SIEVE METHODS AND EXPONENTIAL SUMS: AN INTERPLAY BETWEEN COMBINATORICS AND HARMONIC ANALYSIS

ANGEL V. KUMCHEV

ABSTRACT. Exponential sums over primes have many applications in analytic number theory. The first estimates for such sums were obtained in the late 1930's by I.M. Vinogradov, who used an elaborate combinatorial method to reduce the estimation of sums over primes to the estimation of "double sums." By the early 1980's, Vinogradov's combinatorial method was supplanted in most applications by combinatorial identities discovered by R.C. Vaughan and D.R. Heath-Brown. Subsequent work has revealed that when considered within the context of combinatorial sieve theory, those two approaches have much more in common than meets the eye. Moreover, the sieve viewpoint to such estimates leads to quantitative improvements in many applications, since it often allows for a more efficient deployment of the available harmonic analytic tools. In these notes, I explain the philosophy of the combinatorial sieve approach towards exponential sums over primes and present some applications to additive prime number theory.

INTRODUCTION

A great deal of our knowledge about the distribution of primes in various arithmetic sequences depends on estimates for exponential sums of the form

$$\sum_{p\sim X} e(f(p)), \tag{1}$$

where $e(x) = \exp(2\pi i x)$, f is a "well-behaved" function, and the summation is over primes p with $X/2 \le p < X$. The first estimates for such exponential sums were obtained by I.M. Vinogradov in 1937 and were the crucial ingredient in his proof of the Goldbach–Vinogradov three primes theorem [31, 32]. Vinogradov's idea was to reduce the estimation of the exponential sum (1) to the estimation of double exponential sums

$$\sum_{m\sim M}\sum_{mn\sim X} a_m b_n e(f(mn))$$

of the following two types:

- Type I: $|a_m| \ll 1$, $b_n = 1$, and M is not "too large;"
- Type II: $|a_m| \ll 1$, $|b_n| \ll 1$, and M is "neither small, nor large."

To achieve the transition from (1) to double sums, Vinogradov developed an intricate combinatorial technique. Nowadays, Vinogradov's combinatorial method can be classified as a combinatorial sieve and be understood within the context of sieve theory, but back in its day it was shrouded in mystery for all but a few "specialists." That remained the state of the subject until 1977, when R.C. Vaughan [28] published an identity for von Mangoldt's function $\Lambda(n)$ (see (7) below) that demystified the transition from sums over primes to double sums.

The discovery of Vaughan's identity (and of a related identity by D.R. Heath-Brown [10]) was part of a period of great excitement in analytic number theory. Another development that originated at about the same time (but really flourished during the 1990s) is the idea of prime-detecting sieves. These are sieve methods that utilize bounds for double sums to establish the existence of primes in various arithmetic sequences. While most prime-detecting sieves in the literature draw on the same basic ideas, their formulations differ from one circle of applications to another. The purpose of these notes is to describe in reasonable detail the applications of one such method—G. Harman's "alternative sieve"—to additive problems with prime variables. The reader will find an excellent introduction to applications of the same method to other problems in Harman's monograph [7].

The plan of the survey is as follows. We start with a motivating example from the theory of Diophantine approximations: we study the distribution modulo one of the sequence (αp), where α is an irrational real and p is a prime. This question provides a convenient setting to introduce an inexperienced reader to Vaughan's identity and to the main idea behind Harman's sieve. Indeed, Harman [4, 5] developed his method while studying this very problem. We also use the opportunity to explain some of the technical details involved in the sieve so that those details do not distract from more important things in later sections. The second part of the paper discusses the application of the sieve method to the problem of estimation of the size of the exceptional sets for sums of squares of primes. We sketch the application of the circle method to this problem and provide (without proofs) the necessary background on quadratic exponential sums over primes. We then present the details of the proofs of two applications of Harman's sieve to sums of four squares of primes. First, we present the proof of the earliest such theorem due to Harman and the author [8]. Then, we present the proof of a recent result of L. Zhao and the author [17], which is the sharpest result of this kind to date. We also describe a clever idea of Zhao that greatly simplifies the handling of the major arcs in the underlying application of the circle method.

1. A BASIC EXAMPLE: THE DISTRIBUTION OF αp MODULO ONE

Let $\|x\|$ denote the distance from a real number x to the nearest integer. In this section, we address the following question.

Question 1. *Suppose that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. For what $\theta > 0$ does the Diophantine inequality*

$$\|\alpha p\| < p^{-\theta} \tag{2}$$

have infinitely many solutions with p prime?

This problem has a long history [33, 29, 4, 13, 5, 14, 11, 25] that begins with the work of I.M. Vinogradov [33] and culminates in a recent result of K. Matomäki [25] that answers Question 1 for all $\theta < 1/3$. Here, we sketch the proofs of a result of R.C. Vaughan [29] that the range $0 < \theta < 1/4$ is acceptable and of Harman's first result [4] on this question, where he extends Vaughan's range to $0 < \theta < 3/10$.

1.1. First result. Vaughan's identity.

We first use harmonic analysis to reduce the above question to one about exponential sums. Let X be a large real number, and let $S_\alpha(X)$ denote the number of solutions of (2) with $p \sim X$. Set $\Delta = X^{-\theta}$, and choose a function $\phi \in C^\infty(\mathbb{R})$ that satisfies

$$0 \le \phi(x) \le \mathbb{1}(x; [-\Delta, \Delta]), \quad \int_{\mathbb{R}} \phi(x)\, dx = \Delta;$$

here, $\mathbb{1}(x;A)$ denotes the indicator function of a set A. Then $\Phi(x) = \phi(\|x\|)$ is 1-periodic and has a Fourier expansion of the form

$$\Phi(x) = \Delta + \sum_{h \neq 0} \hat{\Phi}(h) e(hx),$$

with Fourier coefficients satisfying

$$|\hat{\Phi}(h)| \ll_k \frac{\Delta}{(1+\Delta|h|)^k} \qquad (k \geq 0). \tag{3}$$

We will use Φ and its Fourier expansion to estimate $S_\alpha(X)$. By the construction of Φ,

$$S_\alpha(X) \geq \sum_{p \sim X} \Big(\frac{\log p}{\log X}\Big)\Phi(\alpha p) = \sum_{p \sim X} \Big(\frac{\log p}{\log X}\Big)\Big(\Delta + \sum_{h \neq 0} \hat{\Phi}(h) e(\alpha h p)\Big).$$

Let $H_0 = X^{\theta+\varepsilon}$. By (3) with k sufficiently large, the contribution to the last sum from h with $|h| > H_0$ is bounded. Hence, by the prime number theorem,

$$S_\alpha(X) \geq \frac{X^{1-\theta}}{2\log X}(1+o(1)) + \sum_{0<|h|\leq H_0} \frac{\hat{\Phi}(h)}{\log X} \sum_{p \sim X} (\log p) e(\alpha h p). \tag{4}$$

Thus, we have reduced the original problem to the estimation of an exponential sum. It now suffices to show that

$$\max_{1 \leq H \leq H_0} \sum_{h \sim H} \Big| \sum_{p \sim X} (\log p) e(\alpha h p) \Big| \ll X^{1-\varepsilon}. \tag{5}$$

Next, we turn attention to the above exponential sum bound. We will use Vaughan's identity to establish the closely related inequality

$$\max_{1 \leq H \leq H_0} \sum_{h \sim H} \Big| \sum_{x \sim X} \Lambda(x) e(\alpha h x) \Big| \ll X^{1-\varepsilon}, \tag{6}$$

where Λ is von Mangoldt's function. Suppose that U is a parameter to be chosen later subject to $U \leq X^{1/2-\varepsilon}$. The simplest form of Vaughan's identity states that

$$\Lambda(x) = \sum_{\substack{mn=x \\ m \leq U}} \mu(m)(\log n) - \sum_{\substack{rst=x \\ r,s \leq U}} \Lambda(r)\mu(s) + \sum_{\substack{rst=x \\ r,s>U}} \Lambda(r)\mu(s), \tag{7}$$

where μ is the Möbius function. If we put

$$a_m = \sum_{\substack{rs=m \\ r,s \leq U}} \Lambda(r)\mu(s), \quad b_n = \sum_{\substack{rt=n \\ r>U}} \Lambda(r),$$

we can use (7) to derive the identity

$$\begin{aligned} \sum_{x \sim X} \Lambda(x) e(\alpha h x) = & \sum_{m \leq U} \sum_{mn \sim X} \mu(m)(\log n) e(\alpha h m n) \\ & - \sum_{m \leq U^2} \sum_{mn \sim X} a_m e(\alpha h m n) \\ & + \sum_{m,n>U} \sum_{mn \sim X} \mu(m) b_n e(\alpha h m n). \end{aligned}$$

Note that $|a_m| \leq \log m$ and $|b_n| \leq \log n$. Hence, the first two sums on the right side of the last identity can be split into subsums of Type I with $M \leq U^2$, while the third sum can be split into subsums of Type II with $U \leq M \leq X/U$. We have thus reduced the problem to the following:

Can we choose $U \le X^{1/2-\varepsilon}$ so that we can estimate the contributions to the left side of (6) from all those Type I and Type II sums?

The techniques for estimation of double sums fall outside the scope of these notes, so we simply state the needed exponential sum bounds and treat them as "black boxes." In the present application, after a suitable choice of the large parameter X, we can report the estimate[1]

$$\sum_{h\sim H}\Big|\sum_{m\sim M}\sum_{mn\sim X} a_m b_n e(\alpha hmn)\Big| \ll X^{1-2\varepsilon},$$

provided that $1 \le H \le H_0$ and one of the following holds:

- the sum over m, n is of Type I with $M \le X^{1-\theta}$;
- the sum over m, n is of Type II with $X^{\theta} \le M \le X^{1-2\theta}$ or $X^{2\theta} \le M \le X^{1-\theta}$.

Suppose now that $\theta < 1/4$. Then the two ranges for M in the Type II sum bound overlap, and we can estimate contributions of Type I sums with $M \le X^{1-\theta}$ and of Type II sums with $X^{\theta} \le M \le X^{1-\theta}$. Let us compare the constraints

$$M \le X^{1-\theta}, \quad X^{\theta} \le M \le X^{1-\theta} \tag{8}$$

with the ranges

$$M \le U^2, \quad U \le M \le X/U \tag{9}$$

that emerged from the application of Vaughan's identity. We notice that if we choose $U = X^{\theta}$, inequalities (8) imply inequalities (9). Thus, when $\theta < 1/4$, we can apply Vaughan's identity with $U = X^{\theta}$ and the above triple-sum estimates to prove (6), and hence, also (5). Finally, combining (4) and (5), we easily get[2]

$$S_\alpha(X) \ge \frac{X^{1-\theta}}{2\log X}(1+o(1)).$$

This completes the proof of the following

Theorem 1.1 (Vaughan, 1977)**.** *When $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ and $0 < \theta < 1/4$, the Diophantine inequality* (2) *has infinitely many solutions with p prime.*

1.2. **Preparation for the sieve.** We would like to extend the range of θ in Theorem 1.1. To that end, let us examine its proof and the source of the restriction $\theta < 1/4$. We established an asymptotic formula,

$$\sum_{p\sim X}(\log p)\Phi(\alpha p) = X^{1-\theta}(0.5+o(1)).$$

Our proof had two parts: harmonic-analytic (Fourier series) and combinatorial (Vaughan's identity). The harmonic analysis was relatively simple and came first. It produced the main term in our asymptotic formula and left us to deal with an error term in the form of the exponential sum (5). The combinatorial analysis then transformed (5) into a number of sums of the form

$$\sum_{h\sim H}\Big|\sum_{m\sim M}\sum_{mn\sim X} a_m b_n e(\alpha hmn)\Big|,$$

[1] For proofs, see Vaughan [29] or Harman [7, §2.3].

