HUA'S LEMMA AND EXPONENTIAL SUMS OVER BINARY FORMS

Trevor D. Wooley*

Abstract. We establish mean value estimates for exponential sums over binary forms of strength comparable with the bounds attainable via classical, single variable estimates for diagonal forms. These new mean value estimates strengthen earlier bounds of the author when the degree d of the form satisfies $5 \leqslant d \leqslant 10$, the improvements stemming from a basic lemma which provides uniform estimates for the number of integral points on affine plane curves in mean square. Exploited by means of the Hardy-Littlewood method, these estimates permit one to establish asymptotic formulae for the number of integral zeros of equations defined as sums of binary forms of the same degree d, provided that the number of variables exceeds $\frac{17}{16}2^d$, improving significantly on what is attainable either by classical additive methods, or indeed the general methods of Birch and Schmidt.

1. Introduction. Rather general versions of the Hardy-Littlewood method due to Birch [2] and Schmidt [13] offer remarkably successful approaches to estimating the number of integral zeros of prescribed height satisfying a given homogeneous polynomial with integral coefficients. Both approaches require the polynomial under investigation to possess many variables in terms of its degree, and there are further hypotheses to be negotiated involving, directly or indirectly, the singular locus of the associated hypersurface. These unfortunate deficiencies of the method are significantly less pronounced when the polynomial under investigation is diagonal, which is to say, of the shape $a_1x_1^d + \cdots + a_sx_s^d$ (see Chapter 9 of Vaughan [19]), and such is also the case when the polynomial diagonalises over $\mathbb{C}$ (see Birch and Davenport [3]). The availability of superior analytic methods for the diagonal situation motivates investigation of polynomials intermediate in complexity between the diagonal ones, and the quite general homogeneous polynomials investigated by Birch and Schmidt, the hope being that insight will be obtained relevant to the general situation. One such intermediate situation is that in which the polynomial splits as a sum of binary homogeneous polynomials, and such has been investigated with some success for cubic forms by Chowla and Davenport [7], and more recently by Brüdern and Wooley [6]. The author [22] has rather recently obtained analogues

1991 *Mathematics Subject Classification.* 11D72, 11L07, 11E76, 11P55.

Key words and phrases. Exponential sums, binary forms, diophantine equations.

*Packard Fellow and supported in part by NSF grant DMS-9970440.

of Weyl's inequality and Hua's lemma for exponential sums over binary forms of higher degree, and thereby has made progress on problems involving sums of binary forms of arbitrary degree. This work was hindered by our lack of good uniform estimates for the number of integral points on affine plane curves. The object of this paper is to sharpen our earlier conclusions, and this we achieve by developing useful mean square estimates for the number of integral points on certain families of affine plane curves. It is to be hoped that progress will be stimulated in problems involving higher degree forms in many variables.

Before proceeding to the main thrust of this paper, it seems worthwhile to recall the conclusions stemming from the classical additive theory, and the work of Birch and Schmidt, so far as the density of integer points on hypersurfaces is concerned. First, on combining estimates of Weyl and Hua, one obtains the following classical conclusion (see Chapter 9 of Vaughan [19]).

Theorem A (Classical). *Let $a_1,\dots,a_s\in\mathbb{Z}\setminus\{0\}$ and write*

$$F(\mathbf{x})=a_1x_1^d+\dots+a_sx_s^d.$$

Then whenever $s>2^d$, one has

$$card(\{\mathbf{x}\in[-B,B]^s\cap\mathbb{Z}^s\ :\ F(\mathbf{x})=0\})\sim CB^{s-d},$$

where C denotes the "product of local densities" within the box $[-B,B]^s$.

In order to save space at this point, we avoid explaining what is meant by the term "product of local densities", and instead note merely that this number is positive and uniformly bounded away from zero whenever the equation $F(\mathbf{x})=0$ possesses non-singular real and p-adic solutions for every prime p. We refer the reader to Vaughan [17], [18], Heath-Brown [10] and Boklan [4] for the theory underlying the latest developments concerning the asymptotic formula in the diagonal situation. In order to describe Birch's theorem (see [2]), we recall that the singular locus of the hypersurface defined by the homogeneous equation $F(x_1,\dots,x_s)=0$ is the set of points $\mathbf{y}\in\mathbb{C}^s$ satisfying the equations

$$\frac{\partial F}{\partial x_1}(\mathbf{y})=\dots=\frac{\partial F}{\partial x_s}(\mathbf{y})=0.$$

Theorem B (Birch). *Let $F(\mathbf{x})\in\mathbb{Z}[x_1,\dots,x_s]$ be homogeneous of degree d, and suppose that the variety defined by the equation $F(\mathbf{x})=0$ has a singular locus of dimension at most D. Then whenever $s-D>(d-1)2^d$, one has*

$$card(\{\mathbf{x}\in[-B,B]^s\cap\mathbb{Z}^s\ :\ F(\mathbf{x})=0\})\sim CB^{s-d},$$

where C denotes the "product of local densities" within the box $[-B,B]^s$.

Mention of the singular locus is removed by Schmidt [13] at the cost of introducing an invariant h associated with the polynomial under consideration. When $F(\mathbf{x})\in\mathbb{Q}[x_1,\dots,x_s]$ is a form of degree $d>1$, write $h(F)$ for the least number h such that F may be written in the form

$$F=A_1B_1+A_2B_2+\dots+A_hB_h,$$

with A_i,B_i forms in $\mathbb{Q}[\mathbf{x}]$ of positive degree for $1\leqslant i\leqslant h$.

Theorem C (Schmidt). *Let d be an integer exceeding 1, and write $\chi(d) = d^2 2^{4d} d!$. Let $F(\mathbf{x}) \in \mathbb{Z}[x_1, \dots, x_s]$ be homogeneous of degree d, and suppose that $h(F) \geqslant \chi(d)$. Then one has*

$$\mathit{card}(\{\mathbf{x} \in [-B, B]^s \cap \mathbb{Z}^s \;:\; F(\mathbf{x}) = 0\}) \sim CB^{s-d},$$

where C denotes the "product of local densities" within the box $[-B, B]^s$.

We reiterate that the relative simplicity and strength of Theorem A over Theorems B and C seems to us to justify the investment of further effort in investigations which carry successful elements of the classical methods over to more general situations. We are now at liberty to focus on the topics central to this paper.

Over sixty years ago, Hua [11] greatly simplified the analysis of the asymptotic formula in Waring's problem and allied additive problems with the introduction of a new mean value estimate which, to this day, remains central to the theory of exponential sums of small degree in a single variable. Roughly speaking, Hua observed that by Weyl differencing half of the exponential sums in a suitable mean value, and interpreting the result in terms of the underlying diophantine equation, one obtains a recursive estimate for successive mean values in terms of divisor sum estimates of particularly simple type. The author has recently obtained a version of Hua's lemma for exponential sums of the type

$$\sum_{0 \leqslant x, y \leqslant P} e(\alpha \Phi(x, y)),$$

in which $\Phi(x, y)$ is a non-degenerate binary form with integral coefficients, and as usual, we write $e(z)$ to denote $e^{2\pi i z}$ (see [22]). By means of a carefully orchestrated differencing procedure, we are able to engineer a recursion similar to that of Hua in the situation of a single variable. Unfortunately, however, the divisor sum estimates are complicated by the presence of estimates for the number of integral points on affine plane curves, and our relative ignorance of such matters somewhat weakens the ensuing mean value estimates. In this paper we sharpen our analogue of Hua's lemma by means of an enhanced treatment of the affine curves that arise from the differencing process at the heart of our treatment.

In order to describe our version of Hua's lemma, we require some notation. Suppose that $\Phi(x, y) \in \mathbb{Z}[x, y]$ is a binary form of degree d exceeding 1. Then we say that Φ is *degenerate* if there exist complex numbers α and β such that $\Phi(x, y)$ is identically equal to $(\alpha x + \beta y)^d$. It is easily verified that when $\Phi(x, y)$ is degenerate, then there exist integers a, b and c with $\Phi(x, y) = a(bx + cy)^d$. Finally, define the exponential sum

$$f_\Phi(\alpha; P) = \sum_{0 \leqslant x, y \leqslant P} e(\alpha \Phi(x, y)). \tag{1.1}$$

Theorem 1. *Suppose that $\Phi(x, y) \in \mathbb{Z}[x, y]$ is a non-degenerate form of degree $d \geqslant 3$. Then the following estimates hold.*

(i) When $d=3$ or 4 and j is an integer with $1\leqslant j\leqslant d$, or when $d\geqslant 5$ and $j=1$ or 2, one has for each positive number ε the bound

$$\int_0^1 |f_\Phi(\alpha;P)|^{2^{j-1}}d\alpha \ll P^{2^j-j+\varepsilon}.$$

(ii) When $d=5$, one has for each positive number ε the bounds

$$\int_0^1 |f_\Phi(\alpha;P)|^4 d\alpha \ll P^{21/4+\varepsilon},\quad \int_0^1 |f_\Phi(\alpha;P)|^8 d\alpha \ll P^{49/4+\varepsilon},$$

$$\int_0^1 |f_\Phi(\alpha;P)|^{10} d\alpha \ll P^{127/8+\varepsilon},\quad \int_0^1 |f_\Phi(\alpha;P)|^{17} d\alpha \ll P^{29+\varepsilon}.$$

(iii) When $6\leqslant d\leqslant 10$ and j is an integer with $3\leqslant j\leqslant d-2$, then for each positive number ε one has

$$\int_0^1 |f_\Phi(\alpha;P)|^{2^{j-1}}d\alpha \ll P^{2^j-j+1/(d-j+2)+\varepsilon}.$$

Also, when $6\leqslant d\leqslant 10$, one has for each $\varepsilon>0$ the bounds

$$\int_0^1 |f_\Phi(\alpha;P)|^{\frac{9}{32}2^d}d\alpha \ll P^{\frac{9}{16}2^d-d+1+\varepsilon}$$

and

$$\int_0^1 |f_\Phi(\alpha;P)|^{\frac{17}{32}2^d}d\alpha \ll P^{\frac{17}{16}2^d-d+\varepsilon}.$$

