy = 0$ has solutions $x = cv$, $y = -cu$ as claimed. Therefore

$$\mathcal{D}(\alpha) = \sum_{c \in \mathbb{Z}} \sum_w{}^* g(w)(\alpha * \alpha)(cw) = \mathcal{D}_0(\alpha) + 2\mathcal{D}^*(\alpha), \tag{9.7}$$

say, where $\mathcal{D}_0(\alpha)$ denotes the contribution of $c = 0$ and $\mathcal{D}^*(\alpha)$ that of all $c > 0$. Thus

$$\mathcal{D}_0(\alpha) = \|\alpha\|^2 \sum_w{}^* g(w) \ll \|\alpha\|^2 B^2 \tag{9.8}$$

and

$$\mathcal{D}^*(\alpha) = \sum_{z \ne 0} g(z^*)(\alpha * \alpha)(z) \tag{9.9}$$

where z^* denotes the primitive kernel of z (warning: even if α is supported on primitive numbers, the convolution $\alpha * \alpha$ need not be so). We trade the primitivity condition for congruence conditions by means of Möbius inversion getting

$$\mathcal{D}^*(\alpha) = \sum_{b,c>0} \sum \mu(b)\mathcal{D}(\alpha; bc) \tag{9.10}$$

where

$$\mathcal{D}(\alpha; bc) = \sum_{z \equiv 0 \,(\mathrm{mod}\, bc)} g(z/c)(\alpha * \alpha)(z). \tag{9.11}$$

Note that $|z| \le 2A$ (from the support of α) and $cB/2 < |z| < 3cB$ (from the support of g); hence $c < 4AB^{-1}$, otherwise $\mathcal{D}(\alpha; bc)$ is void. Let

$$1 \le \Lambda \le 4AB^{-1} = C, \tag{9.12}$$

say (we shall choose the best Λ later). By the trivial bound

$$\mathcal{D}(\alpha; bc) \ll \|\alpha\|^2 B^2 b^{-2}$$

we see that the terms with $b \ge \Lambda$ or $c \le C\Lambda^{-1}$ contribute to $\mathcal{D}^*(\alpha)$ at most $O(\|\alpha\|^2 AB\Lambda^{-1})$ so

$$\mathcal{D}^*(\alpha) = \sum_{b \le \Lambda} \mu(b) \sum_{C\Lambda^{-1} < c < C} \mathcal{D}(\alpha; bc) + O(\|\alpha\|^2 AB\Lambda^{-1}). \tag{9.13}$$

Now having the above restrictions for b, c at hand we no longer need to refer to the support of $g(z/c)$ to control the ranges so we can separate z from c quite freely. However, in the process of separation we must be careful not to introduce highly oscillatory factors since these would interfere with the sign change of $\alpha(z)$, which is vital at the end of our arguments (we have in mind $\alpha(z) = \mu(r|z|^2)$). There are plenty of ways to do the job. We take advantage of the convolution structure (9.6) so we represent $g(w)$ by its Fourier transform

$$g(w) = \int_{\mathbb{R}^2} f(\omega) e(\omega \cdot w)\, d\omega.$$

By Fourier inversion (spectral resolution of the Laplace operator)

$$f(\omega) = \int_{\mathbb{R}^2} g(w) e(-\omega \cdot w)\, dw.$$

Since g is radial, so is f (the Laplacian is rotation invariant!). More precisely, if $g(w) = G(|w|^2)$, then $f(\omega) = F(|\omega|^2)$ where F is the Hankel transform of G,

$$F(s) = \pi \int_0^\infty J_0(2\pi\sqrt{st}) G(t)\, dt.$$

Here $J_0(x)$ is the Bessel function. By repeated partial integration we infer

$$F(s) \ll B^2 (1 + sB^2)^{-3/2}. \tag{9.14}$$

Applying the above transformations we write

$$g(z/c) = \int_{\mathbb{R}^2} c^2 f(\omega c) e(\omega \cdot z)\, d\omega;$$

then we insert this to (9.11) getting

$$\mathcal{D}(\alpha; bc) = \int_{\mathbb{R}^2} c^2 f(\omega c) S_{bc}(\omega)\, d\omega$$

where

$$S_d(\omega) = \sum_{z\equiv 0 \,(\mathrm{mod}\, d)} (\alpha * \alpha)(z) e(\omega \cdot z) \tag{9.15}$$

$$= \sum_{z_1 \equiv z_2 \,(\mathrm{mod}\, d)} \sum \alpha(z_1)\overline{\alpha}(z_2) e(\omega \cdot (z_1 - z_2))$$

$$= \sum_{\delta \,(\mathrm{mod}\, d)} \Big| \sum_{z\equiv \delta \,(\mathrm{mod}\, d)} \alpha(z) e(\omega \cdot z)\Big|^2.$$