[2] In fact, with a little more effort, we can turn this lower bound into an asymptotic formula with the same main term.

where the double sums are either Type I or Type II. A successful proof required that we be able to estimate all the resulting triple sums, and their estimation for $\theta < 1/4$ was pretty straightforward. On the other hand, when $\theta > 1/4$, we do not have an estimate for Type II sums with $X^{1-2\theta} \le M \le X^{2\theta}$, and so there is no choice of U such that we can estimate all Type II sums with $U \le M \le X/U$. Thus, the approach from §1.1 fails when $\theta > 1/4$. This is especially frustrating when $\theta = 1/4 + \varepsilon$, as in that case we can estimate all the triple sums except for a small number of Type II sums with $X^{1/2-\varepsilon} \le M \le X^{1/2+\varepsilon}$. Yet, even though we have lost control over just a few sums, we have lost the result completely.

Harman's alternative sieve is designed to achieve further progress in situations such as the just described. The main idea is to interchange the order of the harmonic and combinatorial analyses. That is, we start with the combinatorial analysis and apply it directly to $\sum_p \Phi(\alpha p)$ (as opposed to the exponential sums that appear in the harmonic analysis of this sum). Our goal will be to organize the combinatorics so that the "bad" Type II sums can be avoided.

Let us define, for $z \ge 2$,

$$\psi(n,z) = \begin{cases} 1 & \text{if } n \text{ has no prime divisor } p \text{ with } p \le z, \\ 0 & \text{otherwise.} \end{cases}$$

It is also convenient to extend the definition of $\psi(n,z)$ to all positive *real* n by setting $\psi(n,z) = 0$ when $n \notin \mathbb{Z}$. For integers n with $n \sim X$, $\psi(n, X^{1/2})$ is simply the indicator function of the primes. Hence,

$$\sum_{p\sim X} \Phi(\alpha p) = \sum_{n\sim X} \psi(n, X^{1/2})\Phi(\alpha n). \tag{10}$$

Harman's combinatorial argument is based on *Buchstab's identity*,

$$\psi(n,z_2) = \psi(n,z_1) - \sum_{z_1<p\le z_2} \psi(n/p,p) \qquad (2 \le z_1 < z_2), \tag{11}$$

which is merely a form of the inclusion-exclusion principle. Applying Buchstab's identity to the right side of (10), we get

$$\begin{aligned} \sum_{n\sim X} \psi(n,X^{1/2})\Phi(\alpha n) &= \sum_{n\sim X} \psi(n,z)\Phi(\alpha n) - \sum_{n\sim X} \sum_{z<p\le X^{1/2}} \psi(n/p,p)\Phi(\alpha n) \\ &= \sum_{n\sim X} \psi(n,z)\Phi(\alpha n) - \sum_{z<p\le X^{1/2}} \sum_{mp\sim X} \psi(m,p)\Phi(\alpha mp) \\ &= \Sigma_1 - \Sigma_2, \quad \text{say.} \end{aligned}$$

Let us now take a look at the double sum Σ_2 in the above decomposition. Using the Fourier expansion of Φ and the bound (3) for its Fourier coefficients, we obtain

$$\begin{aligned} \Sigma_2 &= \sum_{z<p\le X^{1/2}} \sum_{mp\sim X} \psi(m,p)\Big(\Delta + \sum_{0<|h|\le H_0} \hat{\Phi}(h) e(\alpha hmp)\Big) + O(1) \\ &= \Delta T_2(0) + O\Big(\Delta \sum_{h\le H_0} |T_2(h)| + 1\Big), \end{aligned}$$

where

$$T_2(h) = \sum_{z<p\le X^{1/2}} \sum_{mp\sim X} \psi(m,p) e(\alpha hmp).$$

The sum $T_2(0)$ that appears in the main term can be evaluated using standard techniques from prime number theory. The prime number theorem and the following lemma (itself a consequence of the prime number theorem) are the main tools for evaluating $T_2(0)$ and other similar sums that appear in the sequel.

Lemma 1.2. *Suppose that x is large and $z = x^{\zeta}$, with $0 < \varepsilon \le \zeta \le 1$. Then, for any fixed $A > 0$,*

$$\begin{aligned}\sum_{n\le x} \psi\big(n, x^{\zeta}\big) &= \frac{1}{\log z}\sum_{n\le x}\omega\Big(\frac{\log n}{\log z}\Big) + O\big(x(\log x)^{-A}\big)\\ &= \frac{x}{\log x}\big(\zeta^{-1}\omega(1/\zeta) + O\big((\log x)^{-1}\big)\big),\end{aligned}$$

where ω is Buchstab's function*: the continuous solution of the differential delay equation*

$$\begin{cases}\omega(u) = 1/u & \text{when } 1 \le u \le 2,\\ (u\omega(u))' = \omega(u-1) & \text{when } u > 2.\end{cases}$$

Using the above lemma, we get

$$T_2(0) = \frac{X}{2\log X}(c_2 + o(1)), \qquad c_2 = \int_{\zeta}^{1/2}\omega\Big(\frac{1-t}{t}\Big)\frac{dt}{t^2} = \int_2^{1/\zeta}\omega(u-1)\,du, \tag{12}$$

where $z = X^{\zeta}$. Superficial technical details aside, this evaluation is standard, and we are really interested in the sums $T_2(h)$, with $1 \le h \le H_0$, appearing in the remainder term above. Those sums resemble Type II sums but for one important detail: the coefficients $\psi(m, p)$ are not products of the form $a_m b_p$. Since that structure of the coefficients of the Type II sum plays a central role in its estimation, this difference is an obstacle to the direct estimation of $T_2(h)$. However, by a simple trick using Perron's formula (see Harman [7, §3.2]), we can show that

$$T_2(h) \ll (\log X)\Big|\sum_{z<p\le X^{1/2}}\sum_{mp\sim X} a_m b_p e(\alpha h m p)\Big|, \tag{13}$$

where the coefficients a_m and b_p are complex numbers with $|a_m| \le 1$ and $|b_p| \le 1$. The latter sum can be split into $O(\log X)$ subsums, each of which is a genuine Type II sum with $z \le M \le X^{1/2}$. Therefore, when $0 < \theta < 1/4$ and $z \ge X^{\theta}$, we can use our Type II sum bound to deduce that

$$\Sigma_2 = \frac{X^{1-\theta}}{2\log X}(c_2 + o(1)).$$

Is it possible to use the same approach to obtain an approximation for Σ_1? If we were to try, we would discover that we need an asymptotic formula similar to (12) for $T_1(0)$ and an upper bound for $T_1(h)$, with $1 \le h \le H_0$, where

$$T_1(h) = \sum_{n\sim X}\psi(n, z)e(\alpha h n).$$

The sum $T_1(h)$ is neither of Type I nor of Type II, and thus, we currently know no bound for it. It turns out, however, that we can derive a bound for this sum (and for more general sums) from bounds for Type I and Type II sums.

Lemma 1.3. *Let α, X and H_0 be as above. Suppose that $M \le X^{1-\theta}$, $1 \le H \le H_0$, $z \le X^{1-3\theta}$, and (a_m) is a complex sequence, with $|a_m| \le 1$. Then*

$$\sum_{h\sim H}\Big|\sum_{m\sim M}\sum_{mn\sim X} a_m\psi(n, z)e(\alpha h m n)\Big| \ll X^{1-1.5\varepsilon}.$$

We will not give a complete proof of the lemma and will be content with a sketch of the main idea instead.

Sketch of the proof. Let $\Pi(z) = \prod_{p\le z} p$. By the properties of the Möbius function, the given exponential sum is

$$\sum_{m\sim M}\sum_{mn\sim X}\sum_{d|(n,\Pi(z))} a_m\mu(d)e(\alpha hmn) \ll (\log X)\Big|\sum_{m\sim M}\sum_{\substack{d|\Pi(z)\\ d\sim D}}\sum_{mkd\sim X} a_m\mu(d)e(\alpha hmkd)\Big|,$$

for some $D \ll X/M$. We consider two cases:

Case 1: $MD \ll X^{1-\theta}$. Then the above sum can be rewritten as a Type I sum.

Case 2: $MD \gg X^{1-\theta}$. Then $d = p_1p_2\cdots p_s$, $s\ge 1$, where the p_j's are primes with

$$p_1 < p_2 < \cdots < p_s \le z, \quad mp_1p_2\cdots p_s \gg X^{1-\theta}.$$

Let $1\le t\le s$ be such that $mp_1p_2\cdots p_t \gg X^{1-\theta} \gg mp_1p_2\cdots p_{t-1}$. Recalling that $p_t \le z \le X^{1-3\theta}$, we deduce that

$$X^{2\theta} \ll mp_1p_2\cdots p_{t-1} \ll X^{1-\theta}.$$

We can use this observation and the argument behind (13) to bound the given sum by a linear combination of $O\big((\log X)^3\big)$ Type II sums. □

The kind of sum that appears in Lemma 1.3 is not quite as general as a Type II sum, but it is more general than a Type I sum. Indeed, if $z = 1$, $\psi(n,z) = 1$ and the above sum turns into a Type I sum. We shall refer to this type of sum as a Type I/II sum. Note that the sum $T_1(h)$ above is a Type I/II sum with $M = 1$ (which is acceptable in the lemma). Thus, if $z \le X^{1-3\theta}$, we can use the lemma to estimate $T_1(h)$ and to show that

$$\Sigma_1 = \frac{X^{1-\theta}}{2\log X}(c_1 + o(1)), \qquad c_1 = \zeta^{-1}\omega(1/\zeta). \tag{14}$$

When $0<\theta<1/4$, we can choose $z = X^{1/4}$ in the above analysis of Σ_1 and Σ_2 to obtain an alternative proof of the asymptotic formula

$$\sum_{p\sim X}\Phi(\alpha p) = \frac{X^{1-\theta}}{2\log X}(c_1 - c_2 + o(1)) = \frac{X^{1-\theta}}{2\log X}(1+o(1)),$$

which of course we already knew. However, the real gain from the above discussion will come in the next section, where we will replace the asymptotic formula for Σ_2 with an upper bound valid in a wider range for θ.