Of course, bounds for moments of $f_\Phi(\alpha;P)$ intermediate between those recorded in the statement of Theorem 1 may be obtained by applying Hölder's inequality to interpolate between those above. For comparison, Theorem 2 of Wooley [22] shows that when $d\geqslant 5$ and j is an integer with $1\leqslant j\leqslant d-1$, one has

$$\int_0^1 |f_\Phi(\alpha;P)|^{2^{j-1}}d\alpha \ll P^{2^j-j+\frac{1}{2}+\varepsilon},$$

and also provides the estimates

$$\int_0^1 |f_\Phi(\alpha;P)|^{\frac{5}{16}2^d}d\alpha \ll P^{\frac{5}{8}2^d-d+1+\varepsilon}\quad\text{and}\quad \int_0^1 |f_\Phi(\alpha;P)|^{\frac{9}{16}2^d}d\alpha \ll P^{\frac{9}{8}2^d-d+\varepsilon}.$$

Case (iii) of Theorem 1 above plainly provides estimates superior to the latter bounds. On the other hand, case (i) of Theorem 1 is simply a restatement of the first estimate of [22, Theorem 2]. We note also that when d is greater than or equal to 11, it is possible to apply a trivial variant of Vinogradov's methods in order to obtain conclusions superior to those stemming from Theorem 1 (see [22, §8] for details). Since we are interested primarily in ideas likely to generalise successfully to homogeneous forms in many variables, we discuss Vinogradov's methods no further herein.

There are immediate consequences of the estimates recorded in Theorem 1 for the solubility of homogeneous diophantine equations which split as sums of binary forms. We confine ourselves here to a routine conclusion discussed in detail in [22].

Theorem 2. *Let d be an integer with $3 \leqslant d \leqslant 10$, and define $s_0(d)$ by*

$$s_0(d) = \begin{cases} 2^{d-1}, & \text{when } d = 3, 4, \\ \frac{17}{32}2^d, & \text{when } 5 \leqslant d \leqslant 10. \end{cases}$$

Let $s > s_0(d)$, and let $\Phi_j \in \mathbb{Z}[x, y]$ $(1 \leqslant j \leqslant s)$ be homogeneous forms of degree d with non-zero discriminants. Let $\mathcal{N}(B) = \mathcal{N}_s(B; \mathbf{\Phi})$ denote the number of solutions of the diophantine equation

$$\Phi_1(x_1, y_1) + \cdots + \Phi_s(x_s, y_s) = 0, \tag{1.2}$$

subject to $|x_j| \leqslant B$ and $|y_j| \leqslant B$ $(1 \leqslant j \leqslant s)$. Then provided that the form $\Phi_1(x_1, y_1) + \cdots + \Phi_s(x_s, y_s)$ is indefinite, one has

$$\mathcal{N}_s(B; \mathbf{\Phi}) = \mathcal{C}\mathfrak{S}B^{2s-d} + O_{\mathbf{\Phi}}(B^{2s-d-\delta}),$$

for some positive number δ. Here, $\mathcal{C}$ denotes the volume of the $(2s-1)$-dimensional hypersurface determined by the equation (1.2) contained in the box $[-1, 1]^{2s}$. Also, $\mathfrak{S}$ denotes the singular series $\prod_p v_p$, where the product is over prime numbers,

$$v_p = \lim_{h \to \infty} p^{h(1-2s)} M_s(p^h; \mathbf{\Phi}),$$

and $M_s(p^h; \mathbf{\Phi})$ denotes the number of solutions of the congruence

$$\Phi_1(x_1, y_1) + \cdots + \Phi_s(x_s, y_s) \equiv 0 \pmod{p^h},$$

with $1 \leqslant x_j, y_j \leqslant p^h$ $(1 \leqslant j \leqslant s)$.

We note that the expression $\mathcal{C}\mathfrak{S}$ explicitly describes the "product of local densities", for the problem at hand, previously mentioned in Theorems A, B and C. Given the existence of non-singular real and p-adic solutions of the equation (1.2), the proof of Theorem 2 follows precisely the argument of the proof of [22, Theorem 3], and hence we omit details in the interest of saving space.

Following some preliminary reductions in §2, we grapple with basic estimates for the number of integral points on affine plane curves in §3. We discuss the main induction in §4, thereby establishing the majority of the estimates recorded in Theorem 1. The closing stages of the induction have a different flavour, and this we defer to §5, completing the proof of Theorem 1.

Throughout this paper, implicit constants occurring in Vinogradov's notation $\ll$ and $\gg$ will depend at most on the coefficients of the implicit binary forms, a small positive number ε, exponents d and k, and quantities occurring as subscripts to the latter notations, unless otherwise indicated. We write $f \asymp g$ when $f \ll g$ and $g \ll f$. When x is a real number, we write $[x]$ for the greatest integer not exceeding x. Also, we use vector notation for brevity. Thus, for example, the s-tuple $(\Phi_1, \ldots, \Phi_s)$ will be abbreviated simply to $\mathbf{\Phi}$. In an effort to simplify our exposition, we adopt the convention that whenever ε appears in a statement, we are implicitly asserting that the statement holds for each $\varepsilon > 0$. Note that the "value" of ε may consequently change from statement to statement.

The author is grateful to the referee for useful comments.

2. Preliminary reductions. Let k be an integer with $k \geqslant 3$ and let $\Phi(x,y) \in \mathbb{Z}[x,y]$ be a non-degenerate homogeneous polynomial of degree k. Let P be a large real number, and define the exponential sum $f(\alpha) = f_\Phi(\alpha;P)$ as in (1.1). We aim initially to transform $f(\alpha)$ into an associated exponential sum amenable to our differencing procedure, and the latter goal we achieve by following closely the argument of [22, §2].

When $\Phi(x,y) \in \mathbb{Z}[x,y]$, we describe the polynomial Ψ as being a *condensation* of Φ when the following condition $(\mathcal{C})$ is satisfied.

$(\mathcal{C})$ We have $\Psi(u,v) \in \mathbb{Z}[u,v]$, and the coefficients of Ψ depend at most on those of Φ. Further, the polynomial $\Psi(u,v)$ has the same degree as $\Phi(x,y)$, and takes the shape

$$\Psi(u,v) = Au^k + Bu^{k-t}v^t + \sum_{j=t+1}^{k} C_j u^{k-j}v^j, \tag{2.1}$$

with $AB \neq 0$ and $2 \leqslant t \leqslant k$.

Lemma 2.1. *There is a condensation Ψ of Φ, and a positive real number X with $X \asymp P$, with the property that for every natural number s one has*

$$\int_0^1 |f_\Phi(\alpha;P)|^{2s} d\alpha \ll \int_0^1 |H_\Psi(\alpha;X)|^{2s} d\alpha,$$

where we write

$$H_\Psi(\theta;X) = \sum_{|u|\leqslant X} \sum_{|v|\leqslant X} e(\theta\Psi(u,v)). \tag{2.2}$$

Proof. This is [22, Lemma 2.3].

The work of [22, §5] takes care of certain special cases that arise in our treatment. We summarise the relevant conclusions of this discussion in the following two lemmata.

Lemma 2.2. *Suppose that $k = 3$ or 4 and j is an integer with $1 \leqslant j \leqslant k$, or else that $k \geqslant 5$ and $j = 1$ or 2. Then for each positive number ε, one has*

$$\int_0^1 |f_\Phi(\alpha;P)|^{2^{j-1}} d\alpha \ll P^{2^j - j + \varepsilon}.$$

Proof. This estimate is recorded as the first conclusion of [22, Theorem 2].

Lemma 2.3. *Suppose that $\Psi(u,v) \in \mathbb{Z}[u,v]$ has the shape (2.1). Suppose also that $k \geqslant 5$, that X is a large real number, and that $H_\Psi(\alpha;X)$ is defined as in (2.2). Then for $1 \leqslant j \leqslant k$, and for each positive number ε, one has the upper bound*

$$\int_0^1 |H_\Psi(\alpha;X)|^{2^{j-1}} d\alpha \ll X^{2^j - j + \varepsilon},$$

provided either that $t = k$, or else that $t = k-1$ and $C_k = 0$. When $t = k-1$ and $C_k \neq 0$, meanwhile, then there is a condensation Υ of Ψ with the property that Υ has the shape

$$\Upsilon(x,y) = A'x^k + B'x^{k-2}y^2 + \sum_{j=3}^{k} C'_j x^{k-j} y^j,$$

with $A'B' \neq 0$, and there is a positive real number Y with $Y \asymp X$, and Υ and Y satisfy the property that for each natural number s, one has

$$\int_0^1 |H_\Psi(\alpha;X)|^{2s} d\alpha \ll \int_0^1 |H_\Upsilon(\alpha;Y)|^{2s} d\alpha.$$

Proof. The situations in which $t = k$, or else $t = k-1$ and $C_k = 0$, are dealt with, respectively, in Lemmata 5.2 and 5.3 of [22]. The alternative situation in which $t = k-1$ and $C_k \neq 0$, on the other hand, is discussed in the preamble to Lemma 5.3 of [22].

Our deliberations are also greatly simplified through a manoeuvre that transforms a polynomial of the shape (2.1) with $t = k-2$ into a corresponding polynomial in which $t = 2$ or 3. We begin with an analogue of Lemma 5.3 of [22]. Suppose, temporarily, that $\Psi(u,v)$ has the shape (2.1) with $t = k-2$, so that for some integers a, b, c, d with $ab \neq 0$, one has

$$\Psi(x,y) = ax^k + bx^2y^{k-2} + cxy^{k-1} + dy^k. \tag{2.3}$$

Lemma 2.4. *Suppose that $k \geqslant 4$, and that $\Psi(u,v) \in \mathbb{Z}[u,v]$ has the shape (2.3) with $ab \neq 0$ and $d = 0$. Define the exponential sum $H_\Psi(\alpha;X)$ as in (2.2). Then for $1 \leqslant j \leqslant k$, and for each positive number ε, one has the upper bound*

$$\int_0^1 |H_\Psi(\alpha;X)|^{2^{j-1}} d\alpha \ll X^{2^j - j + \varepsilon}. \tag{2.4}$$

Proof. Our argument is a variant of the proof of Lemma 5.3 of [22]. We abbreviate $H_\Psi(\alpha;X)$ simply to $H(\alpha)$. Also, when $1 \leqslant j \leqslant k$, we write

$$I_j(X) = \int_0^1 |H(\alpha)|^{2^{j-1}} d\alpha. \tag{2.5}$$

The bound (2.4) is immediate from Lemma 2.2 when $j = 1, 2$. Suppose then that j is an integer with $2 \leqslant j \leqslant k-1$, and that the inequality (2.4) holds. We seek to show that (2.4) holds with j replaced by $j+1$, whence the desired conclusion follows for $1 \leqslant j \leqslant k$ by induction.