Here δ runs over $(\mathbb{Z}/d\mathbb{Z})^2$. Note that $S_d(\omega)$ is real, non-negative. By (9.14) we derive the following estimate:

$$c^2 f(\omega c) = c^2 F(|\omega|^2 c^2) \ll \Lambda A^2 h(\omega)$$

where $h(\omega) = (1 + |\omega|^2 A^2)^{-3/2}$. Hence

$$\mathcal{D}(\alpha; bc) \ll \Lambda A^2 \int_{\mathbb{R}^2} h(\omega) S_{bc}(\omega)\, d\omega. \tag{9.16}$$

Note that $d = bc$ (for b, c in the range of (9.13)) lies in the segment $C\Lambda^{-1} < d < C\Lambda = 4\Lambda AB^{-1} = D$, say. Putting

$$S(\omega) = \sum_{d\le D} d^2 S_d(\omega), \tag{9.17}$$

by (9.13) and (9.16) we conclude that

$$\mathcal{D}^*(\alpha) \ll \Lambda^3 B^2 \int_{\mathbb{R}^2} h(\omega) S(\omega)\, d\omega + \|\alpha\|^2 AB\Lambda^{-1}. \tag{9.18}$$

It remains to estimate $S(\omega)$. This is an attractive proposition per se so it deserves a separate section. In the next section we establish a general inequality for sums of type (9.17), which yields (see Proposition 15)

$$S(\omega) \le 2D \sum_{d\le G} d^2 S_d(\omega) + O(A^2 DG^{\varepsilon-1}\|\alpha\|^2) \tag{9.19}$$

where G can be chosen arbitrarily subject to $DG < A^{1-\varepsilon}$. We assume $B > \Lambda^7 A^\varepsilon$, choose $G = \Lambda^6$ and insert (9.19) into (9.18) getting

$$\mathcal{D}^*(\alpha) \ll \Lambda^4 AB \sum_{d\le G} d^2 \mathcal{D}_d(\alpha) + \|\alpha\|^2 AB\Lambda^{-1} \tag{9.20}$$

where

$$\mathcal{D}_d(\alpha) = \int_{\mathbb{R}^2} h(\omega) S_d(\omega)\, d\omega. \tag{9.21}$$

Next we insert (9.15) into (9.21) getting

$$\mathcal{D}_d(\alpha) = \sum_{z\equiv 0 \,(\mathrm{mod}\, d)} (\alpha * \alpha)(z) H(z) \tag{9.22}$$

where $H(z)$ is the Fourier transform of $h(\omega)$,

$$H(z) = \int_{\mathbb{R}^2} h(\omega) e(\omega \cdot z)\, d\omega = 2\pi A^{-2} \exp(-2\pi|z|/A).$$

Check this via Hankel transform with the formula (see [GR], 6.554.4)

$$\int_0^\infty J_0(xy) x (1+x^2)^{-3/2} dx = e^{-y}.$$

Hence (9.22) becomes

$$\mathcal{D}_d(\alpha) = 2\pi A^{-2} \sum_{z_1\equiv z_2 \,(\mathrm{mod}\, d)} \alpha(z_1)\overline{\alpha}(z_2) \exp(-2\pi|z_1 - z_2|A^{-1}). \tag{9.23}$$

Collecting (9.7), (9.8) and (9.20) we infer

$$\mathcal{D}(\alpha) \ll \Lambda^4 AB \sum_{d\le \Lambda^6} d^2 \mathcal{D}_d(\alpha) + \|\alpha\|^2 (B^2 + AB\Lambda^{-1}). \tag{9.24}$$