1.3. **A simple lower-bound sieve.** Let us consider again how things change as θ crosses over the threshold $\theta = 1/4$. The argument in §1.2 required that we choose z with $X^\theta \le z \le X^{1-3\theta}$. When $\theta > 1/4$, this is not possible, so we set $z = X^{1-3\theta}$. The analysis of Σ_1 then remains the same for $\theta < 1/3$, but we need to revisit Σ_2. We write

$$\begin{aligned}\Sigma_2 &= \Big\{\sum_{z<p<X^\theta} + \sum_{X^\theta\le p\le X^{1-2\theta}} + \sum_{X^{1-2\theta}<p\le X^{1/2}}\Big\}\sum_{mp\sim X}\psi(m,p)\Phi(\alpha mp)\\ &= \Sigma_3+\Sigma_4+\Sigma_5, \quad \text{say.}\end{aligned}$$

We can use our Type II sum bound to evaluate Σ_4 similarly to how we evaluated the entire Σ_2 in the case $\theta < 1/4$. This yields

$$\Sigma_4 = \frac{X^{1-\theta}}{2\log X}(c_4 + o(1)), \qquad c_4 = \int_\theta^{1-2\theta} \omega\Big(\frac{1-t}{t}\Big)\frac{dt}{t^2}.$$

We now turn to Σ_3. The harmonic analysis of this sum produces exponential sums of Type II that we cannot estimate. However, we can apply Buchstab's identity to Σ_3 to get

$$\begin{aligned}\Sigma_3 &= \sum_{z<p<X^\theta}\sum_{mp\sim X}\psi(m,z)\Phi(\alpha mp) - \sum_{z<p<X^\theta}\sum_{mp\sim X}\sum_{z<q\le p}\psi(m/q,q)\Phi(\alpha mp)\\ &= \sum_{z<p<X^\theta}\sum_{mp\sim X}\psi(m,z)\Phi(\alpha mp) - \sum_{z<q\le p<X^\theta}\sum_{kpq\sim X}\psi(k,q)\Phi(\alpha kpq) = \Sigma_6 - \Sigma_7, \quad \text{say;}\end{aligned}$$

here q also denotes a prime number. Note that Σ_6 is a Type I/II sum, which can be evaluated similarly to Σ_1; we have

$$\Sigma_6 = \frac{X^{1-\theta}}{2\log X}(c_6 + o(1)), \qquad c_6 = \frac{1}{1-3\theta}\int_{1-3\theta}^{\theta} \omega\Big(\frac{1-t}{1-3\theta}\Big)\frac{dt}{t}.$$

We can give a similar further decomposition for Σ_5. However, before we do that, we note that in the case of Σ_5, the only integers m with $mp\sim X$ and $\psi(m,p)\ne 0$ are the primes $q\sim X/p$. Hence, on writing $Y_p = (X/p)^{1/2}$, we have

$$\begin{aligned}\Sigma_5 &= \sum_{X^{1-2\theta}<p\le X^{1/2}}\sum_{mp\sim X}\psi(m,Y_p)\Phi(\alpha mp)\\ &= \sum_{X^{1-2\theta}<p\le X^{1/2}}\sum_{mp\sim X}\psi(m,z)\Phi(\alpha mp) - \sum_{\substack{X^{1-2\theta}<p\le X^{1/2}\\ z<q\le Y_p}}\sum_{kpq\sim X}\psi(k,q)\Phi(\alpha kpq) = \Sigma_8 - \Sigma_9.\end{aligned}$$

Again, Σ_8 is a Type I/II sum and we have

$$\Sigma_8 = \frac{X^{1-\theta}}{2\log X}(c_8 + o(1)), \qquad c_8 = \frac{1}{1-3\theta}\int_{1-2\theta}^{1/2} \omega\Big(\frac{1-t}{1-3\theta}\Big)\frac{dt}{t}.$$

Combining all the decompositions and evaluations, we now obtain

$$\sum_{p\sim X}\Phi(\alpha p) = \frac{X^{1-\theta}}{2\log X}(c_1 - c_4 - c_6 - c_8 + o(1)) + \Sigma_7 + \Sigma_9. \tag{15}$$

We still do not know how to evaluate Σ_7 and Σ_9, but they have one important advantage over their exponential sum counterparts—they are non-negative! Hence, (15) gives

$$\sum_{p\sim X}\Phi(\alpha p) \ge \frac{X^{1-\theta}}{2\log X}(c_1 - c_4 - c_6 - c_8 + o(1)), \tag{16}$$

which is non-trivial whenever $c_1 - c_4 - c_6 - c_8 > 0$. It is easy to see that this already supersedes the earlier result. Indeed, when $\theta = 1/4+\varepsilon$, with $\varepsilon > 0$ small, one can show that the sum $c_4 + c_6 + c_8$ is close to c_2, whence $c_1 - c_4 - c_6 - c_8$ is close to 1. Thus, we can get a result with $\theta = 1/4+\varepsilon$ for *some* $\varepsilon > 0$.

Before we try to go any further with this sieve idea, let us analyze the constant $c_1 - c_4 - c_6 - c_8$ appearing in the lower bound (16). We decomposed the left side of (16) as

$$\sum_{p\sim X}\Phi(\alpha p) = \Sigma_1 - \Sigma_4 - \Sigma_6 + \Sigma_7 - \Sigma_8 + \Sigma_9.$$

We expect that

$$\Sigma_i = \frac{X^{1-\theta}}{2\log X}(c_i + o(1))$$

for all six values of i that appear in the above decomposition, but we can prove those asymptotic formulas only for $i = 1,4,6$, and 8. For Σ_7 and Σ_9, we *expect* such asymptotic formulas with

$$c_7 = \int_{1-3\theta}^{\theta}\int_{1-3\theta}^{u} \omega\Big(\frac{1-u-v}{v}\Big)\frac{dv\,du}{uv^2} \quad\text{and}\quad c_9 = \int_{1-2\theta}^{1/2}\int_{1-3\theta}^{(1-u)/2} \omega\Big(\frac{1-u-v}{v}\Big)\frac{dv\,du}{uv^2},$$

respectively, but we cannot prove them. However, it follows easily from our combinatorial decomposition and the prime number theorem that

$$c_1 - c_4 - c_6 + c_7 - c_8 + c_9 = 1,$$

whence

$$c_1 - c_4 - c_6 - c_8 = 1 - c_7 - c_9.$$

We note that c_7 and c_9 are the constants in the main terms of the two expected asymptotic formulas that we missed. One may think of this as follows: Each sum (like Σ_7 and Σ_9 above) which we cannot evaluate and estimate trivially results in a "loss" being subtracted from the expected asymptotic formula for the original sum; the resulting lower bound is non-trivial if the total of such losses does not exceed the expected main term (which has a constant coefficient equal to 1). This observation is useful in the final stage of the application of the method. Indeed, the sum $c_7 + c_9$ is an increasing function of θ, and so the real limit of our simple sieve is the solution θ_0 of the transcendental equation $c_7(\theta) + c_9(\theta) = 1$. It is very difficult (perhaps, even impossible) to solve this equation in closed form, but a numerical approximation to θ_0 is definitely within reach. Numerical integration yields

$$c_7(0.284) + c_9(0.284) < 0.951, \qquad c_7(0.285) + c_9(0.285) > 1.003,$$

so $\theta = 0.284$ presents a reasonable lower estimate for θ_0. Thus, we obtain an affirmative answer to Question 1 when $\theta \le 0.284$.

We can easily improve on the above result. For example, there is no need to discard all of Σ_9. Recall that

$$\Sigma_9 = \sum_{X^{1-2\theta} < p_1 \le X^{1/2}} \sum_{z < p_2 < Y_{p_1}} \sum_{kp_1p_2 \sim X} \psi(k, p_2)\Phi(\alpha k p_1 p_2).$$

We can use our Type II sum bound to evaluate the part of this sum where $X^{2\theta} \le p_1p_2 \le X^{1-\theta}$. The evaluation of this sum will reduce the loss from Σ_9 from c_9 to

$$b_9 = \iint_{\mathcal{D}_9} \omega\Big(\frac{1-u-v}{v}\Big)\frac{dv\,du}{uv^2},$$

where $\mathcal{D}_9$ is the two-dimensional region defined by the inequalities

$$1-2\theta \le u \le 1/2, \quad 1-3\theta \le v \le (1-u)/2, \quad u+v \notin [2\theta, 1-\theta].$$

Similarly, we can use Type II sum bounds to evaluate the part of Σ_7 sum where $X^{\theta} \le p_1p_2 \le X^{1-2\theta}$. Let us examine further the two parts of Σ_7 which are subject to the conditions

$$p_1p_2 < X^{\theta} \quad\text{or}\quad p_1p_2 > X^{1-2\theta}.$$

Let Σ_{10} denote the part of Σ_7 subject to $p_1p_2 < X^\theta$. This sum may be empty (it *is* empty when $\theta \le 2/7$), but even when Σ_{10} does not vanish, we can evaluate most of it by further use of Buchstab's identity. Two more appeals to Buchstab's identity yield

$$\begin{aligned}\sum_{\substack{z<p_2\le p_1\\ p_1p_2<X^\theta}}\sum_{mp_1p_2\sim X}\psi(m,p_2)\Phi(\alpha mp_1p_2) &= \sum_{\substack{z<p_2\le p_1\\ p_1p_2<X^\theta}}\sum_{mp_1p_2\sim X}\psi(m,z)\Phi(\alpha mp_1p_2)\\ &\quad - \sum_{\substack{z<p_3\le p_2\le p_1\\ p_1p_2<X^\theta}}\sum_{mp_1p_2p_3\sim X}\psi(m,z)\Phi(\alpha mp_1p_2p_3)\\ &\quad + \sum_{\substack{z<p_4\le p_3\le p_2\le p_1\\ p_1p_2<X^\theta}}\sum_{mp_1\cdots p_4\sim X}\psi(m,p_4)\Phi(\alpha mp_1\cdots p_4).\end{aligned}$$

Note that the first two sums on the right can be evaluated using our Type I/II sum bound. Furthermore, we can use a Type II sum estimate to evaluate any part of the quintuple sum over $m,p_1,\dots,p_4$ in which a subproduct of $p_1p_2p_3p_4$ lies in one of the ranges $[X^\theta, X^{1-2\theta}]$ or $[X^{2\theta}, X^{1-\theta}]$. The remaining, "bad" parts of the quintuple sum we discard. At the end, the total loss from discarding those "bad" sums will be measured by the quadruple integral

$$b_{10} = \iiiint_{\mathcal{D}_{10}} \omega\Big(\frac{1-u_1-u_2-u_3-u_4}{u_4}\Big)\frac{du_4\cdots du_1}{u_1u_2u_3u_4^2},$$

where the domain of integration is defined by the conditions

$$1-3\theta \le u_4 \le u_3 \le u_2 \le u_1 < u_1+u_2 \le \theta,$$
$$\text{no subsum of } u_1+\cdots+u_4 \text{ lies in } [\theta, 1-2\theta]\cup[2\theta,1-\theta].$$

A similar argument can be applied to the part of Σ_7 where[3] $p_1p_2 > X^{1-2\theta}$ and $p_1p_2^2 \le X^{1-\theta}$. The ensuing loss from discarded quintuple sums is given by the integral

$$b_{11} = \iiiint_{\mathcal{D}_{11}} \omega\Big(\frac{1-u_1-u_2-u_3-u_4}{u_4}\Big)\frac{du_4\cdots du_1}{u_1u_2u_3u_4^2},$$

where the domain of integration is defined by the conditions

$$1-3\theta \le u_4 \le u_3 \le u_2 \le u_1 \le \theta, \quad u_1+u_2 \ge 1-2\theta, \quad u_1+2u_2 \le 1-\theta,$$
$$\text{no subsum of } u_1+\cdots+u_4 \text{ lies in } [\theta, 1-2\theta]\cup[2\theta,1-\theta].$$

Finally, the total loss from discarding the part of Σ_7 where $p_1p_2 > X^{1-2\theta}$ and $p_1p_2^2 > X^{1-\theta}$ is

$$b_{12} = \iint_{\mathcal{D}_{12}} \omega\Big(\frac{1-u-v}{v}\Big)\frac{dv\,du}{uv^2},$$

where the domain of integration is defined by the conditions

$$1-3\theta \le v \le u \le \theta, \quad u+v > 1-2\theta, \quad u+2v > 1-\theta.$$

Altogether, we reduced the loss from Σ_7 and Σ_9 from c_7+c_9 to $b_9+\cdots+b_{12}$. This is a substantial saving: a rather crude numerical estimation shows that

$$c_7(0.3)+c_9(0.3) > 3.28, \qquad b_9(0.3)+\cdots+b_{12}(0.3) < 0.71.$$

In particular, this establishes the following

[3]The latter inequality ensures the respective quadruple sum over m, p_1, p_2, p_3 is an acceptable Type I/II sum.