Observe first that

$$|H(\alpha)| \ll X + \sum_{1\leqslant |x|\leqslant X} \Bigl| \sum_{|y|\leqslant X} e(\alpha(ax^k + bx^2y^{k-2} + cxy^{k-1}))\Bigr|.$$

Define the exponential sum $h_l(\alpha) = h_l(\alpha; X)$ by

$$h_l(\alpha; X) = \sum_{|y|\leqslant X} e(\alpha(bly^{k-2} + cy^{k-1})).$$

Then it follows from (2.5) via Hölder's inequality that

$$\begin{aligned} I_{j+1}(X) &\ll X^{2^{j-1}} I_j(X) + \int_0^1 \Bigl(|H(\alpha)| \sum_{1\leqslant |x|\leqslant X} |h_x(x\alpha)| \Bigr)^{2^{j-1}} d\alpha \\ &\ll X^{2^{j-1}} I_j(X) + X^{2^{j-1}-1} N(X), \end{aligned} \tag{2.6}$$

where

$$N(X) = \int_0^1 |H(\alpha)|^{2^{j-1}} \sum_{1\leqslant |x|\leqslant X} |h_x(x\alpha)|^{2^{j-1}} d\alpha. \tag{2.7}$$

In the special situation in which $j = k-1$ and $c = 0$, we instead note that by Cauchy's inequality, one has

$$\begin{aligned} &\Bigl| \sum_{|y|\leqslant X} \sum_{1\leqslant |x|\leqslant X} e(\alpha(ax^k + bx^2y^{k-2}))\Bigr|^2 \\ &\qquad \ll X \sum_{|y|\leqslant X} \Bigl| \sum_{1\leqslant |x|\leqslant X} e(\alpha(ax^k + bx^2y^{k-2}))\Bigr|^2 \\ &\qquad \ll X^3 + X \sum_{\substack{1\leqslant |x_1|, |x_2|\leqslant X \\ x_1\neq x_2}} \Bigl| \sum_{|y|\leqslant X} e(\alpha b(x_1^2 - x_2^2)y^{k-2})\Bigr|. \end{aligned}$$

Thus, on applying Hölder's inequality within (2.5), we now obtain

$$\begin{aligned} I_k(X) &\ll X^{3\cdot 2^{k-3}} I_{k-1}(X) \\ &\quad + X^{2^{k-3}} \int_0^1 |H(\alpha)|^{2^{k-2}} \Bigl(\sum_{\substack{1\leqslant |x_1|, |x_2|\leqslant X \\ x_1\neq x_2}} |h_{x_1^2 - x_2^2}(\alpha)| \Bigr)^{2^{k-3}} d\alpha \\ &\ll X^{3\cdot 2^{k-3}} I_{k-1}(X) + X^{3\cdot 2^{k-3}-2} M(X), \end{aligned} \tag{2.8}$$

where

$$M(X)=\int_0^1 |H(\alpha)|^{2^{k-2}} \sum_{\substack{1\leqslant |x_1|,|x_2|\leqslant X\\ x_1\neq x_2}} |h_{x_1^2-x_2^2}(\alpha)|^{2^{k-3}}\,d\alpha. \tag{2.9}$$

By orthogonality, it follows from (2.7) that $N(X)$ is equal to the number of integral solutions of the equation

$$x\sum_{i=1}^{2^{j-2}} \left(bx(y_i^{k-2}-z_i^{k-2})+c(y_i^{k-1}-z_i^{k-1})\right)=\sum_{i=1}^{2^{j-2}} \left(\Psi(u_i,v_i)-\Psi(t_i,w_i)\right), \tag{2.10}$$

with $1\leqslant |x|\leqslant X$, and with each of y_i, z_i, u_i, v_i, t_i, w_i $(1\leqslant i\leqslant 2^{j-2})$ bounded in absolute value by X. Let $N_0(X)$ denote the number of such solutions of (2.10) in which the right hand side of the equation is equal to zero, and let $N_1(X)$ denote the corresponding number of solutions with the latter expression non-zero. Then one has

$$N(X)=N_0(X)+N_1(X). \tag{2.11}$$

We first estimate $N_0(X)$. On considering the underlying diophantine equations and recalling (2.5), we have

$$N_0(X)\ll I_j(X)\sum_{1\leqslant |x|\leqslant X}\int_0^1 |h_x(\alpha)|^{2^{j-1}}\,d\alpha.$$

But a classical version of Hua's lemma (see Lemma 2.5 of Vaughan [19]) shows that for $2\leqslant j\leqslant k-1$, one has

$$\int_0^1 |h_x(\alpha)|^{2^{j-1}}\,d\alpha\ll X^{2^{j-1}-j+1+\varepsilon},$$

uniformly in $x\neq 0$. Thus we deduce that for $2\leqslant j\leqslant k-1$, one has

$$N_0(X)\ll X^{2^{j-1}-j+2+\varepsilon}I_j(X). \tag{2.12}$$

In order to dispose of $N_1(X)$, we introduce some additional notation. For each integer l, we denote by $r_j(n;l)$ the number of representations of the integer n in the form

$$n=l\sum_{i=1}^{2^{j-2}} \left(bl(y_i^{k-2}-z_i^{k-2})+c(y_i^{k-1}-z_i^{k-1})\right),$$

with $|y_i|\leqslant X$ and $|z_i|\leqslant X$ $(1\leqslant i\leqslant 2^{j-2})$. Similarly, for each integer n we write $R_j(n)$ for the number of representations of n in the form

$$n=\sum_{i=1}^{2^{j-2}} \left(\Psi(u_i,v_i)-\Psi(t_i,w_i)\right),$$

with each of u_i, v_i, t_i, w_i $(1\leqslant i\leqslant 2^{j-2})$ bounded in absolute value by X. Then on writing $\gamma=(|b|+|c|)2^j$, we find that

$$N_1(X)\leqslant \sum_{1\leqslant |n|\leqslant \gamma X^k} R_j(n)\sum_{\substack{l|n\\ |l|\leqslant X}} r_j(n;l).$$

On applying an elementary estimate for the divisor function, we therefore deduce from Cauchy's inequality that

$$\begin{aligned}N_1(X)&\leqslant \Bigl(\sum_{n\in\mathbb{Z}}R_j(n)^2\Bigr)^{1/2}\Bigl(\sum_{1\leqslant |n|\leqslant \gamma X^k}\Bigl(\sum_{\substack{l|n\\ |l|\leqslant X}} r_j(n;l)\Bigr)^2\Bigr)^{1/2}\\ &\ll X^\varepsilon\Bigl(\sum_{n\in\mathbb{Z}}R_j(n)^2\Bigr)^{1/2}\Bigl(\sum_{n\in\mathbb{Z}}\sum_{1\leqslant |l|\leqslant X} r_j(n;l)^2\Bigr)^{1/2}. \qquad (2.13)\end{aligned}$$

However, on considering the underlying diophantine equations, it is apparent from (2.13) that

$$N_1(X)\ll X^\varepsilon\,(I_{j+1}(X))^{1/2}\Bigl(\sum_{1\leqslant |l|\leqslant X}\int_0^1 |h_l(\alpha)|^{2^j}\,d\alpha\Bigr)^{1/2}.$$

But the classical version of Hua's lemma (see Lemma 2.5 of [19]) shows that for $1\leqslant j\leqslant k-2$, one has

$$\int_0^1 |h_l(\alpha)|^{2^j}\,d\alpha\ll X^{2^j-j+\varepsilon},$$

uniformly in $l\neq 0$. Moreover, the latter conclusion remains valid for $j=k-1$ whenever c is non-zero. In either circumstance, we deduce that

$$N_1(X)\ll X^\varepsilon\,(I_{j+1}(X))^{1/2}\Bigl(\sum_{1\leqslant |l|\leqslant X} X^{2^j-j+\varepsilon}\Bigr)^{1/2}. \qquad (2.14)$$

On combining (2.6), (2.11), (2.12) and (2.14), we find that for $2\leqslant j\leqslant k-2$, and also when $j=k-1$ and $c\neq 0$, one has

$$I_{j+1}(X)\ll \Bigl(X^{2^{j-1}}+X^{2^j-j+1+\varepsilon}\Bigr)I_j(X)+X^{2^j-(j+1)/2+\varepsilon}\,(I_{j+1}(X))^{1/2},$$

whence our inductive hypothesis (2.4) leads to the upper bound

$$I_{j+1}(X)\ll X^{2^{j+1}-j-1+\varepsilon}+X^{2^j-(j+1)/2+\varepsilon}(I_{j+1}(X))^{1/2}.$$

Thus the estimate (2.4) follows with $j+1$ in place of j in the current circumstances, and so the conclusion of the lemma has been established in all cases but that in which $c=0$ and $j=k-1$.