Finally, inserting (9.24) into (9.3) we obtain

PROPOSITION 11. *Suppose A, B and $\Lambda \ge 1$ are such that*

$$\Lambda^7 A^\varepsilon < B < A\Lambda^{-1} \tag{9.25}$$

for some $0 < \varepsilon < 1$. Then the bilinear form (9.1) *satisfies the following estimate*:

$$\begin{aligned} \mathcal{C}(\alpha,\beta;\Lambda) &\ll \|\alpha\| \cdot \|\beta\| \cdot \|\lambda\| (AB)^{1/2} \Lambda^{-1/2} \\ &\quad + \|\beta\| \cdot \|\lambda\| (AB)^{1/2} \Lambda^2 \Big(\sum_{d\le \Lambda^6} d^2 \mathcal{D}_d(\alpha) \Big)^{1/2} \end{aligned} \tag{9.26}$$

where the $\mathcal{D}_d(\alpha)$ are given by (9.23), *and the implied constant depends only on ε.*

Proposition 11 holds true for general sequences α, β, λ; still as it stands the estimate (9.26) is not quite ready to use because the sums $\mathcal{D}_d(\alpha)$ are there. Trivially by Cauchy's inequality one gets

$$\mathcal{D}_d(\alpha) \ll \|\alpha\|^2 d^{-2} \tag{9.27}$$

but we need slightly better bounds. A better bound could result by observing cancellation of terms due to presumed sign change of $\alpha(z)$ while the congruence $z_1 \equiv z_2 \pmod d$ is not a strong interference since the modulus is relatively small, $d \le \Lambda^6$. It is reasonable to make the following

HYPOTHESIS. *There is a $j > 0$ such that*

$$\mathcal{D}_d(\alpha) \ll \|\alpha\|^2 (\log A)^{-j} \tag{9.28}$$

for all $d \geq 1$, with the implied constant depending only on j.

Notice that (9.28) is non-trivial only for $d^2 < (\log A)^j$. Inserting (9.28) into (9.26) and choosing $\Lambda = (\log A)^{j/23}$ we derive

COROLLARY 12. *Let $\varepsilon > 0$ and $j > 0$. Suppose $A^\varepsilon < B < A(\log A)^{-j/23}$. Then assuming* (9.28) *we have*

$$\mathcal{C}_d(\alpha, \beta; \lambda) \ll \|\alpha\| \cdot \|\beta\| \cdot \|\lambda\| (AB)^{1/2} (\log A)^{-j/46} \tag{9.29}$$

where the implied constant depends on ε and j only.

As an example of an $\alpha(z)$ which satisfies (9.28) we take

$$\alpha(z) = \mu(r|z|^2). \tag{9.30}$$

This is just the Möbius function on ideals (z) prime to r such that $(z, \overline{z}) = (1)$ or $(1+i)$. In this case the hypothesis holds for any $j > 0$; it is analogous to the celebrated Siegel–Walfisz theorem in prime number theory. By modern standards the proof of (9.28) in the above case is a routine traversal throughout the zero-free region for L-functions with Hecke Grossencharacters yet it is very long in detail so we skip it entirely (see [K], [F] for related arguments).

One can verify (9.28) for $\alpha(z) = \mu(r|z|^2)$ indirectly by reducing to the more familiar result for Gaussian primes in a box and in an arithmetic progression. Though such a reduction can be made on an elementary level, it still requires considerable skill with sieve methods.

10. A mean-value theorem for Gaussian integers in arithmetic progressions. Let f be a complex-valued function on $\mathbb{Z}[i]$ supported on the disc $|z| \leq A$. Our aim is to estimate

$$\mathcal{S}_f(D) = \sum_{d \leq D} d^2 \sum_{\delta \,(\mathrm{mod}\, d)} \Big| \sum_{z \equiv \delta \,(\mathrm{mod}\, d)} f(z) \Big|^2. \tag{10.1}$$

For $f(z) = \alpha(z) e(\omega \cdot z)$ this is the sum $S(\omega)$ which emerged at the end of the previous section. We put

$$\mathfrak{S}_f(d, \delta) = \sum_{z \equiv \delta \,(\mathrm{mod}\, d)} f(z).$$

Throughout we have in mind a function $f(z)$ whose argument varies wildly so we do not expect to see main terms for any of these sums. Using additive

characters modulo d we get

$$\mathfrak{S}_f^2(d) = d^2 \sum_{\delta \,(\mathrm{mod}\, d)} |\mathfrak{S}_f(d,\delta)|^2 = \sum_{s \,(\mathrm{mod}\, d)} \Big| \sum_z f(z) e\Big(\frac{s\cdot z}{d}\Big)\Big|^2.$$