Theorem 1.4 (Harman, 1983)**.** *When $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ and $0 < \theta \le 3/10$, the Diophantine inequality* (2) *has infinitely many solutions with p prime.*

We should pause here to make an important observation about the method. Whereas in Theorem 1.1 the exponent $\theta = 1/4$ is a "hard threshold" at which the proof breaks (hence, the strict inequality $\theta < 1/4$), the exponent $\theta = 3/10$ in this result is just a nice-looking approximation for the real breaking point in the proof. It is clear from the numerical work, however, that a Jia's result [13] with $\theta \le 4/13$ is out of reach with the above sieve construction.

2. An application to additive number theory: Exceptional sets for sums of squares of primes

The applications of the alternative sieve to problems in Diophantine approximation (such as the one above) and to the distribution of primes in short intervals and in arithmetic progressions are discussed in detail and at great length in Harman's monograph [7]. In the remainder of these notes, we shall focus on the applications of the method to additive problems with prime variables. Our exposition will focus on the following two questions.

Question 2. *Let X be a large real number, and write*

$$E_3(X) = \#\{n \sim X : n \equiv 3 \pmod{24},\ 5 \nmid n,\ n \ne p_1^2 + p_2^2 + p_3^2\}.$$

For what $\theta > 0$ does the inequality $E_3(X) \ll X^\theta$ hold?

Question 3. *Let X be a large real number, and write*

$$E_4(X) = \#\{n \sim X : n \equiv 4 \pmod{24},\ n \ne p_1^2 + \cdots + p_4^2\}.$$

For what $\theta > 0$ does the inequality $E_4(X) \ll X^\theta$ hold?

The bounds

$$E_k(X) \ll X(\log X)^{-A} \qquad (k = 3, 4)$$

are classical, having been proved by L.K. Hua [12] for some fixed $A > 0$, and by W. Schwarz [27] for all fixed $A > 0$. The first to give an affirmative answer to Questions 2 or 3 for a fixed $\theta < 1$ were M.C. Leung and M.C. Liu [18], who established the bound $E_3(X) \ll X^{1-\delta}$ for some (very small) fixed $\delta > 0$. A big breakthrough occurred in the late 1990's, when J.Y. Liu and T. Zhan [22] discovered a new technique for dealing with the major arcs in applications of the circle method to additive problems with prime variables. A flurry of activity ensued that produced series of results on sums of three and four squares (see [2, 23, 24, 16] and [20, 19, 34, 21], respectively) that culminated in the results of the author [16] and of J.Y. Liu, T.D. Wooley, and G. Yu [21] that, for any fixed $\varepsilon > 0$,

$$E_3(X) \ll X^{7/8+\varepsilon} \quad \text{and} \quad E_4(X) \ll X^{3/8+\varepsilon}. \tag{17}$$

It transpired from that body of work that solutions to Questions 2 and 3 obtained by the circle method follow similar paths and one should, in general, expect simultaneous bounds of the form

$$E_3(X) \ll X^{1-\sigma} \quad \text{and} \quad E_4(X) \ll X^{1/2-\sigma} \tag{18}$$

for a given $\sigma > 0$.

Subsequently, G. Harman and the author [8, 9] adapted the sieve idea described in §1 to obtain further improvements on (17). In [8] and [9], we established the bounds (18) for all

$\sigma < 1/7$ and $\sigma < 3/20$, respectively, thus reaffirming the conventional wisdom that results in the two problems should move in lockstep. Our work revealed that applications of sieve methods to additive problems with prime variables may run into technical obstacles not present in more traditional settings, like our introductory example. In particular, in both of our papers, the breaking points $\sigma = 1/7$ and $\sigma = 3/20$ resulted from reaching technical barriers (hence, the strict inequalities). On the other hand, the presence of multiple prime variables in Questions 2 and 3 allowed us to use J.R. Chen's "switching trick" in ways not readily available in conventional applications of the alternative sieve.

In a recent private communication, L. Zhao made a very clever observation about the way the circle method is applied to Question 3. His observation sidesteps the technical difficulties that limited what Harman and the author could achieve in [8, 9]. It allows one to push the result on four squares in [9] to $\sigma \le 0.153$ and to simplify the proof in process. Using that idea and new exponential sum bounds for quadratic exponential sums (see Lemma 2.4 below), L. Zhao and the author [17] recently proved that

$$E_4(X) \ll X^{11/32}. \tag{19}$$

Note that since Zhao's idea is not applicable to sums of three squares of primes, we obtain no improvement on the bound for $E_3(X)$ in [9].

2.1. **A canonical application of the circle method.** We begin the discussion of sums of squares of primes with a sketch of the proof of the result of J.Y. Liu, T.D. Wooley, an G. Yu [21] on sums of four squares.

Theorem 2.1 (J.Y. Liu *et al.*, 2004)**.** *Let $\varepsilon > 0$ be fixed. Then $E_4(X) \ll X^{3/8+\varepsilon}$.*

Let $\mathcal{E} = \mathcal{E}(X)$ denote the set of integers counted by $E_4(X)$. For the remainder of §2, we suppress the index 4 and write simply $E(X) = \#\mathcal{E}$. To estimate $E(X)$, we set

$$R(n) = \sum_{\substack{p_1^2+\cdots+p_4^2=n \\ p_i \sim N}} 1, \quad N = \tfrac{2}{3}X^{1/2}.$$

For each $n \sim X$, we have

$$R(n) = \int_0^1 S(\alpha)^4 e(-\alpha n)\, d\alpha, \quad S(\alpha) = \sum_{p\sim N} e(\alpha p^2).$$

Suppose that $1 \le P \le Q \le X$, and define the sets of major and minor arcs by

$$\mathfrak{M} = \bigcup_{1\le q\le P} \bigcup_{\substack{1\le a\le q \\ (a,q)=1}} \left(\frac{a}{q} - \frac{1}{qQ}, \frac{a}{q} + \frac{1}{qQ} \right), \quad \mathfrak{m} = \left[Q^{-1}, 1+Q^{-1}\right] \setminus \mathfrak{M}.$$

First, we estimate the contribution from the major arcs, which yields the main term in the expected asymptotic formula for $R(n)$. For any fixed $A > 4$, we have

$$\int_{\mathfrak{M}} S(\alpha)^4 e(-\alpha n)\, d\alpha = \kappa_n N^2 (\log N)^{-4} + O\big(N^2 (\log N)^{-A}\big) \tag{20}$$

provided that PQ is "close" to N^2. Here, κ_n is a function of n which satisfies

$$1 \ll \kappa_n \ll \log\log X$$

for $n \equiv 4 \pmod{24}$ with $n \sim X$. When $P = (\log N)^{B_1}$ and $Q = N^2(\log N)^{-B_2}$, with $B_i = B_i(A) > 0$ sufficiently large, the proof of this result is an exercise using the Siegel–Walfisz theorem and partial summation.[4] Using the technique that he and T. Zhan introduced in [22], J.Y. Liu [19] showed that, in fact, one can choose any P and Q such that

$$P \le N^{2/5-\varepsilon}, \quad Q \ge N^{8/5+\varepsilon}, \quad PQ \le N^2.$$

Setting the difficulty of its proof aside, we view (20) as the "easy" part of $R(n)$, similar to the contribution from the zeroth Fourier coefficient in the study of αp modulo one.

The estimation of the contribution from the minor arcs is harder, and we can obtain a bound only on average over n. To this end, we use a device introduced by T.D. Wooley [34]. We first note that, for any exceptional $n \in \mathcal{E}$, we have

$$-\int_{\mathfrak{m}} S(\alpha)^4 e(-\alpha n)\, d\alpha = \int_{\mathfrak{M}} S(\alpha)^4 e(-\alpha n)\, d\alpha \gg N^2 (\log N)^{-4}.$$

Hence,

$$-\int_{\mathfrak{m}} S(\alpha)^4 Z(\alpha)\, d\alpha = -\sum_{n\in\mathcal{E}} \int_{\mathfrak{m}} S(\alpha)^4 e(-\alpha n)\, d\alpha \gg E(X) N^2 (\log N)^{-4}, \tag{21}$$

where

$$Z(\alpha) = \sum_{n\in\mathcal{E}} e(-\alpha n).$$

We now estimate the left side of (21) from above. We have

$$\begin{aligned}\int_{\mathfrak{m}} S(\alpha)^4 Z(\alpha)\, d\alpha &\ll \big(\max_{\alpha\in\mathfrak{m}} |S(\alpha)|\big)\Big(\int_0^1 |S(\alpha)|^4\, d\alpha\Big)^{1/2}\Big(\int_0^1 |S(\alpha)Z(\alpha)|^2\, d\alpha\Big)^{1/2} \\ &\ll \big(\max_{\alpha\in\mathfrak{m}} |S(\alpha)|\big) N^{1+\varepsilon}\big(E(X)^{1/2}N^{1/2} + E(X)\big),\end{aligned} \tag{22}$$

where the last inequality uses some standard mean-value estimates (e.g., Hua's lemma: see R.C. Vaughan [30, Theorem 2.5]). Therefore, if we assume a bound of the form

$$\max_{\alpha\in\mathfrak{m}} |S(\alpha)| \ll N^{1-\sigma}$$

for some fixed $\sigma > 0$, we can combine (21) and (22) to show that

$$E(X) \ll N^{-2}(\log N)^4 N^{5/2-\sigma+\varepsilon} E(X)^{1/2} \ll N^{1/2-\sigma+3\varepsilon/2} E(X)^{1/2},$$

whence, upon readjusting the choice of ε,

$$E(X) \ll X^{1/2-\sigma+\varepsilon}. \tag{23}$$

This reduces the problem of estimating $E(X)$ to that of estimating the exponential sum $f(\alpha)$ on the minor arcs. The best known estimate for $f(\alpha)$ goes back to work of A. Ghosh [3] in the early 1980s and states that

$$S(a/q+\beta) \ll N^{1+\varepsilon}\big(q^{-1/4} + N^{-1/8}\big), \tag{24}$$

provided that $(a,q)=1$ and β is "small" (essentially, $|\beta| < q^{-2}$). On the bulk of the minor arcs, the term $N^{-1/8}$ in the above bound dominates the term $q^{-1/4}$. On the small set where the situation is reversed, Ghosh's bound has been improved by X. Ren [26] and the author [16], so we actually have

$$S(a/q+\beta) \ll N^{1+\varepsilon}\big(q^{-1/2} + N^{-1/8}\big), \tag{25}$$

[4] The motivated reader may find it worthwhile to work his/her way through this exercise. The proof of the respective result for sums of three primes in R.C. Vaughan [30, Theorem 3.3] provides a good blueprint how to proceed.

for $a/q+\beta$ on a minor arc. Therefore, upon choosing P, Q in the definition of the minor arcs so that

$$N^{1/4} \ll P \ll N^{2/5-\varepsilon}, \quad Q = N^2 P^{-1},$$

we obtain the bound (23) with $\sigma = 1/8$.