We now turn to the final elusive case wherein $c=0$ and $j=k-1$. By orthogonality, it follows from (2.9) that $M(X)$ is equal to the number of integral solutions of the equation

$$b(x_1^2-x_2^2)\sum_{i=1}^{2^{k-4}}(y_i^{k-2}-z_i^{k-2})=\sum_{i=1}^{2^{k-3}}(\Psi(u_i,v_i)-\Psi(t_i,w_i)), \tag{2.15}$$

with $1\leqslant |x_1|,|x_2|\leqslant X$ and $x_1\neq x_2$, and with each of y_i, z_i $(1\leqslant i\leqslant 2^{k-4})$, and u_i, v_i, t_i, w_i $(1\leqslant i\leqslant 2^{k-3})$ bounded in absolute value by X. Let $M_0(X)$ denote the number of such solutions of (2.15) in which the right hand side of the equation is equal to zero, and let $M_1(X)$ denote the corresponding number of solutions with the latter expression non-zero. Then plainly one has

$$M(X)=M_0(X)+M_1(X). \tag{2.16}$$

We first estimate $M_0(X)$. On considering the underlying diophantine equations and recalling (2.5), we have

$$M_0(X)\ll I_{k-1}(X)\sum_{1\leqslant l,m\leqslant 2X}\int_0^1 |h_{lm}(\alpha)|^{2^{k-3}}d\alpha.$$

But a classical version of Hua's lemma shows that

$$\int_0^1 |h_{lm}(\alpha)|^{2^{k-3}}d\alpha\ll X^{2^{k-3}-k+3+\varepsilon},$$

uniformly in $lm\neq 0$, whence we obtain

$$M_0(X)\ll X^{2^{k-3}-k+5+\varepsilon}I_{k-1}(X). \tag{2.17}$$

Meanwhile, recycling the notation introduced to treat $N_1(X)$, we see that

$$M_1(X)\leqslant \sum_{1\leqslant |n|\leqslant \gamma X^k}R_{k-1}(n)\sum_{\substack{l|n\\ |l|\leqslant 2X}}\sum_{\substack{m|n\\ |m|\leqslant 2X}}T(n;lm),$$

where we write $T(n;\lambda)$ for the number of representations of the integer n in the form

$$n=b\lambda\sum_{i=1}^{2^{k-4}}(y_i^{k-2}-z_i^{k-2}),$$

with $|y_i|\leqslant X$ and $|z_i|\leqslant X$ $(1\leqslant i\leqslant 2^{k-4})$. Again applying an elementary estimate for the divisor function, we deduce from Cauchy's inequality that

$$\begin{aligned}M_1(X)&\leqslant \Bigl(\sum_{n\in\mathbb{Z}}R_{k-1}(n)^2\Bigr)^{1/2}\Bigl(\sum_{1\leqslant|n|\leqslant\gamma X^k}\Bigl(\sum_{\substack{l|n\\ |l|\leqslant 2X}}\sum_{\substack{m|n\\ |m|\leqslant 2X}}T(n;lm)\Bigr)^2\Bigr)^{1/2}\\ &\ll X^{\varepsilon}\Bigl(\sum_{n\in\mathbb{Z}}R_{k-1}(n)^2\Bigr)^{1/2}\Bigl(\sum_{n\in\mathbb{Z}}\sum_{1\leqslant|l|\leqslant 2X}\sum_{1\leqslant|m|\leqslant 2X}T(n;lm)^2\Bigr)^{1/2}.\end{aligned} \tag{2.18}$$

On considering the underlying diophantine equations, we find from (2.18) that

$$M_1(X) \ll X^\varepsilon (I_k(X))^{1/2} \Big(\sum_{1\leqslant |l|\leqslant 2X} \sum_{1\leqslant |m|\leqslant 2X} \int_0^1 |h_{lm}(\alpha)|^{2^{k-2}} d\alpha \Big)^{1/2}.$$

Again applying the classical version of Hua's lemma, one has

$$\int_0^1 |h_{lm}(\alpha)|^{2^{k-2}} d\alpha \ll X^{2^{k-2}-k+2+\varepsilon},$$

uniformly in $lm \neq 0$, whence

$$M_1(X) \ll X^\varepsilon (I_k(X))^{1/2} \Big(\sum_{1\leqslant |l|\leqslant 2X} \sum_{1\leqslant |m|\leqslant 2X} X^{2^{k-2}-k+2+\varepsilon} \Big)^{1/2}. \tag{2.19}$$

On combining (2.8), (2.16), (2.17) and (2.19), we find that when $c = 0$, one has

$$I_k(X) \ll \left(X^{3\cdot 2^{k-3}} + X^{2^{k-1}-k+3+\varepsilon} \right) I_{k-1}(X) + X^{2^{k-1}-k/2+\varepsilon} (I_k(X))^{1/2},$$

whence our inductive hypothesis (2.4) with $j = k-1$ leads to the upper bound

$$I_k(X) \ll X^{2^k-k+\varepsilon} + X^{2^{k-1}-k/2+\varepsilon} (I_k(X))^{1/2}.$$

We therefore conclude that (2.4) holds with $j = k$ even when $c = 0$, and this completes the proof of the lemma.

Lemma 2.5. *Suppose that $k \geqslant 4$, and that $\Psi(u,v) \in \mathbb{Z}[u,v]$ has the shape (2.3) with $abd \neq 0$. Define the exponential sum $H_\Psi(\alpha; X)$ as in (2.2). Then there is a condensation Υ of Ψ with the property that Υ has the shape*

$$\Upsilon(x,y) = A'x^k + B'x^{k-t}y^t + \sum_{j=t+1}^{k} C'_j x^{k-j} y^j, \tag{2.20}$$

with $A'B' \neq 0$ and $2 \leqslant t \leqslant 3$, and there is a positive real number Y with $Y \asymp X$, and Υ and Y satisfy the property that for each natural number s, one has

$$\int_0^1 |H_\Psi(\alpha; X)|^{2s} d\alpha \ll \int_0^1 |H_\Upsilon(\alpha; Y)|^{2s} d\alpha. \tag{2.21}$$

Proof. By hypothesis, the coefficient d is non-zero, and thus we may make the non-singular change of variable $u = kdy + cx$, $v = x$. Write

$$\Upsilon(u,v) = \Psi(kdv, u - cv), \tag{2.22}$$

so that one has $\Upsilon(u,v)=(kd)^k\Psi(x,y)$. Then it follows from the argument of the proof of Lemma 2.3 of [22] that for some positive real number Y with $Y\asymp X$, and for every natural number s, one has the upper bound (2.21). The proof of the lemma will therefore be completed on establishing that the polynomial $\Upsilon(x,y)$, defined in (2.22), has the shape (2.20). In order to establish the latter conclusion, we apply Taylor's theorem to determine whether or not various coefficients of $\Upsilon(u,v)$ vanish.

Write ∂_z for the differential operator $\partial/\partial z$. Then the coefficient of u^k in $\Upsilon(u,v)$ is equal to

$$\frac{1}{k!}\partial_u^k\Psi(kdv,u-cv)\Big|_{(u,v)=(0,0)}.$$

On writing $\Psi_{i,j}$ for

$$\partial_x^i\partial_y^j\Psi(x,y)\Big|_{(x,y)=(0,0)},$$

one finds by the chain rule, therefore, that the coefficient of u^k in $\Upsilon(u,v)$ is equal to

$$\frac{1}{k!}\Psi_{0,k}=d, \tag{2.23}$$

and this is non-zero by hypothesis. Similarly, the coefficient of $u^{k-1}v$ in $\Upsilon(u,v)$ is equal to

$$\begin{aligned}\frac{1}{(k-1)!}\partial_u^{k-1}\partial_v\Psi(kdv,u-cv)\Big|_{(u,v)=(0,0)}&=\frac{1}{(k-1)!}\left(kd\Psi_{1,k-1}-c\Psi_{0,k}\right)\\&=kdc-ckd=0.\end{aligned} \tag{2.24}$$

Next, the coefficient of $u^{k-2}v^2$ in $\Upsilon(u,v)$ is equal to

$$\begin{aligned}&\frac{1}{2!(k-2)!}\partial_u^{k-2}\partial_v^2\Psi(kdv,u-cv)\Big|_{(u,v)=(0,0)}\\&\qquad=\frac{1}{2!(k-2)!}\left((kd)^2\Psi_{2,k-2}-2kdc\Psi_{1,k-1}+c^2\Psi_{0,k}\right)\\&\qquad=bk^2d^2-\tfrac{1}{2}k(k-1)dc^2.\end{aligned} \tag{2.25}$$

Finally, the coefficient of $u^{k-3}v^3$ in $\Upsilon(u,v)$ is equal to

$$\begin{aligned}&\frac{1}{3!(k-3)!}\partial_u^{k-3}\partial_v^3\Psi(kdv,u-cv)\Big|_{(u,v)=(0,0)}\\&=\frac{1}{3!(k-3)!}\left((kd)^3\Psi_{3,k-3}-3(kd)^2c\Psi_{2,k-2}+3kdc^2\Psi_{1,k-1}-c^3\Psi_{0,k}\right)\\&=-bk^2(k-2)d^2c+\tfrac{1}{3}k(k-1)(k-2)dc^3.\end{aligned} \tag{2.26}$$

When $c=0$, one finds from (2.25) that the coefficient of $u^{k-2}v^2$ is bk^2d^2, and this is non-zero by hypothesis. When $c\neq 0$, on the other hand, it follows from

(2.25) and (2.26) that when the coefficients of both $u^{k-2}v^2$ and $u^{k-3}v^3$ are zero, then necessarily

$$2kbd = (k-1)c^2 \quad \text{and} \quad 3kbd = (k-1)c^2,$$

whence $bd = c = 0$, contrary to hypothesis. We therefore conclude from equations (2.23)–(2.26) that $\Upsilon(x, y)$ does indeed take the shape (2.20), wherein $A'B' \neq 0$ and $t = 2$ or 3. This completes the proof of the lemma.

We next recall the Weyl differencing lemma. Let Δ_j denote the jth iterate of the forward differencing operator, so that for any function Ω of a real variable α, one has

$$\Delta_1(\Omega(\alpha); \beta) = \Omega(\alpha + \beta) - \Omega(\alpha),$$

and when j is a natural number,

$$\Delta_{j+1}(\Omega(\alpha); \beta_1, \dots, \beta_{j+1}) = \Delta_1(\Delta_j(\Omega(\alpha); \beta_1, \dots, \beta_j); \beta_{j+1}).$$

We adopt the convention that $\Delta_0(\Omega(\alpha); \boldsymbol{\beta}) = \Omega(\alpha)$.