Put $s = a + bi$, $z = m + ni$ and write the fractions a/d, b/d in the lowest terms to get

$$\mathfrak{S}_f^2(d) = \sum_{\substack{k|d\\ l|d}} \sum_{\substack{a \,(\mathrm{mod}\, k)\\ b \,(\mathrm{mod}\, l)}}^{*} \Big| \sum_z f(z) e\Big(\frac{am}{k} + \frac{bn}{l}\Big)\Big|^2$$

where $*$ restricts the summations to primitive residue classes. Hence

$$\mathcal{S}_f(D) \le D \sum_{[k,l]\le D} [k,l]^{-1} \sum_{\substack{a \,(\mathrm{mod}\, k)\\ b \,(\mathrm{mod}\, l)}} \Big| \sum_z f(z) e\Big(\frac{am}{k} + \frac{bn}{l}\Big)\Big|^2$$

where $[k,l] = kl/(k,l)$ denotes the least common multiple of k, l.

If $g = [k,l]$ is small, we estimate crudely by writing the fractions a/k, b/l with common denominator g and by expanding the outer summations to all residue classes modulo g. This way we get

$$\sum_{a \,(\mathrm{mod}\, k)}^{*} \sum_{b \,(\mathrm{mod}\, l)}^{*} \Big| \sum_z f(z) e\Big(\frac{am}{k} + \frac{bn}{l}\Big)\Big|^2 \le g^2 \sum_{z_1 \equiv z_2 \,(\mathrm{mod}\, g)} f(z_1)\overline{f}(z_2) = \mathfrak{S}_f^2(g).$$

Hence the terms with $g = [k,l] \le G$, say, contribute to $\mathcal{S}_f(D)$ at most

$$D \sum_{g\le G} g^{-1} \Big(\sum_{[k,l]=g} 1\Big) \mathfrak{S}_f^2(g) \le 2D\mathcal{S}_f(G) \tag{10.2}$$

since the number of k, l with $[k,l] = g$ is less than $2g$.

For estimating the contribution of larger moduli we use the following two-dimensional large sieve inequality:

Lemma 13. *For any complex numbers c_{mn} we have*

$$\sum_{h\le H} \sum_{\substack{k\le K\\ l\le L}} \tau(hk)^{-1} \sum_{\substack{a \,(\mathrm{mod}\, hk)\\ b \,(\mathrm{mod}\, hl)}}^{*} \Big| \sum_{\substack{m\le M\\ n\le N}} c_{mn} e\Big(\frac{am}{hk} + \frac{bn}{hl}\Big)\Big|^2 \le (H^2K^2 + M)(HL^2 + N) \sum_{mn} |c_{mn}|^2.$$

Proof. First we apply the one-dimensional large sieve inequality

$$\sum_{l\le L} \sum_{b \,(\mathrm{mod}\, hl)}^{*} \Big| \sum_{n\le N} c_n e\Big(\frac{bn}{hl}\Big)\Big|^2 \le (HL^2 + N) \sum_n |c_n|^2$$

for the numbers

$$c_n = \sum_m c_{mn} e\Big(\frac{am}{hk}\Big).$$

Then we apply the one-dimensional large sieve inequality

$$\sum_{h\le H}\sum_{k\le K}\tau(hk)^{-1}\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,hk)}\Big|\sum_m c_{mn}e\Big(\frac{an}{hk}\Big)\Big|^2 \le (H^2K^2+M)\sum_m |c_{mn}|^2$$

and the result follows.

The factor $\tau(hk)^{-1}$ in the above result is annoying. Using the estimate $\tau(hk) \ll (hk)^\varepsilon$ we deduce from Lemma 13

Corollary 14. *For $X > Y \ge 1$, $M \ge 1$, $N \ge 1$ and any complex numbers c_{mn} we have*

$$\sum_{Y<hkl\le X}(hkl)^{-1}\sideset{}{^*}\sum_{\substack{a\,(\mathrm{mod}\,hk)\\ b\,(\mathrm{mod}\,hl)}}\Big|\sum_{\substack{m\le M\\ n\le N}} c_{mn}e\Big(\frac{am}{hk}+\frac{bn}{hl}\Big)\Big|^2 \ll \{(M+N+X)X^{1+\varepsilon}+MNY^{\varepsilon-1}\}\sum_{mn}|c_{mn}|^2$$

for any $\varepsilon > 0$, with the implied constant depending only on ε.