2.2. **Some background on quadratic exponential sums.** The exponent 3/8 in Theorem 2.1 was set by the limitations of an exponential sum bound similar to that in Vaughan's result on αp. In later sections, we sketch how the alternative sieve and the circle method can be combined to yield improvements on Theorem 2.1. However, before we launch that discussion, we take a short detour to equip the reader with the required background about exponential sums.

The standard estimation of $S(\alpha)$ on the minor arc uses ideas similar to the estimation of the exponential sum in §1. Let $\mathfrak{M}_\sigma$ and $\mathfrak{m}_\sigma$ be the major and minor arcs corresponding to $P = N^{4\sigma}$ and $Q = N^2 P^{-1}$, where $0 < \sigma < 1/6$. We can show that

$$\sum_{m \sim M} \sum_{mk \sim N} a_m b_k e(\alpha m^2 k^2) \ll N^{1-\sigma+\varepsilon}, \tag{26}$$

provided that $\alpha \in \mathfrak{m}_\sigma$ and one of the following holds:

- the sum is of Type I with $M \le N^{1/2-\sigma}$;
- the sum is of Type II with $N^{2\sigma} \le M \le N^{1-4\sigma}$ or $N^{4\sigma} \le M \le N^{1-2\sigma}$.

When $0 < \sigma < 1/8$, the two ranges for M in our Type II sum estimate overlap, and we have a Type II bound whenever $N^{2\sigma} \le M \le N^{1-2\sigma}$. Thus, we can use Vaughan's identity in a similar fashion to §1 to derive a bound for $S(\alpha)$. This is the essence of Ghosh's original proof of (24).

One may hope that when $\sigma > 1/8$ one can use the above Type I and II bounds through the alternative sieve in a similar manner to §1.3, but that is not straightforward. Indeed, when $\sigma > 1/8$, we can estimate neither a Type I nor a Type II sum with $N^{1-4\sigma} \ll M \ll N^{4\sigma}$. The reader should compare this to the situation at the beginning of §1.2, where the Type II information required by Vaughan's identity disappeared, but Type I sum bounds were still available. In general, such breakdowns in analytic information pose significant problems to Harman's alternative sieve. Here, however, we can avoid those problems, thanks to an alternative Type I sum bound due to G. Harman [6]. Harman's result extends the range of Type I information to $M \le N^{1-3\sigma}$. Moreover, when $\sigma < 1/7$, our Type II sum bound boosts the range of the Type I bound to $M \le N^{1-2\sigma}$. Once we have an estimate for Type I sums with such a range, our Type I and Type II information mirrors the information available in §1.2 (with N and 2σ in place of X and θ). We can then easily obtain a version of Lemma 1.3. On the other hand, when $1/7 < \sigma < 1/6$, the argument behind Lemma 1.3 yields a Type I/II bound only for $M \le X^{1-4\sigma}$. We summarize all these observations in the next lemma.

Lemma 2.2. *Suppose that $0 < \sigma < 1/6$ and $\alpha \in \mathfrak{m}_\sigma$, and suppose that (a_m) and (b_k) are complex sequences with $|a_m| \le 1$ and $|b_k| \le 1$. Then* (26) *holds, provided that any of the following sets of conditions is satisfied:*

- Type I: $M \le N^{1-3\sigma}$ *and* $b_k = 1$*;*
- Type II: $N^{2\sigma} \le M \le N^{1-4\sigma}$ *or* $N^{4\sigma} \le M \le N^{1-2\sigma}$*;*
- Type I/II: $M \le N^{1-4\sigma}$ *and* $b_k = \psi(k,z)$*, with* $z \le N^{1-6\sigma}$*.*

Furthermore, if $0 < \sigma < 1/7$*, then* (26) *holds when*

- Type I/II: $M \le N^{1-2\sigma}$ *and* $b_k = \psi(k,z)$*, with* $z \le N^{1-6\sigma}$*.*

When $1/7 < \sigma < 1/6$, we can still use the longer range of the above Type I bound to obtain a stronger result for Type I/II sums with some additional structure. In [9, Lemma 10], G. Harman and the author established the following result, in which the product $M = RS$ can be as large as $N^{1-3\sigma}$.

Lemma 2.3. *Suppose that* $0 < \sigma < 1/6$*, and assume the notation of Lemma 2.2. Then* (26) *holds, provided that* $b_k = \psi(k, z)$*, with* $z \le N^{1-6\sigma}$*, and* a_m *is a convolution of the form*

$$a_m = \sum_{\substack{rs=m \\ r\sim R, s\sim S}} c_r d_s,$$

where $R \le N^{2\sigma}$*,* $S \le N^{1-5\sigma}$*, and* (c_r) *and* (d_s) *are complex sequences with* $|c_r| \le 1$ *and* $|d_s| \le 1$.

This lemma and earlier results on cubic sums [15, 16, 35] motivated L. Zhao and the author to seek a further improvement on Harman's Type I estimate. In [17, Lemma 3.2], we prove the following bound for Type I sums with "extra structure" to the unknown coefficients.

Lemma 2.4. *Suppose that* $0 < \sigma < 1/6$*, and assume the notation of Lemma 2.2. Then* (26) *holds, provided that* $b_k = 1$ *and* a_m *is a convolution of the form*

$$a_m = \sum_{\substack{rs=m \\ r\sim R, s\sim S}} c_{r,s},$$

where $R \le N^{1-3\sigma}$*,* $RS^2 \le \varepsilon N^{1-2\sigma}$*, and* $(c_{r,s})$ *is a complex sequence with* $|c_{r,s}| \le 1$.

The reader may compare this estimate with a result of R.C. Baker and G. Harman [1] on the special case where $\alpha = a/q$, with q large. In that context, Baker and Harman were able to deal with general Type I sums with $M \le N^{1-5\sigma/2}$, whereas Lemma 2.4 allows for products $M = RS$ that can be as large as $N^{1-5\sigma/2}$ only in some special cases. Nonetheless, the above lemma proves quite useful in extending Lemma 2.3. We have the following result [17, Lemma 3.4].

Lemma 2.5. *Suppose that* $0 < \sigma < 1/6$*, and assume the notation of Lemma 2.2. Then* (26) *holds, provided that* $b_k = \psi(k, z)$*, with* $z \le N^{1-6\sigma}$*, and* a_m *is a convolution of the form*

$$a_m = \sum_{\substack{rs=m \\ r\sim R, s\sim S}} c_r d_s,$$

where $R \le N^{2\sigma}$*,* $S \le N^{2\sigma}$*,* $RS \le N^{1-3\sigma}$*, and* (c_r) *and* (d_s) *are complex sequences with* $|c_r| \le 1$ *and* $|d_s| \le 1$.

Sketch of the proof. As in the proof of Lemma 1.3, let $\Pi(z) = \prod_{p \le z} p$, and bound the given sum by

$$(\log X) \bigg| \sum_{r\sim R} \sum_{s\sim S} \sum_{\substack{d|\Pi(z) \\ d\sim D}} \sum_{rsdk\sim N} c_r d_s \mu(d) e(\alpha r^2 s^2 d^2 k^2) \bigg|,$$

for some $D \ll N/(RS)$. We consider three cases:

Case 1: $DRS \le N^{1-3\sigma}$. Then the above sum is a Type I sum covered by Lemma 2.2.

Case 2: $DRS > N^{4\sigma}$, or $DR > N^{2\sigma}$, or $DS > N^{2\sigma}$. Then we can argue as in Case 2 of the proof of Lemma 1.3 to bound the above sum by a linear combination of $O\big((\log N)^3\big)$ Type II sums.

Case 3: $N^{1-3\sigma} < DRS \le N^{4\sigma}$, $DR \le N^{2\sigma}$, and $DS \le N^{2\sigma}$. Then have

$$RS \le N^{1-3\sigma}, \quad RSD^2 = (DR)(DS) \le N^{4\sigma} < 0.1 N^{1-2\sigma},$$

so we can refer to Lemma 2.4 with $(r, s) = (rs, d)$. □

2.3. **A first sieve result.** In this section, we outline an application of the sieve ideas from §1.3 to the present problem and establish the following result due to G. Harman and the author [8].