Lemma 2.6. *Let X be a positive real number, and let $\Omega(x)$ be an arbitrary arithmetical function. Write*

$$T(\Omega) = \sum_{|x| \leqslant X} e(\Omega(x)).$$

Then for each natural number j there exist intervals $I_i = I_i(\mathbf{h})$ $(1 \leqslant i \leqslant j)$, possibly empty, satisfying

$$I_1(h_1) \subseteq [-X, X] \quad \text{and} \quad I_i(h_1, \dots, h_i) \subseteq I_{i-1}(h_1, \dots, h_{i-1}) \quad (2 \leqslant i \leqslant j),$$

with the property that

$$|T(\Omega)|^{2^j} \leqslant (4X+1)^{2^j - j - 1} \sum_{|h_1| \leqslant 2X} \cdots \sum_{|h_j| \leqslant 2X} T_j,$$

and here we write

$$T_j = \sum_{x \in I_j \cap \mathbb{Z}} e(\Delta_j(\Omega(x); h_1, \dots, h_j)).$$

Proof. This trivial variant of Lemma 2.3 of Vaughan [19] is recorded as Lemma 3.2 of [22].

We must also make use of a two dimensional forward differencing operator $\Delta_{i,j}$ defined as follows. When $\Omega(x, y)$ is a function of the real variables x and y, one defines

$$\Delta_{1,0}(\Omega(x, y); \beta) = \Omega(x + \beta, y) - \Omega(x, y)$$

and

$$\Delta_{0,1}(\Omega(x, y); \gamma) = \Omega(x, y + \gamma) - \Omega(x, y).$$

When i and j are non-negative integers, one then defines

$$\Delta_{i,j}(\Omega(x,y);\beta_1,\ldots,\beta_i;\gamma_1,\ldots,\gamma_j)$$

by taking $\Delta_{0,0}(\Omega(x,y);\boldsymbol{\beta};\boldsymbol{\gamma})=\Omega(x,y)$, and in general by means of the relations

$$\begin{aligned}\Delta_{i+1,j}(\Omega(x,y);\beta_1,\ldots,\beta_{i+1};\gamma_1,\ldots,\gamma_j)\\ =\Delta_{1,0}(\Delta_{i,j}(\Omega(x,y);\beta_1,\ldots,\beta_i;\gamma_1,\ldots,\gamma_j);\beta_{i+1})\end{aligned}$$

and

$$\begin{aligned}\Delta_{i,j+1}(\Omega(x,y);\beta_1,\ldots,\beta_i;\gamma_1,\ldots,\gamma_{j+1})\\ =\Delta_{0,1}(\Delta_{i,j}(\Omega(x,y);\beta_1,\ldots,\beta_i;\gamma_1,\ldots,\gamma_j);\gamma_{j+1}).\end{aligned}$$

3. Integral points on affine plane curves. Essential to the main body of our argument are estimates for the number of integral points on affine plane curves, and in this section we record the estimates required for later use. Our basic tool is the following result of Bombieri and Pila [5].

Lemma 3.1. *Let $\mathcal{C}$ be the curve defined by the equation $F(x,y)=0$, where $F(x,y)\in\mathbb{R}[x,y]$ is an absolutely irreducible polynomial of degree $d\geqslant 2$. Also, let $N\geqslant \exp(d^6)$. Then the number of integral points on $\mathcal{C}$, and inside a square $[0,N]\times[0,N]$, does not exceed*

$$N^{1/d}\exp(12(d\log N\log\log N)^{1/2}).$$

Proof. This is Theorem 5 of Bombieri and Pila [5]. We note that slightly sharper estimates are now available through work of Pila [12], though these new estimates have no impact on the present work.

At the request of the referee, we point out that applications of this result of Bombieri and Pila (of a rather different flavour, involving slicing arguments) may be found in [1], [14], [15] and [16]. An application more akin to that at hand may be examined in §3 of [21] (the argument therein was in fact inspired by the proof of Lemma 3.2 below). We avoid detailed discussion of the absolute irreducibility condition occurring in the above lemma by careful averaging. Here the initial stages of our argument are modelled closely on the method of the proof of [22, Lemma 4.2].

Lemma 3.2. *Let X denote a large real number. Suppose that $F(x,y)\in\mathbb{Z}[x,y]$ is a non-degenerate polynomial of degree $d\geqslant 2$, and that X is sufficiently large in terms of d. Suppose also that for some fixed positive number A, one has that the coefficients of F are each bounded in absolute value by X^A. Given a polynomial $T(x,y)\in\mathbb{R}[x,y]$, denote by $r_T(n;X)$ the number of solutions of the diophantine equation $T(x,y)=n$, with $(x,y)\in[-X,X]^2\cap\mathbb{Z}^2$. Then one of the following two situations must occur, and in each of the bounds which follows, implicit constants* *depend at most on d, ε and A, and otherwise are independent of the coefficients of F.*

(i) *There exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the following conditions.*

(a) *G is non-degenerate of degree exceeding 1;*

(b) *g has degree exceeding 1;*

(c) *the equation $F(x,y)=g(G(x,y))$ is satisfied identically;*

(d) *one has*

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2 \ll X^{2+1/d+\varepsilon} + X^{\varepsilon}\sum_{n\in\mathbb{Z}} r_G(n;X)^2.$$

(ii) *No polynomials G, g exist satisfying the conditions (a), (b), (c), (d) above. Then one has*

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2 \ll X^{2+1/d+\varepsilon}.$$

Proof. Consider an integer $n\in\mathbb{N}$ with $r_F(n;X)\neq 0$. In view of the hypotheses of the statement of the lemma, we may suppose that for some fixed positive number B, one has $|n|\leqslant X^B$. When i is a non-negative integer, write

$$\mathcal{Z}_i=\{n\in\mathbb{Z}\ :\ |n|\leqslant 2^iX^B\}.$$

Also, let $\mathcal{N}_1$ denote the set of integers $n\in\mathcal{Z}_0$ for which the polynomial $F(x,y)-n$ is absolutely irreducible. Then an application of Lemma 3.1 reveals that for each $n\in\mathcal{N}_1$, one has $r_F(n;X)=O(X^{1/d+\varepsilon})$, whence

$$\sum_{n\in\mathcal{N}_1} r_F(n;X)^2 \ll X^{1/d+\varepsilon}\sum_{n\in\mathcal{Z}_0} r_F(n;X) \ll X^{2+1/d+\varepsilon}. \tag{3.1}$$

Suppose next that $n\notin\mathcal{N}_1$, so that the polynomial $F(x,y)-n$ factors as a product of absolutely irreducible factors, say

$$F(x,y)-n=\prod_{j=1}^{l} g_j(x,y)\prod_{k=1}^{m} h_k(x,y),$$

where $l+m\geqslant 2$, and where $g_j(x,y)\in\mathbb{R}[x,y]$ $(1\leqslant j\leqslant l)$, and

$$h_k(x,y)=u_k(x,y)+v_k(x,y)\sqrt{-1}\quad (1\leqslant k\leqslant m),$$

with $u_k,v_k\in\mathbb{R}[x,y]$ for each k. Since $h_k(x,y)$ is presumed to be absolutely irreducible, we may suppose that $u_k(x,y)$ and $v_k(x,y)$ have no non-trivial polynomial common divisor over $\mathbb{C}[x,y]$. It therefore follows from Bezout's theorem that the number of solutions of the simultaneous equations $u_k(x,y)=v_k(x,y)=0$ is bounded above by d^2. By considering real and imaginary components, therefore, the number of integral solutions of the equation $h_k(x,y)=0$ is also bounded above by d^2. Next, if $g_j(x,y)$ is not some constant multiple of a $\mathbb{Q}$-rational polynomial, then since $g_j(x,y)$ is necessarily a constant multiple of a polynomial with algebraic coefficients, we deduce that the number of integral solutions of the equation $g_j(x,y)=0$ is at most d^2. For we may remove the aforementioned constant factor and consider components with respect to some basis for the field extension containing the coefficients of $g_j(x,y)$. Then since $g_j(x,y)$ is not a constant multiple of a $\mathbb{Q}$-rational polynomial, we find that the integral zeros of $g_j(x,y)=0$ necessarily satisfy at least two linearly independent $\mathbb{Q}$-rational polynomial equations of degree at most d, whence the desired conclusion follows as in the complex case. Let $\mathcal{N}_2$ denote the set of integers $n\in\mathcal{Z}_0\setminus\mathcal{N}_1$ for which the polynomial $F(x,y)-n$ possesses no non-trivial, absolutely irreducible $\mathbb{Q}$-rational polynomial factor. Then the above argument shows that for each $n\in\mathcal{N}_2$, one has $r_F(n;X)=O(1)$, whence

$$\sum_{n\in\mathcal{N}_2} r_F(n;X)^2 \ll \sum_{n\in\mathcal{Z}_0} r_F(n;X) \ll X^2. \tag{3.2}$$

Suppose next that the set $\mathcal{Z}_0\setminus(\mathcal{N}_1\cup\mathcal{N}_2)$ is non-empty, so that there exists some integer $n_0\in\mathcal{Z}_0$ with the property that $F(x,y)-n_0$ possesses a non-trivial, absolutely irreducible $\mathbb{Q}$-rational polynomial factor. Since $F(x,y)$ has integer coefficients, it follows that $F(x,y)-n_0$ may be written as a product

$$F(x,y)-n_0=\psi_1(x,y)\dots\psi_m(x,y), \tag{3.3}$$

with each $\psi_i(x,y)\in\mathbb{Z}[x,y]$ irreducible of degree d_i, say. Moreover, we may suppose without loss of generality that $m\geqslant 2$ and that $d_1+\dots+d_m=d$. Furthermore, on writing

$$R(\mathbf{u};\boldsymbol{\phi};X)=R(u_1,\dots,u_m;\phi_1,\dots,\phi_m;X)$$

for the number of integer solutions of the system of equations

$$\phi_i(x,y)=u_i \quad (1\leqslant i\leqslant m), \tag{3.4}$$

with $|x|,|y|\leqslant X$, it follows from (3.3) that when $\mathcal{Z}_0\setminus(\mathcal{N}_1\cup\mathcal{N}_2)$ is non-empty, one has

$$\sum_{n\in\mathcal{Z}_0\setminus\{n_0\}} r_F(n;X)^2 \leqslant \sum_{n\in\mathcal{Z}_1\setminus\{0\}} \Biggl(\sum_{u_1\dots u_m=n} R(\mathbf{u};\boldsymbol{\psi};X)\Biggr)^2.$$