By Corollary 14 for $Y = G$, $X = D$, $M = N = A$, $c_{mn} = f(m+ni)$ we deduce that the moduli k, l with $g = [k,l]$ in the segment $G < g \le D$ contribute to $\mathcal{S}_f(D)$ at most

$$D(AD^{1+\varepsilon} + D^{2+\varepsilon} + A^2G^{\varepsilon-1})\|f\|^2. \tag{10.3}$$

Adding (10.3) to (10.2) we conclude:

Proposition 15. *Suppose $A \ge D \ge 1$. For any $G \ge 1$ we have*

$$\mathcal{S}_f(D) \le 2D\mathcal{S}_f(G) + O(AD(D^{1+\varepsilon} + AG^{\varepsilon-1})\|f\|^2) \tag{10.4}$$

with any $\varepsilon > 0$, with the implied constant depending only on ε.

Remarks. If $DG < A^{1-\varepsilon}$ the second term reduces to $O(A^2DG^{\varepsilon-1}\|f\|^2)$. In applications one needs only slightly better than the trivial bound $\mathcal{S}_f(D) \ll A^2D\|f\|^2$, and (10.4) for $G = (\log A)^j$ with j sufficiently large does the job. One still needs to estimate $\mathcal{S}_f(G)$, but only for G small, and this is a problem of Siegel–Walfisz type for which classical techniques of analytic number theory can be used.

References

[B1] E. Bombieri, *On twin almost primes*, Acta Arith. 28 (1975), 177–193; ibid. 28 (1976), 457–461.

[B2] —, *The asymptotic sieve*, Rend. Accad. Naz. XL (5), 1/2 (1975/76).

[C] M. Coleman, *The Rosser–Iwaniec sieve in number fields, with an application*, Acta Arith. 65 (1993), 53–83.

[DH] H. Davenport and H. Halberstam, *The values of a trigonometric polynomial at well spaced points*, Mathematika 13 (1966), 91–96.

[D] W. Duke, *Some problems in multidimensional analytic number theory*, Acta Arith. 52 (1989), 203–228.

[DFI] W. Duke, J. Friedlander and H. Iwaniec, *Equidistribution of roots of a quadratic congruence to prime moduli*, Ann. of Math. 141 (1995), 423–441.

[F] E. Fogels, *On the zeros of Hecke's L-functions I*, *II*, *III*, Acta Arith. 7 (1962), 87–106, 131–147, 225–240.

[FI] J. Friedlander and H. Iwaniec, *Bombieri's sieve*, in: Analytic Number Theory, Proceedings of a Conference in Honor of Heini Halberstam, Progr. Math. 138, Birkhäuser, 1996, 411–430.

[GR] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, London, 1965.

[H] E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen II*, Math. Z. 6 (1920), 11–51.

[IR] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, 2nd ed., Springer, 1982.

[I] H. Iwaniec, *Rosser's sieve*, Acta Arith. 36 (1980), 171–202.

[J] E. Jacobsthal, *Über die Darstellung der Primzahlen der Form* $4n+1$ *als Summe zweier Quadrate*, J. Reine Angew. Math. 132 (1907), 238–245.

[K] J. P. Kubilius, *On some problems in geometry of prime numbers*, Mat. Sb. 31 (73) (1952), 507–542 (in Russian).

[MV] H. L. Montgomery and R. C. Vaughan, *Hilbert's inequality*, J. London Math. Soc. (2) 8 (1974), 73–82.

[P] J. Pomyka/la, *Cubic norms represented by quadratic sequences*, Colloq. Math. 66 (1994), 283–297.

[R] G. J. Rieger, *Über die Summe aus einem Quadrat und einem Primzahlquadrat*, J. Reine Angew. Math. 231 (1968), 89–100.

[S] A. Selberg, *Lectures on Sieves*, Collected Papers, Vol. II, Springer, 1991.

[V] R. C. Vaughan, *Mean value theorems in prime number theory*, J. London Math. Soc. 10 (1975), 153–162.

Bât. 425 – Mathématique
Université de Paris-Sud
91405 Orsay Cedex, France
E-mail: etienne.fouvry@math.u-psud.fr

Department of Mathematics
Rutgers University
New Brunswick, New Jersey 08903
U.S.A.
E-mail: iwaniec@math.rutgers.edu

Received on 18.11.1996 (3076)