Theorem 2.6 (Harman & K., 2006)**.** *Let $\varepsilon > 0$ be fixed. Then $E_4(X) \ll X^{5/14+\varepsilon}$.*

The sieve method employed in §1.3 is based on an inequality of the form

$$\psi(m, X^{1/2}) = \rho_1(m) + \rho_2(m) \ge \rho_1(m), \tag{27}$$

where $\rho_2(m) \ge 0$. The function ρ_1 in (27) represents the total contribution of terms in the decomposition that give rise to acceptable sums of Types I/II or II: the coefficient $\psi(m,z)$ of Σ_1 and the coefficients of Σ_4 and Σ_6 are terms in $\rho_1(m)$. The function ρ_2 represents the contribution of "bad" triple and quintuple sums, whose removal generated the "losses" measured by the integrals $b_9, \dots, b_{12}$. The above identity was constructed by pure combinatorics driven by two competing objectives: to minimize the total loss generated by the dismissal of ρ_2; and to ensure that the exponential sum with coefficients ρ_1 satisfies a bound similar to (26) above. Observe that when $0 < \sigma < 1/7$, the size restrictions on Type I/II and Type II sums in Lemma 2.2 match exactly those in §1 with 2σ and N in place of θ and X. Thus, the combinatorial argument from §1.3 (with $N, 2\sigma$ in place of X, θ) yields also an arithmetic function $\rho = \rho_1$ such that:

- We have $\rho(m) = O(1)$ and $\psi(m, N^{1/2}) \ge \rho(m)$ for all $m \sim N$.
- When $0 < \sigma < 1/7$,

$$\max_{\alpha \in \mathfrak{m}_\sigma} \Big| \sum_{m \sim N} \rho(m) e(\alpha m^2) \Big| \ll N^{1-\sigma+\varepsilon}. \tag{28}$$

Let ρ be the above arithmetic function. We use the notation introduced in §2.1. We have

$$R(n) \ge R(n;\rho) = \sum_{\substack{m^2+p_1^2+p_2^2+p_3^2=n \\ m, p_i \sim N}} \rho(m), \tag{29}$$

and by orthogonality,

$$R(n;\rho) = \int_0^1 S(\alpha)^3 T(\alpha) e(-\alpha n)\, d\alpha, \quad T(\alpha) = T(\alpha;\rho) = \sum_{m \sim N} \rho(m) e(\alpha m^2). \tag{30}$$

As a primary dissection of the unit interval into major and minor arcs, we use $\mathfrak{M} = \mathfrak{M}_{\sigma/2}$ and $\mathfrak{m} = \mathfrak{m}_{\sigma/2}$. We consider also an auxiliary dissection into $\mathfrak{N} = \mathfrak{M}_\sigma$ and $\mathfrak{n} = \mathfrak{m}_\sigma$. We define a function Δ on $\mathfrak{N}$ by setting

$$\Delta(\alpha) = (q + N^2|q\alpha - a|)^{-1} \qquad \text{when } |q\alpha - a| < N^{4\sigma - 2}.$$

When $0 < \sigma < 1/6$, a result of X. Ren [26] yields

$$S(\alpha) \ll N^{1+\varepsilon}\Delta(\alpha)^{1/2} + N^{5/6+\varepsilon} \qquad (\alpha \in \mathfrak{N}). \tag{31}$$

Note that the first term in this bound is $\ll N^{1-\sigma+\varepsilon}$ when $\alpha \in \mathfrak{m} \cap \mathfrak{N}$. Hence, we can combine (28) and (31) to show that

$$\max_{\alpha \in \mathfrak{m}} \min\big(|S(\alpha)|, |T(\alpha)|\big) \ll N^{1-\sigma+\varepsilon}.$$

Then, arguing similarly to (22), we have

$$\int_{\mathfrak{m}} S(\alpha)^3 T(\alpha) Z(\alpha)\, d\alpha \ll N^{1-\sigma+\varepsilon}\Big(\int_0^1 |S(\alpha)|^4\, d\alpha\Big)^{1/2}\Big(\int_0^1 \big(|S(\alpha)|^2 + |T(\alpha)|^2\big)|Z(\alpha)|^2\, d\alpha\Big)^{1/2} \\ \ll N^{2-\sigma+\varepsilon}\big(E(X)^{1/2}N^{1/2} + E(X)\big). \tag{32}$$

Earlier, we declared the integral over the major arcs to be the "easy" one, but we need to return to that and examine how that work is affected when we replace one of the exponential sums $S(\alpha)$ by $T(\alpha)$. The proof of (20) has two stages. The first, where the bulk of the effort is spent, is to show that

$$\int_{\mathfrak{M}} \big(S(\alpha)^4 - S^*(\alpha)^4\big) e(-\alpha n)\, d\alpha \ll N^2 (\log N)^{-A} \tag{33}$$

for any fixed $A > 4$. Here, $S^*(\alpha)$ is the expected major arc approximation to $S(\alpha)$, defined for $\alpha = a/q + \beta$ by

$$S^*(a/q+\beta) = \phi(q)^{-1}\Big(\sum_{h\in\mathbb{Z}_q^*} e(ah^2/q)\Big) \sum_{m\sim N} \frac{e(\beta m^2)}{\log m}.$$

Once we have (33), it is relatively easy to show that

$$\int_{\mathfrak{M}} S^*(\alpha)^4 e(-\alpha n)\, d\alpha \approx \kappa_n N^2 (\log N)^{-4}, \tag{34}$$

which completes the proof of (20). To illustrate how this program is affected by the sieve, we will focus on one of the terms in the explicit definition of ρ, but the same ideas can be used to deal with all the other terms. Let $\psi_1(m) = \psi(m,z)$, $z = N^{1-6\sigma}$, and write $S_1(\alpha) = T(\alpha;\psi_1)$. The quantity $R_1(n) = R(n;\psi_1)$ is then a variant of the sum Σ_1 in §1.2.

In the analysis of $R_1(n)$, (33) and (34) are replaced by

$$\int_{\mathfrak{M}} \big(S(\alpha)^3 S_1(\alpha) - S^*(\alpha)^3 S_1^*(\alpha)\big) e(-\alpha n)\, d\alpha \ll N^2 (\log N)^{-A}, \tag{35}$$

$$\int_{\mathfrak{M}} S^*(\alpha)^3 S_1^*(\alpha) e(-\alpha n)\, d\alpha \approx c_1 \kappa_n N^2 (\log N)^{-4}, \tag{36}$$

respectively. Here, $c_1 = \zeta^{-1}\omega(1/\zeta)$ with $\zeta = 1-6\sigma$, and $S_1^*(\alpha)$ is the expected major arc approximation to $S_1(\alpha)$, defined for $\alpha = a/q+\beta$ by

$$S_1^*(a/q+\beta) = \phi(q)^{-1}\Big(\sum_{h\in\mathbb{Z}_q^*} e(ah^2/q)\Big) \sum_{m\sim N} \frac{e(\beta m^2)}{\log z}\,\omega\Big(\frac{\log m}{\log z}\Big),$$

where ω is Buchstab's function. The constant c_1 in (36) is similar to the constants c_i in §1 and satisfies

$$\sum_{m\sim N} \frac{e(\beta m^2)}{\log z}\,\omega\Big(\frac{\log m}{\log z}\Big) \approx c_1 \sum_{m\sim N} \frac{e(\beta m^2)}{\log m} \tag{37}$$

for small β. (Since we are using essentially the same combinatorial construction as in §1.3, c_1 is, in fact, the constant c_1 in §1.3 with 2σ in place of θ.) Between (35) and (36), (36) is the easier by far. Indeed, it is a standard exercise to deduce (36) from (34) and (37). The proof of (35) is another story. The methods from the proof of (33) can be adapted without too much extra difficulty when the parameter P in the definition of $\mathfrak{M}$ satisfies $1 \le P \le z^{1-\varepsilon}$, but this inequality fails for our chosen P and z. To get around this obstacle, we show in the next section that (35) fails for $\ll X^{\varepsilon}$ integers $n \equiv 4 \pmod{24}$ with $n \sim X$. Assuming that result, we can now use (30) and (32) (with $\rho = \psi_1$) to state that when $0 < \sigma < 1/7$, we have

$$R_1(n) \approx c_1 \kappa_n N^2 (\log N)^{-4},$$

with $\ll X^{1/2-\sigma+\varepsilon}$ exceptions (henceforth, we reserve the term "exception" for integers $n \equiv 4 \pmod{24}$ with $n \sim X$). This approximation is an analogue of (14). Using the same ideas, we can use (28) to show that

$$R(n;\rho) \approx (1 - b_9(2\sigma) - \cdots - b_{12}(2\sigma))\kappa_n N^2 (\log N)^{-4}, \tag{38}$$

with $\ll X^{1/2-\sigma+\varepsilon}$ exceptions. Here, $b_9, \dots, b_{12}$ are the double and quadruple integrals defined in §1.3 to measure the losses incurred in the sieve from discarding the contributions of "bad" subsums of Σ_7 and Σ_9. In particular, it is known from the numerical work behind the proof of Theorem 1.4 that

$$b_9(2/7) + \cdots + b_{12}(2/7) < 0.11.$$

Consequently, on choosing $\sigma = 1/7 - \varepsilon$, we deduce from (29) and (38) that

$$R(n) \geq 0.89 \kappa_n N^2 (\log N)^{-4}$$

with $\ll X^{5/14+\varepsilon}$ exceptions.

Remark. The reader should take a note of the breakdown that occurs at $\sigma = 1/7$ in the present problem: at that point, the last claim of Lemma 2.2 no longer holds, and the exponential sum estimates in that lemma no longer match the respective estimates in §1. Even worse, as σ crosses over 1/7, the acceptable range for the parameter M in the definition of a Type I/II sum drops from $M \leq N^{5/7+\varepsilon}$ to $M \leq N^{3/7-\varepsilon}$. This poses some serious challenges to the sieve method. We will deal with those issues in §2.5, but first we explain L. Zhao's clever idea for proving approximations like (35) with only "a few" exceptions.

2.4. **Zhao's major-arc idea.** We now assume that $0 < \sigma < 1/6$ and the major arcs $\mathfrak{M}$ are defined as in §2.3. We also consider an auxiliary set of major arcs, $\mathfrak{M}_0$ defined with $P = N^{\delta}$ and $Q = N^{2-\delta}$, where $\delta > 0$ is a (small) fixed number.

The proof of (33) relies on two basic facts about the primes:

(P_1) Primes are well-distributed in arithmetic progressions with small moduli. The analytic formulation of this fact, the Siegel–Walfisz theorem, states: For any fixed $A, B > 0$, and any primitive Dirichlet character χ with a modulus $\leq (\log N)^B$, one has

$$\sum_{p \sim N} \chi(p) = E(\chi) \sum_{m \sim N} (\log m)^{-1} + O\big(N(\log N)^{-A}\big),$$

where $E(\chi)$ is 1 when χ is the trivial character, and $E(\chi) = 0$ otherwise.

(P_2) One can estimate certain mean values of Dirichlet polynomials over primes: If $\mathcal{H}$ is a set of primitive Dirichlet characters with moduli q, where $q \leq Q$ and $r \mid q$, then

$$\sum_{\chi \in \mathcal{H}} \int_T^{2T} \Big| \sum_{m \sim N} \Lambda(m)\chi(m) m^{-1/2-it} \Big| \, dt \ll \big(N^{1/2} + H^{1/2}N^{3/10} + H\big)(\log N)^c, \tag{39}$$

where $H = r^{-1}Q^2T$ and $c > 0$ is an absolute constant. (This bound was established by J.Y. Liu [19].) In particular, the condition $P \leq N^{2/5-\varepsilon}$ imposed in §2.1 results from the term $H^{1/2}N^{3/10}$ in this bound.

To make use of these two facts, one first uses the orthogonality of Dirichlet characters to express $S(a/q+\beta)$ as a linear combination over the Dirichlet characters modulo q. If $\alpha = a/q+\beta$, with β small and $q \le P$, we write

$$S(a/q+\beta) = \sum_{h\in\mathbb{Z}_q^*} e(ah^2/q) \sum_{\substack{p\sim N\\ p\equiv h \pmod q}} e(\beta p^2) = \phi(q)^{-1} \sum_{\chi \bmod q} \sum_{h\in\mathbb{Z}_q^*} \bar{\chi}(h) e(ah^2/q) \sum_{p\sim N} \chi(p) e(\beta p^2). \tag{40}$$

This is justified, because when $P < N/2$, the primes $p \sim N$ fall only in residue classes $h \bmod q$ with $(h,q) = 1$.