Notice here that on the right hand side of the last inequality, we are implicitly applying a shift by $-n_0$ to $\mathcal{Z}_0$, and then we note that this shifted set is contained in $\mathcal{Z}_1$. Thus, on combining an application of Cauchy's inequality with an elementary estimate for the divisor function, we obtain

$$\begin{aligned}\sum_{n\in\mathcal{Z}_0} r_F(n;X)^2 &\ll r_F(n_0;X)^2+X^\varepsilon \sum_{n\in\mathcal{Z}_1\setminus\{0\}}\ \sum_{u_1\dots u_m=n} R(\mathbf{u};\boldsymbol{\psi};X)^2\\ &\ll X^2+X^\varepsilon\sum_{\mathbf{u}\in\mathcal{Z}_1^m} R(\mathbf{u};\boldsymbol{\psi};X)^2. \end{aligned}\tag{3.5}$$

Suppose now that $m \geqslant 2$, and that for $1 \leqslant i \leqslant m$ the polynomials $\phi_i(x,y) \in \mathbb{Z}[x,y]$ have degree $d_i \geqslant 1$. Suppose also that these polynomials satisfy the condition that $d_1 + \cdots + d_m \leqslant d$, that $F(x,y)$ is a polynomial in $\phi_1, \ldots, \phi_m$, and that for some j with $1 \leqslant j < d$, one has the upper bound

$$\sum_{n \in \mathcal{Z}_0} r_F(n;X)^2 \ll X^{2+\varepsilon} + X^{\varepsilon} \sum_{\mathbf{u} \in \mathcal{Z}_j^m} R(\mathbf{u};\boldsymbol{\phi};X)^2. \tag{3.6}$$

Note that by (3.3) and (3.5), this condition is already met when $\boldsymbol{\phi} = \boldsymbol{\psi}$, wherein we take $j = 1$. It is possible that the intersection (3.4) is proper for every available choice of $\mathbf{u}$, by which we mean that the intersection over $\mathbb{C}$ consists of isolated points only, and in such circumstances an application of Bezout's theorem leads to the bound $R(\mathbf{u};\boldsymbol{\phi};X) = O(1)$ uniformly in $\mathbf{u}$, whence

$$\sum_{\mathbf{u} \in \mathcal{Z}_j^m} R(\mathbf{u};\boldsymbol{\phi};X)^2 \ll \sum_{\mathbf{u} \in \mathcal{Z}_j^m} R(\mathbf{u};\boldsymbol{\phi};X) \ll X^2.$$

If, on the other hand, there exists a choice of $\mathbf{u}$ in the summation for which the intersection defined by (3.4) is improper, say $\mathbf{u} = \mathbf{u}^*$, then the polynomials $\phi_i - u_i^*$ $(1 \leqslant i \leqslant m)$ must possess a non-trivial common factor $\chi_{m+1} \in \mathbb{Z}[x,y]$. Denote by $\chi_1, \ldots, \chi_m \in \mathbb{Z}[x,y]$ the quotient polynomials satisfying the equations

$$\phi_i(x,y) - u_i^* = \chi_{m+1}(x,y)\chi_i(x,y) \quad (1 \leqslant i \leqslant m). \tag{3.7}$$

Then it is apparent that

$$\begin{aligned}\sum_{\mathbf{u} \in \mathcal{Z}_j^m} R(\mathbf{u};\boldsymbol{\phi};X)^2 \ll &\sum_{i=1}^m R(u_i^*;\phi_i;X)^2 \\ &+ \sum_{\mathbf{u} \in (\mathcal{Z}_{j+1}\setminus\{0\})^m} \Biggl(\sum_{\substack{\mathbf{v} \in \mathcal{Z}_{j+1}^{m+1} \\ v_i v_{m+1} = u_i \ (1 \leqslant i \leqslant m)}} R(\mathbf{v};\boldsymbol{\chi};X) \Biggr)^2,\end{aligned}$$

whence by combining Cauchy's inequality with an elementary divisor function estimate, one obtains

$$\begin{aligned}\sum_{\mathbf{u} \in \mathcal{Z}_j^m} R(\mathbf{u};\boldsymbol{\phi};X)^2 &\ll X^2 + X^{\varepsilon} \sum_{\mathbf{u} \in (\mathcal{Z}_{j+1}\setminus\{0\})^m} \sum_{\substack{\mathbf{v} \in (\mathcal{Z}_{j+1}\setminus\{0\})^{m+1} \\ v_i v_{m+1} = u_i \ (1 \leqslant i \leqslant m)}} R(\mathbf{v};\boldsymbol{\chi};X)^2 \\ &\leqslant X^2 + X^{\varepsilon} \sum_{\mathbf{v} \in \mathcal{Z}_{j+1}^{m+1}} R(\mathbf{v};\boldsymbol{\chi};X)^2. \end{aligned} \tag{3.8}$$

Let $\{\chi_{i_1}, \ldots, \chi_{i_l}\}$ denote the subset of $\{\chi_1, \ldots, \chi_{m+1}\}$ in which constant polynomials are omitted. Then it is apparent from (3.7) and our initial hypothesis that

$F(x,y)$ is a polynomial in $\chi_{i_1},\dots,\chi_{i_l}$. If the degrees of the latter polynomials are respectively $e_1,\dots,e_l$, then it is clear from (3.7) also that

$$e_1+\dots+e_l<d_1+\dots+d_m\leqslant d.$$

Also, on combining the hypothesis (3.6) with (3.8), one deduces that

$$\sum_{n\in\mathcal{Z}_0} r_F(n;X)^2\ll X^{2+\varepsilon}+X^\varepsilon\sum_{\mathbf{w}\in\mathcal{Z}_{j+1}^l} R(\mathbf{w};\chi_{i_1},\dots,\chi_{i_l};X)^2. \tag{3.9}$$

In view of the above discussion, therefore, we infer from the hypotheses concluding with (3.6) either that

$$\sum_{n\in\mathcal{Z}_0} r_F(n;X)^2\ll X^{2+\varepsilon}, \tag{3.10}$$

or that (3.9) holds with $l=1$, or else that these initial hypotheses again hold, but with j replaced by $j+1$, and with the m-tuple $\boldsymbol{\phi}$ replaced by an m'-tuple of polynomials with strictly smaller degree in the sense that their sum of degrees is strictly smaller. Since the sum of the degrees of the ϕ_i must always be at least 1, we conclude that repeated application of this reduction must terminate after at most d steps either with the conclusion (3.10), or else with the conclusion that (3.9) holds with $l=1$ and $j=d$. In the former case we deduce that

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2\ll X^{2+\varepsilon}. \tag{3.11}$$

In the latter situation, meanwhile, we may conclude that polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ exist satisfying the conditions (b), (c) of the statement of Lemma 3.2. If $G(x,y)$ has degree 1, or else is degenerate of degree exceeding 1, moreover, then it follows from conditions (b) and (c) that $F(x,y)$ is itself degenerate, contrary to our earlier hypotheses. Thus condition (a) is also satisfied. Furthermore, our above discussion also yields the bound

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2\ll X^{2+\varepsilon}+X^\varepsilon\sum_{n\in\mathbb{Z}} r_G(n;X)^2. \tag{3.12}$$

On combining the estimates (3.1), (3.2), (3.11) and (3.12), we find that the conclusion of the lemma follows in all cases.

In the later stages of our argument we are reduced to equations quadratic with respect to a subset of the variables. These we handle with the aid of the following elementary estimate.

Lemma 3.3. *Let a,b,c be integers with $abc\neq 0$, and let $S(a,b,c;P)$ denote the number of integral solutions of the equation $ax^2+by^2=c$, with $|x|\leqslant P$ and $|y|\leqslant P$. Then for each positive number ε, one has $S(a,b,c;P)\ll 1+(|abc|P)^\varepsilon$.*

Proof. This well-known estimate can be found in Estermann [8] or Vaughan and Wooley [20, Lemma 3.5].

We now provide the refinement of Lemma 3.2 of such utility in quadratic cases, basing our argument on that occurring in the proof of Lemma 7.1 of [22].

Lemma 3.4. *Let X denote a large real number. Suppose that $F(x,y) \in \mathbb{Z}[x,y]$ is a non-degenerate polynomial of degree $d \geqslant 2$, and suppose also that $F(x,y)$ has degree precisely 2 in terms of x. Suppose in addition that for no rational numbers λ and μ is it true that there exists a polynomial $f(x,y) \in \mathbb{Z}[x,y]$ for which the equation*

$$F(x,y) = \lambda f(x,y)^2 + \mu$$

is satisfied identically. Further, suppose that for some fixed positive number A, the coefficients of F are each bounded in absolute value by X^A. Then in the notation defined in the statement of Lemma 3.2*, one has*

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2 \ll X^{2+\varepsilon}.$$

Proof. We may rewrite the polynomial $F(x,y)$ in the form

$$F(x,y) = \alpha(y)x^2 + \beta(y)x + \gamma(y), \tag{3.13}$$

where $\alpha(y)$ is a polynomial in y with integral coefficients which is not identically zero, though possibly constant, and $\beta(y), \gamma(y) \in \mathbb{Z}[y]$. Let $R_1(X)$ denote the number of solutions of the equation

$$F(x_1,y_1) = F(x_2,y_2), \tag{3.14}$$

with $|x_i| \leqslant X$, $|y_i| \leqslant X$ $(i = 1,2)$, in which $\alpha(y_i) = 0$ for $i = 1$ or 2. Define the polynomial $\Delta(y)$ by

$$\Delta(y) = \beta(y)^2 - 4\alpha(y)\gamma(y), \tag{3.15}$$

and let $R_2(X)$ denote the corresponding number of solutions of (3.14) in which $\alpha(y_i) \neq 0$ $(i = 1,2)$, and one has that $\Delta(y)$ is identically zero as a polynomial in y. Let $R_3(X)$ denote the corresponding number of solutions in which $\alpha(y_i) \neq 0$ $(i = 1,2)$, and $\Delta(y)$ is not identically zero as a polynomial in y, and moreover one has

$$\alpha(y_2)\Delta(y_1) = \alpha(y_1)\Delta(y_2). \tag{3.16}$$

Finally, let $R_4(X)$ denote the corresponding number of solutions with $\alpha(y_i) \neq 0$ $(i = 1,2)$, and for which the equation (3.16) does not hold. Then plainly,

$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2 \leqslant \sum_{i=1}^{4} R_i(X). \tag{3.17}$$

We first bound $R_1(X)$. Suppose that $\alpha(y_i) = 0$ for $i = 1,2$. Since $\alpha(y)$ is not identically zero, it follows that there are at most d^2 permissible choices for $\mathbf{y}$. Since there are trivially $O(X^2)$ possible choices for $\mathbf{x}$, we find that the contribution to $R_1(X)$ from this first class of solutions is $O(X^2)$. Consider next the remaining solutions for which $\alpha(y_i) = 0$ for at most one value of i. By relabelling variables, we may suppose that $\alpha(y_1)=0$. There are consequently at most d choices permissible for y_1. Fix any one such choice, and also fix any one of the $O(X)$ available choices for x_1. Since $\alpha(y_2)$ is non-zero, it follows from (3.13) and (3.14) that the latter equation is explicit in both x_2 and y_2, whence a simple counting argument reveals that the number of possible choices for x_2 and y_2 satisfying (3.14) is at most $O(X)$. There are thus $O(X^2)$ solutions of this second type, whence

$$R_1(X) \ll X^2. \tag{3.18}$$

Consider next the solutions counted by $R_2(X)$. There exist non-trivial polynomials $\alpha_1(y), \alpha_2(y) \in \mathbb{Z}[y]$ with the property that $\alpha(y)=\alpha_1(y)\alpha_2(y)^2$, and $\alpha_1(y)$ has no repeated factors over $\mathbb{C}[y]$. Since $\alpha(y)$ is a non-trivial polynomial in y, it follows from (3.15) that if $\Delta(y)$ is identically zero as a polynomial in y, then $\beta(y)$ is divisible by the polynomial $\alpha_1(y)\alpha_2(y)$. Such is immediate when $\gamma(y)$ is non-zero, and when $\gamma(y)$ is equal to zero one has $\beta(y)=0$, and the desired conclusion again follows. But if $\beta(y)$ is divisible by $\alpha_1(y)\alpha_2(y)$, then the vanishing of $\Delta(y)$ ensures, by (3.15), that $\gamma(y)$ is divisible by $\alpha_1(y)$. We therefore deduce that for some non-zero integers κ_1, κ_2, and some polynomial in y with integral coefficients, say $\delta(y)$, one has

$$\kappa_1 F(x,y) = \alpha_1(y)(\kappa_2\alpha_2(y)x+\delta(y))^2 \tag{3.19}$$

identically as a polynomial in x and y. We observe here that since $\alpha_1(y)$ and $\alpha_2(y)$ are divisors of $\alpha(y)$, it follows that their coefficients have absolute values at most $O(X^A)$ (see, for example, Granville [9]). One finds in like manner that the coefficients of $\delta(y)$, and also κ_1 and κ_2, may be chosen with absolute values at most $O(X^{2A})$. Notice also that our hypothesis that F is not a rational multiple of the square of a polynomial ensures that $\alpha_1(y)$ is not a constant polynomial. Let x_2 and y_2 be any one of the $O(X^2)$ permissible choices counted by $R_2(X)$. Since, by an elementary counting argument, the number of solutions of the equation $F(x,y)=0$ with $|x|\leqslant X$ and $|y|\leqslant X$ is $O(X)$, the total number of solutions $\mathbf{x},\mathbf{y}$ counted by $R_2(X)$ with $F(x_2,y_2)=0$ is $O(X^2)$. We may therefore suppose that our aforementioned choice of x_2, y_2 satisfies the condition that $F(x_2,y_2)\neq 0$, whence $\kappa_1 F(x_2,y_2)\neq 0$. But it follows from (3.14) and (3.19) that $\alpha_1(y_1)$ and $\kappa_2\alpha_2(y_1)x_1+\delta(y_1)$ are both divisors of the fixed non-zero integer $\kappa_2 F(x_2,y_2)$. By elementary estimates for the divisor function, therefore, there are at most $O(X^\varepsilon)$ possible choices for integers d_1 and d_2 with $\alpha_1(y_1)=d_1$ and $\kappa_2\alpha_2(y_1)x_1+\delta(y_1)=d_2$. Since $\alpha_1(y)$ is not a constant polynomial, the first of the latter equations shows that there are at most d possible choices for y_1. Given any one fixed such choice of y_1, on noting that the non-vanishing of $\alpha(y_1)$ ensures also that $\alpha_2(y_1)\neq 0$, one finds that x_1 is uniquely determined from the second of these equations. Thus we deduce that

$$R_2(X) \ll X^{2+\varepsilon}. \tag{3.20}$$

Consider next the solutions $\mathbf{x},\mathbf{y}$ counted by $R_3(X)$. If, on the one hand, the polynomial equation (3.16) is non-trivial in y_1 and y_2, then a simple counting argument shows that there are $O(X)$ permissible choices for y_1 and y_2 satisfying

(3.16). Given any one such choice of $\mathbf{y}$, in view of the presumed non-vanishing of $\alpha(y_i)$ $(i=1,2)$, it follows from (3.13) that the equation (3.14) is non-trivial in x_1 and x_2, whence there are $O(X)$ permissible choices of x_1 and x_2 satisfying (3.14). Thus the total number of solutions of this type is $O(X^2)$. If, on the other hand, the polynomial equation (3.16) is trivial in y_1 and y_2, then it follows that $\Delta(y)$ is a non-zero constant multiple of $\alpha(y)$, say $\Delta(y)=\lambda\alpha(y)$. We may again write $\alpha(y)=\alpha_1(y)\alpha_2(y)^2$, with α_1 and α_2 defined as in the treatment of $R_2(X)$. An inspection of (3.15) now reveals that

$$\lambda\alpha_1(y)\alpha_2(y)^2=\beta(y)^2-4\alpha_1(y)\alpha_2(y)^2\gamma(y),$$

whence $\beta(y)$ is a multiple of $\alpha_1(y)\alpha_2(y)$. Write $\beta(y)=\mu^{-1}\beta_1(y)\alpha_1(y)\alpha_2(y)$, where μ is a non-zero integer and $\beta_1(y)\in\mathbb{Z}[y]$. Note here that, as in the above discussion, one may suppose that the coefficients of $\beta_1(y)$, $\alpha_1(y)$ and $\alpha_2(y)$, together with the integer μ, have absolute values at most $O(X^{2A})$. We thus infer that

$$4\mu^2\gamma(y)=\alpha_1(y)\beta_1(y)^2-\lambda\mu^2.$$

On substituting into (3.13), we find that

$$4\mu^2F(x,y)=\alpha_1(y)(2\mu\alpha_2(y)x+\beta_1(y))^2-\lambda\mu^2.$$

In particular, our hypothesis that $F(x,y)$ is not a translation of a rational multiple of a square of a polynomial ensures that $\alpha_1(y)$ is not a constant polynomial. In this way, it follows that the equation (3.14) takes the shape

$$\alpha_1(y_1)(2\mu\alpha_2(y_1)x_1+\beta_1(y_1))^2=\alpha_1(y_2)(2\mu\alpha_2(y_2)x_2+\beta_1(y_2))^2.$$

A comparison between the polynomial $\alpha_1(y)(2\mu\alpha_2(y)x+\beta_1(y))^2$ and that on the right hand side of (3.19) reveals that we may now apply the argument concluding the treatment of $R_2(X)$ above in order to conclude that the number of solutions of this type is $O(X^{2+\varepsilon})$. Thus we have

$$R_3(X)\ll X^{2+\varepsilon}. \tag{3.21}$$

Finally, we discuss the solutions counted by $R_4(X)$. Let $\mathbf{x},\mathbf{y}$ be any solution of (3.14) of the latter type. Then on recalling (3.13), (3.14) and (3.15), we deduce that

$$\begin{aligned}\alpha(y_2)(2\alpha(y_1)x_1+\beta(y_1))^2-\alpha(y_1)(2\alpha(y_2)x_2+\beta(y_2))^2\\ =\alpha(y_2)\Delta(y_1)-\alpha(y_1)\Delta(y_2).\end{aligned} \tag{3.22}$$

But in view of our hypotheses relevant to $R_4(X)$, for each of the $O(X^2)$ permissible values of $\mathbf{y}$, one has that the right hand side of (3.22) is a non-zero integer, say

N. Fix any one such choice of $\mathbf{y}$, and note that our hypotheses ensure also that $\alpha(y_i) \neq 0$ $(i = 1, 2)$. But by Lemma 3.3, the number of solutions of the equation

$$\alpha(y_2)\xi^2 - \alpha(y_1)\eta^2 = N,$$

with ξ and η each bounded in absolute value by a fixed power of X, is $O(X^\varepsilon)$. Consequently, the number of possible x_i $(i = 1, 2)$ is also $O(X^\varepsilon)$, and thus we conclude that

$$R_4(X) \ll X^{2+\varepsilon}. \tag{3.23}$$

The conclusion of the lemma now follows immediately on collecting together the estimates (3.18), (3.20), (3.21), (3.23) with (3.17).