To prove an approximation like (35), one needs versions of (P_1) and (P_2) for sums featuring the coefficients $\psi(m,z)$ in place of the characteristic function of the primes and von Mangoldt's function. Such a generalization of (P_1) is straightforward to obtain. A version of the mean-value estimate (P_2) can also be obtained with the tools used by Liu to obtain (39). However, trying to extend (40) to $S_1(a/q+\beta)$, one encounters a very serious technical obstacle: if p_0 is a prime with $z < p_0 \le P$, then there will be moduli q, $z < q \le P$, which are divisible by p_0, and for such moduli the sum $S_1(a/q+\beta)$ will include terms with $(m,q) = p_0$. Such terms caused a lot of headaches in the author's joint work with Harman [8, 9]. On the other hand, when $z = N^{1-6\sigma}$ and $\delta < 1-6\sigma$, the above issue does not affect the exponential sum $S_1(\alpha)$ on $\mathfrak{M}_0$. Hence, it becomes straightforward to generalize the proof of (33) and to show that

$$\int_{\mathfrak{M}_0} \big(S(\alpha)^3 S_1(\alpha) - S^*(\alpha)^3 S_1^*(\alpha)\big) e(-\alpha n)\, d\alpha \ll N^2 (\log N)^{-A}.$$

Let $\mathfrak{R} = \mathfrak{M} \setminus \mathfrak{M}_0$. A quick calculation gives

$$\int_{\mathfrak{R}} S^*(\alpha)^3 S_1^*(\alpha) e(-\alpha n)\, d\alpha \ll N^{2-\varepsilon},$$

so (35) will follow if we show that

$$\left| \int_{\mathfrak{R}} S(\alpha)^3 S_1(\alpha) e(-\alpha n)\, d\alpha \right| \le N^{2-\varepsilon}. \tag{41}$$

Let $\mathcal{E}_0(X)$ denote the set of integers $n \equiv 4 \pmod{24}$, with $n \sim X$, such that (41) fails. Since the set $\mathfrak{R}$ is symmetric about 0 modulo one and the coefficients of the exponential sums are real, the integral in (41) is real; let $\theta_n \in \{\pm 1\}$ denote its sign. We have

$$|\mathcal{E}_0(X)| N^{2-\varepsilon} \ll \int_{\mathfrak{R}} S(\alpha)^3 S_1(\alpha) Z_0(\alpha)\, d\alpha, \qquad Z_0(\alpha) = \sum_{n\in\mathcal{E}_0(X)} \theta_n e(-\alpha n). \tag{42}$$

Since $\mathfrak{R} \subset \mathfrak{M} \subset \mathfrak{N}$, when $0 < \sigma < 1/6$ and $\alpha \in \mathfrak{R}$, we deduce from (31) that

$$S(\alpha)^2 \ll N^{2+\varepsilon} \Delta(\alpha).$$

Variants of (22) and of [34, (3.28)] then yield

$$\begin{aligned} \int_{\mathfrak{R}} S(\alpha)^3 S_1(\alpha) Z_0(\alpha)\, d\alpha &\ll N^{2+\varepsilon} \Big(\int_{\mathfrak{R}} |Z_0(\alpha)|^2 \Delta(\alpha)^2\, d\alpha \Big)^{1/2} \Big(\int_0^1 \big(|S(\alpha)|^4 + |S_1(\alpha)|^4\big)\, d\alpha \Big)^{1/2} \\ &\ll N^{3+3\varepsilon/2} \Big(\int_{\mathfrak{R}} |Z_0(\alpha)|^2 \Delta(\alpha)^2\, d\alpha \Big)^{1/2} \\ &\ll N^{2+2\varepsilon} \big(|\mathcal{E}_0(X)|^{1/2} + |\mathcal{E}_0(X)| N^{-\delta/2} \big). \end{aligned} \tag{43}$$

Comparing (42) and (43), we see that (41) holds with $\ll X^{\varepsilon}$ exceptions; hence, (35) also holds with $\ll X^{\varepsilon}$ exceptions.

2.5. **A "full-scale" alternative sieve.** In this section, we sketch the proof of the following recent theorem due to L. Zhao and the author [17].

Theorem 2.7 (Zhao & K., 2015+)**.** *One has $E_4(X) \ll X^{11/32}$.*

When $1/7 < \sigma < 1/6$, a simple lower bound sieve of the form (27) no longer suffices, since it is not possible to avoid "bad" exponential sums that correspond to negative terms in the decomposition. Instead, we combine lower and upper sieve functions. We will construct decompositions

$$\psi(m, N^{1/2}) = \rho_1(m) - \rho_2(m) + \rho_3(m) \geq \rho_1(m) - \rho_2(m), \tag{44}$$

$$\psi(m, N^{1/2}) = \rho_4(m) - \rho_5(m) \leq \rho_4(m), \tag{45}$$

where $\rho_i(m) \geq 0$ for $i = 2, 3, 5$ and the bound (28) holds for $\rho = \rho_1, \rho_4$. In the case $\sigma > 1/7$, we will obtain such bounds for Type II sums using Lemma 2.2 and for Type I/II sums using Lemma 2.5. Combining (44) and (45), we obtain

$$\psi(m, N^{1/2})\psi(k, N^{1/2}) \geq \rho_1(m)\psi(k, N^{1/2}) - \rho_2(m)\rho_4(k).$$

Using this inequality in place of (27), we obtain the following variant of (29)–(30):

$$R(n) \geq \int_0^1 S(\alpha)^2 (T_1(\alpha)S(\alpha) - T_2(\alpha)T_4(\alpha)) e(-\alpha n)\, d\alpha,$$

where $T_i(\alpha) = T(\alpha; \rho_i)$. Since our construction will ensure that $T_1(\alpha)$ and $T_4(\alpha)$ satisfy (28), the ideas from the last two sections can be applied to the above integral to show that

$$R(n) \geq (C_1 - C_2 C_4 + o(1))\kappa_n N^2 (\log N)^{-4} \tag{46}$$

with $\ll X^{1/2-\sigma+\varepsilon}$ exceptions. Here, the numbers C_i, $i = 1, 2, 4$, are the constants that appear in the analogues of (37) for the respective sums T_i. Next, we discuss the combinatorial identities (44) and (45).

For $3/20 < \sigma < 1/6$, we put

$$z = N^{1-6\sigma}, \quad V = N^{2\sigma}, \quad W = N^{1-4\sigma}, \quad Y = N^{1-3\sigma}.$$

As before, we treat σ as a numerical parameter to be chosen later in the proof of our theorem; its value will eventually be set to $\sigma = 5/32 + 10^{-4}$.

We first describe the identity (44). By Buchstab's identity (11),

$$\begin{aligned} \psi(m, N^{1/2}) &= \psi(m, z) - \Big\{ \sum_{z \leq p < V} + \sum_{V \leq p \leq W} + \sum_{W < p < N^{1/2}} \Big\} \psi(m/p, p) \\ &= \psi_1(m) - \psi_2(m) - \psi_3(m) - \psi_4(m), \quad \text{say.} \end{aligned} \tag{47}$$

Note that $T(\alpha; \psi_1)$ and $T(\alpha; \psi_3)$ are acceptable Type I/II and Type II sums, respectively. Hence, we include ψ_1 and ψ_3 in ρ_1; we include ψ_4 in ρ_2 and decompose ψ_2 further. Another application of Buchstab's identity gives

$$\psi_2(m) = \sum_{z \leq p_1 < V} \Big\{ \psi(m/p_1, z) - \sum_{z \leq p_2 < p_1 < V} \psi(m/(p_1 p_2), p_2) \Big\} = \psi_5(m) - \psi_6(m), \quad \text{say.} \tag{48}$$

We now write

$$\psi_6(m) = \psi_7(m) + \cdots + \psi_{10}(m), \tag{49}$$

where ψ_i is the part of ψ_6 subject to the following extra conditions on the product p_1p_2:

- $\psi_7(m)$: $p_1p_2 < V$;
- $\psi_8(m)$: $V \le p_1p_2 \le W$;
- $\psi_9(m)$: $W < p_1p_2 \le Y$;
- $\psi_{10}(m)$: $p_1p_2 > Y$.

In our final decomposition, ψ_5 and ψ_8 contribute to ρ_1 and ψ_{10} contributes to ρ_3; we give further decompositions of ψ_7 and ψ_9.

We apply (11) twice more to ψ_7:

$$\begin{aligned}\psi_7(m) &= \sum_{p_1,p_2}\Big\{\psi(m/(p_1p_2),z) - \sum_{z\le p_3<p_2}\psi(m/(p_1p_2p_3),z) \\ &\qquad\qquad + \sum_{z\le p_4<p_3<p_2}\psi(m/(p_1p_2p_3p_4),p_4)\Big\} \\ &= \psi_{11}(m) - \psi_{12}(m) + \psi_{13}(m), \quad \text{say.}\end{aligned} \tag{50}$$

We next apply Buchstab's identity to ψ_9 and obtain

$$\begin{aligned}\psi_9(m) &= \sum_{p_1,p_2}\Big\{\psi(m/(p_1p_2),z) - \sum_{z\le p_3<p_2}\Big(\sum_{p_1p_2p_3\le Y} + \sum_{p_1p_2p_3>Y}\Big)\psi(m/(p_1p_2p_3),p_3)\Big\} \\ &= \psi_{14}(m) - \psi_{15}(m) - \psi_{16}(m), \quad \text{say.}\end{aligned} \tag{51}$$

Note that the summation conditions in ψ_{15} imply $p_2p_3 \le P^{2/3-2\sigma} \le W$. Thus, a final application of (11) yields

$$\begin{aligned}\psi_{15}(m) &= \sum_{p_1,p_2,p_3}\Big\{\sum_{p_2p_3\ge V} + \sum_{p_2p_3<V}\Big\}\psi(m/(p_1p_2p_3),p_3) \\ &= \psi_{17}(m) + \sum_{\substack{p_1,p_2,p_3\\ p_2p_3<V}}\Big\{\psi(m/(p_1p_2p_3),z) - \sum_{z\le p_4<p_3}\psi(m/(p_1\cdots p_4),p_4)\Big\} \\ &= \psi_{17}(m) + \psi_{18}(m) - \psi_{19}(m), \quad \text{say.}\end{aligned} \tag{52}$$

Finally, we split ψ_{13}, ψ_{16}, and ψ_{19} into "good" and "bad" parts, which we denote ψ_j^g and ψ_j^b, respectively. We collect in ψ_j^g the terms in ψ_j in which a subproduct of $p_1p_2p_3p_4$ lies within the ranges $[V,W]$ or $[N/W,N/V]$ (this makes $T(\alpha;\psi_j^g)$ an acceptable Type II sum); those terms will contribute to ρ_1. The remaining terms in ψ_j are placed in ψ_j^b and will contribute to ρ_2 or ρ_3, depending on their respective signs.

Combining (47)–(52), we now have (44) with

$$\begin{aligned}\rho_1(m) &= \psi_1(m) - \psi_3(m) - \psi_5(m) + \psi_8(m) + \psi_{11}(m) - \psi_{12}(m) + \psi_{13}^g(m) \\ &\qquad + \psi_{14}(m) - \psi_{16}^g(m) - \psi_{17}(m) - \psi_{18}(m) + \psi_{19}^g(m), \\ \rho_2(m) &= \psi_4(m) + \psi_{16}^b(m), \qquad \rho_3(m) = \psi_{10}(m) + \psi_{13}^b(m) + \psi_{19}^b(m).\end{aligned}$$

We note (again) that each term ψ_j that appears in ρ_1 leads to an exponential sum that can be estimated using Lemmas 2.2 or 2.5, and that each term $\psi_j^g(m)$ leads to a sum that can be estimated using Lemma 2.2.

We now turn to (45). We have

$$\psi_2(m) = \Big\{ \sum_{z \le p \le Y^{1/2}} + \sum_{Y^{1/2} < p < V} \Big\} \psi(m/p, p) = \psi_{20}(m) + \psi_{21}(m), \quad \text{say.} \tag{53}$$

The term ψ_{21} will contribute to ρ_5; we apply (11) twice to ψ_{20}. That gives

$$\begin{aligned} \psi_{20}(m) = \sum_{z \le p_1 \le Y^{1/2}} \Big\{ \psi(m/p_1, z) - \sum_{z \le p_2 < p_1} \psi(m/(p_1p_2), z) \\ + \sum_{z \le p_3 < p_2 < p_1} \psi(m/(p_1p_2p_3), p_3) \Big\} \\ = \psi_{22}(m) - \psi_{23}(m) + \psi_{24}(m), \quad \text{say.} \end{aligned} \tag{54}$$

We split ψ_{24} into "good" and a "bad" parts, and then further split $\psi_{24}^b(m)$ in two:

$$\psi_{24}^b(m) = \sum_{p_1,p_2,p_3} \Big\{ \sum_{p_1p_2p_3^2 \le Y} + \sum_{p_1p_2p_3^2 > Y} \Big\} \psi(m/(p_1p_2p_3), p_3) = \psi_{25}(m) + \psi_{26}(m), \quad \text{say.} \tag{55}$$

We apply Buchstab's identity two more times to ψ_{25}:

$$\begin{aligned} \psi_{25}(m) = \sum_{p_1,p_2,p_3} \Big\{ \psi(m/(p_1p_2p_3), z) - \sum_{z \le p_4 < p_3} \psi(m/(p_1 \cdots p_4), z) \\ + \sum_{z \le p_5 < p_4 < p_3} \psi(m/(p_1 \cdots p_5), p_5) \Big\} \\ = \psi_{27}(m) - \psi_{28}(m) + \psi_{29}(m), \quad \text{say.} \end{aligned} \tag{56}$$

Finally, we split ψ_{26} and ψ_{29} into "good" and "bad" subsums. We remark that the summation conditions in ψ_{25} imply $p_1p_3 \le W$. (Otherwise, we would have $p_2p_3 \le P^{\sigma}$, whence $p_3 \le P^{\sigma/2}$ and $p_1p_3 \le P^{1/2-\sigma}$; the latter contradicts the assumption $p_1p_3 > W$ when $\sigma < 1/6$.) Therefore, the exponential sums $T(\alpha;\psi_{27})$ and $T(\alpha;\psi_{28})$ can be estimated either by Lemma 2.2 (when $p_1p_3 \ge V$) or by Lemma 2.5 (when $p_1p_3 < V$). Combining (47) and (53)–(56), we deduce (45) with

$$\begin{aligned} \rho_4(m) &= \psi_1(m) - \psi_3(m) - \psi_5(m) - \psi_{22}(m) + \psi_{23}(m) - \psi_{24}^g(m) \\ &\quad - \psi_{26}^g(m) - \psi_{27}(m) + \psi_{28}(m) - \psi_{29}^g(m), \\ \rho_5(m) &= \psi_4(m) + \psi_{21}(m) + \psi_{26}^b(m) + \psi_{29}^b(m). \end{aligned}$$

It is clear from the above decompositions that $T(\alpha;\rho_i)$, $i = 1,4$, are acceptable on the minor arcs and that all three sums $T(\alpha;\rho_i)$, $i = 1,2,4$, are supported on integers not divisible by primes p, with $p \le z$. Hence, we have (46) with $\ll X^{1/2-\sigma+\varepsilon}$ exceptions. The constants C_i in (46) are linear combinations of multiple integrals similar to those that appeared in §1. For example,

$$C_2 = \log\left(\frac{4\sigma}{1-4\sigma}\right) + \iiint_{\mathcal{D}_{16}} \omega\left(\frac{1-u_1-u_2-u_3}{u_3}\right) \frac{du_1du_2du_3}{u_1u_2u_3^2},$$

where $\mathcal{D}_{16}$ is the set in $\mathbb{R}^3$ defined by the conditions

$$1-6\sigma \le u_3 \le u_2 \le u_1 \le 2\sigma, \quad 1-4\sigma \le u_1+u_2 \le 1-3\sigma \le u_1+u_2+u_3,$$
$$\text{no subsum of } u_1+u_2+u_3 \text{ lies in the set } [2\sigma, 1-4\sigma] \cup [4\sigma, 1-2\sigma].$$

Numerical evaluation of the various multiple integrals reveals that when $\sigma = 5/32 + 10^{-4}$, we have

$$C_1 > 1.665, \quad C_2 < 0.769, \quad C_4 < 2.096, \quad C_1 - C_2C_4 > 0.05.$$

Hence, when $\sigma = 5/32 + 10^{-4}$, the lower bound in (46) is non-trivial with $\ll X^{11/32}$ exceptions.[5]

Acknowledgment. The author would like to thank the Morningside Center for Mathematics at the Chinese Academy of Sciences for its hospitality during the Workshop on Number Theory, July 20–28, 2014. In particular, the proofs of Theorem 2.7 and of Lemmas 2.4 and 2.5 were completed during the workshop.

REFERENCES

[1] R.C. Baker and G. Harman, *On the distribution of αp^k modulo one*, Mathematika **38** (1991), 170–184.
[2] C. Bauer, M.C. Liu, and T. Zhan, *On sums of three prime squares*, J. Number Theory **85** (2000), 336–359.
[3] A. Ghosh, *The distribution of αp^2 modulo 1*, Proc. London Math. Soc. (3) **42** (1981), 252–269.
[4] G. Harman, *On the distribution of αp modulo one*, J. London Math. Soc. (2) **27** (1983), 9–18.
[5] G. Harman, *On the distribution of αp modulo one II*, Proc. London Math. Soc. (3) **72** (1996), 241–260.
[6] G. Harman, *The values of ternary quadratic forms at prime arguments*, Mathematika **51** (2004), 83–96.
[7] G. Harman, *Prime-Detecting Sieves*, Princeton University Press, Princeton, NJ, 2007.
[8] G. Harman and A.V. Kumchev, *On sums of squares of primes*, Math. Proc. Cambridge Phil. Soc. **140** (2006), 1–13.
[9] G. Harman and A.V. Kumchev, *On sums of squares of primes II*, J. Number Theory. **130** (2010), 1969–2002.
[10] D.R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), 1365–1377.
[11] D.R. Heath-Brown and C.H. Jia, *The distribution of αp modulo one*, Proc. London Math. Soc. (3) **84** (2002), 79–104.
[12] L.K. Hua, *Some results in additive prime number theory*, Quart. J. Math. Oxford **9** (1938), 68–80.
[13] C.H. Jia, *On the distribution of αp modulo one*, J. Number Theory **45** (1993), no. 3, 241–253.
[14] C.H. Jia, *On the distribution of αp modulo one II*, Sci. China Ser. A **43** (2000), 703–721.
[15] K. Kawada and T.D. Wooley, *On the Waring–Goldbach problem for fourth and fifth powers*, Proc. London Math. Soc. (3) **83** (2001), 1–50.
[16] A.V. Kumchev, *On Weyl sums over primes and almost primes*, Michigan Math. J. **54** (2006), 243–268.
[17] A.V. Kumchev and L. Zhao, *Sums of four squares of primes*, preprint.
[18] M.C. Leung and M.C. Liu, *On generalized quadratic equations in three prime variables*, Monatsh. Math. **115** (1993), 133–169.
[19] J.Y. Liu, *On Lagrange's theorem with prime variables*, Quart. J. Math. Oxford (2) **54** (2003), 453–462.
[20] J.Y. Liu and M.C. Liu, *The exceptional set in the four prime squares problem*, Illinois J. Math. **44** (2000), 272–293.
[21] J.Y. Liu, T.D. Wooley and G. Yu, *The quadratic Waring–Goldbach problem*, J. Number Theory **107** (2004), 298–321.
[22] J.Y. Liu and T. Zhan, *Sums of five almost equal prime squares II*, Sci. China **41** (1998), 710–722.
[23] J.Y. Liu and T. Zhan, *Distribution of integers that are sums of three squares of primes*, Acta Arith. **98** (2001), 207–228.

[5] We remark that while the exponent 11/32 was chosen mostly for convenience, it is quite close to the real limit of the method, which is $\approx 11/32 - 3.6 \times 10^{-4}$.

[24] J.Y. Liu and T. Zhan, *The exceptional set in Hua's theorem for three squares of primes*, Acta Math. Sinica (N.S.) **21** (2005), 335–350.

[25] K. Matomäki, *The distribution of αp modulo one*, Math. Proc. Cambr. Philos. Soc. **147** (2009), 267–283.

[26] X. Ren, *On exponential sum over primes and application in Waring-Goldbach problem*, Sci. China Ser. A Math. (6) **48** (2005), 785–797.

[27] W. Schwarz, *Zur Darstellun von Zahlen durch Summen von Primzahlpotenzen*, J. reine angew. Math. **206** (1961), 78–112.

[28] R.C. Vaughan, *Sommes trigonométriques sure les nombres premiers*, C. R. Acad. Sci. Paris Sér. A **285** (1977), 981–983.

[29] R.C. Vaughan, *On the distribution of αp modulo 1*, Mathematika **24** (1977), 135–141.

[30] R.C. Vaughan, *The Hardy–Littlewood Method*, Second ed., Cambridge University Press, Cambridge, 1997.

[31] I.M. Vinogradov, *Representation of an odd number as the sum of three primes*, Dokl. Akad. Nauk SSSR **15** (1937), 291–294, in Russian.

[32] I.M. Vinogradov, *Some theorems concerning the theory of primes*, Mat. Sbornik N.S. **2** (1937), 179–195, in Russian.

[33] I.M. Vinogradov, *The Method of Trigonometric Sums in the Theory of Numbers*, Wiley–Interscience, 1954, translated from Russian.

[34] T.D. Wooley, *Slim exceptional sets for sums of four squares*, Proc. London Math. Soc. (3) **85** (2002), 1–21.

[35] L. Zhao, *On the Waring–Goldbach problem for fourth and sixth powers*, Proc. London Math. Soc. (3) **108** (2014), 1593–1622.