We note that the treatment of $\mathcal{K}_3(X; \mathbf{h})$ in the proof of Lemma 7.1 of [22] contains an oversight in that $\beta(y; \mathbf{h})$ was presumed to be necessarily zero, as a consequence of the argument presented therein. The treatment of $R_3(X)$ above takes care of this oversight, and the diligent reader will find that there are no substantive difficulties encountered here. Indeed, one may assume in the above treatment that $\alpha_2(y)$ is identically equal to 1 when applying this argument in the context of the treatment of $\mathcal{K}_3(X; \mathbf{h})$ in the aforementioned work.

Before proceeding to the main inductive part of our argument, we require still another estimate of simpler type than those embodied in Lemmata 3.2 and 3.4.

Lemma 3.5. *Let X denote a large real number. Suppose that $F(x, y) \in \mathbb{Z}[x, y]$ is a non-degenerate polynomial of total degree 2, and suppose also that $F(x, y)$ has degree precisely 1 in terms of x. Further, suppose that for some fixed positive number A, the coefficients of F are each bounded in absolute value by X^A. Then in the notation defined in the statement of Lemma 3.2, one has*

$$\sum_{n \in \mathbb{Z}} r_F(n; X)^2 \ll X^{2+\varepsilon}.$$

Proof. We may rewrite the polynomial $F(x, y)$ in the shape

$$F(x, y) = \alpha(y)x + \beta(y),$$

where $\alpha(y)$ is a linear polynomial in y with integral coefficients that is not identically zero, and $\beta(y)$ is a quadratic polynomial in y with integral coefficients. By considering the putative coefficient of x^2, it is apparent that $F(x, y)$ cannot be the translate of a rational multiple of the square of a polynomial. Consequently, when $\beta(y)$ has non-vanishing leading coefficient we may reverse the roles of x and y, and appeal to Lemma 3.4 in order to establish the conclusion of the lemma. When the leading coefficient of $\beta(y)$ is zero, on the other hand, it follows that for some integers a, b, c and d, with $a \neq 0$, one has

$$F(x, y) = axy + bx + cy + d,$$

whence
$$aF(x,y)=(ax+c)(ay+b)+ad-bc.$$
But then one has
$$\sum_{n\in\mathbb{Z}} r_F(n;X)^2=\sum_{n\in\mathbb{Z}} r_G(n;X)^2, \tag{3.24}$$
where $G(x,y)=(ax+c)(ay+b)$.

When n is non-zero and $G(x,y)=n$, an elementary divisor function estimate shows that there are $O(X^\varepsilon)$ possible choices for $ax+c$ and $ay+b$, whence also for x and y. When n is zero, on the other hand, one has that $ax+c=0$ or else that $ay+b=0$, so that the corresponding number of solutions is $O(X)$. Consequently,
$$\sum_{n\in\mathbb{Z}} r_G(n;X)^2 \ll r_G(0;X)^2+X^\varepsilon \sum_{n\in\mathbb{Z}\setminus\{0\}} r_G(n;X)\ll X^{2+\varepsilon},$$
and the desired conclusion is again immediate, in view of (3.24).

4. The inductive step. We are now equipped to discuss the main inductive step in the proof of Theorem 1. Consider a non-degenerate binary form $\Psi(x,y)$ of the shape (2.1), and define the exponential sum $H_\Psi(\theta;X)$ as in (2.2). When X is a large real number and s is a positive number, define
$$\mathcal{I}_s(X)=\int_0^1 |H_\Psi(\alpha;X)|^s d\alpha.$$

Lemma 4.1. *Let $\Psi(x,y)\in\mathbb{Z}[x,y]$ be a non-degenerate form of degree k, with $3\leqslant k\leqslant 10$, of the shape discussed above. Then one of the following statements is true.*

(i) For $1\leqslant j\leqslant k$, and for each positive number ε, one has
$$\mathcal{I}_{2^j-1}(X)\ll X^{2^j-j+\varepsilon}.$$

(ii) For each positive number s, and for each integer j with $1\leqslant j\leqslant k-3$, one has for each $\varepsilon>0$ the upper bound
$$\mathcal{I}_{s+2^j}(X)\ll X^{2^{j+1}-1}\mathcal{I}_s(X)+X^{2^{j+1}-\frac{1}{2}(j+2-\delta)+\varepsilon}(\mathcal{I}_{2s}(X))^{1/2},$$
where $\delta=\delta(j)$ is defined by
$$\delta(j)=\begin{cases} 1/(k-j), & \text{when } 1\leqslant j<k-3,\\ 0, & \text{when } j=k-3.\end{cases}$$

Proof. We begin by noting that the conclusion (i) of the lemma is immediate from Lemma 2.2 when $k=3$ or 4, and also when $k\geqslant 5$ and $j=1$ or 2. When $k\geqslant 5$ and $t\geqslant k-1$, moreover, the conclusion of Lemma 2.3 demonstrates either that conclusion (i) holds, or else that conclusion (ii) will follow provided we establish the validity of the latter when Ψ is replaced by a condensation Υ of Ψ of the shape (2.1) wherein $t=2$. There is therefore no loss of generality in supposing that $2\leqslant t\leqslant k-2$. Similarly, when $k\geqslant 5$ and $t=k-2$, the conclusions of Lemmata 2.4 and 2.5 ensure either that conclusion (i) holds, or else that conclusion (ii) will follow provided we establish the validity of the latter when Ψ is replaced by a condensation Υ of Ψ of the shape (2.1) wherein $t=2$ or $t=3$. We therefore deduce that the conclusion of the lemma follows by establishing the inequality recorded in (ii) for those polynomials Ψ for which either $k=5$ and $t=2$ or 3, or else $6\leqslant k\leqslant 10$ and $2\leqslant t\leqslant k-3$. We suppose henceforth that the latter conditions do indeed hold.

We now modify the argument applied in §§6 and 7 of [22], applying a more elaborate differencing procedure, and considering also moments other than the even ones. Let w be a parameter to be chosen later satisfying the inequalities

$$\max\{1,j-k+t+2\}\leqslant w\leqslant \min\{j,t-1\}. \tag{4.1}$$

The significance of these inequalities will become clear in due course. For the moment we remark only that our hypotheses concerning j, t and k ensure that an integral value of w can always be found satisfying (4.1).

We first view the exponential sum $H(\theta)=H_\Psi(\theta;X)$ as an exponential sum over v, so that on applying Hölder's inequality to (2.2), and then making use of Lemma 2.6, we deduce that

$$\begin{aligned}|H(\theta)|^{2^w}&\ll X^{2^w-1}\sum_{|u|\leqslant X}\Bigl|\sum_{|v|\leqslant X}e(\theta\Psi(u,v))\Bigr|^{2^w}\\&\ll X^{2^{w+1}-w-2}\sum_{\mathbf{h}\in[-2X,2X]^w}\ \sum_{v\in I(\mathbf{h})}|K(\theta;\mathbf{h};v)|,\end{aligned} \tag{4.2}$$

where $I=I(h_1,\dots,h_w)$ is an interval of integers contained in $[-X,X]$, and

$$K(\theta;\mathbf{h};v)=\sum_{|u|\leqslant X}e(\Delta_{0,w}(\theta\Psi(u,v);\mathbf{h})).$$

Next applying Lemma 2.6 to the latter exponential sum, we obtain

$$|K(\theta;\mathbf{h};v)|^{2^{j-w}}\ll X^{2^{j-w}-j+w-1}\sum_{\mathbf{g}\in[-2X,2X]^{j-w}}\ \sum_{u\in J(\mathbf{g})}e(\theta p(v;u;\mathbf{g};\mathbf{h})), \tag{4.3}$$

where $J=J(g_1,\dots,g_{j-w})$ is an interval of integers contained in $[-X,X]$, and the polynomial $p(v;u;\mathbf{g};\mathbf{h})$ is defined by

$$p(v;u;\mathbf{g};\mathbf{h})=\Delta_{j-w,w}(\Psi(u,v);\mathbf{g};\mathbf{h}). \tag{4.4}$$

We note for future reference that, on recalling (4.1), and considering the term $Bu^{k-t}v^t$ in (2.1), it is apparent that the polynomial $p(v;u;\mathbf{g};\mathbf{h})$ is not identically zero.

On combining (4.2) and (4.3) via Hölder's inequality, we conclude that

$$|H(\theta)|^{2^j} \ll X^{2^{j+1}-j-2}\mathcal{G}(\theta), \tag{4.5}$$

where

$$\mathcal{G}(\theta) = \sum_{\mathbf{m}\in[-2X,2X]^j} \sum_{u\in J(\mathbf{g})} \sum_{v\in I(\mathbf{h})} e(\theta p(v;u;\mathbf{g};\mathbf{h})), \tag{4.6}$$

and here, and throughout, we adopt the convention that

$$\mathbf{m} = (m_1, \dots, m_j), \quad \mathbf{h} = (m_1, \dots, m_w) \quad \text{and} \quad \mathbf{g} = (m_{w+1}, \dots, m_j). \tag{4.7}$$

Define the exponential sum $\mathcal{G}_1(\alpha)$ by

$$\mathcal{G}_1(\alpha) = \sum_{\mathbf{m}} \sum_{u,v} e(\alpha p(v;u;\mathbf{g};\mathbf{h})),$$

where the summation is restricted to the values of $\mathbf{m}, u, v$ satisfying

$$\mathbf{m} \in [-2X, 2X]^j, \quad u \in J(\mathbf{g}), \quad v \in I(\mathbf{h}), \tag{4.8}$$

with

$$p(v;u;\mathbf{g};\mathbf{h}) \neq 0. \tag{4.9}$$

Since $p(v;u;\mathbf{g};\mathbf{h})$ is not identically zero, it follows from an elementary argument that the number of choices of $\mathbf{m}, u, v$ satisfying (4.8) and $p(v;u;\mathbf{g},\mathbf{h}) = 0$ is at most $O(X^{j+1})$. Consequently, one has

$$|\mathcal{G}(\alpha) - \mathcal{G}_1(\alpha)| \ll X^{j+1}.$$

Then in view of (4.5), we obtain

$$\begin{aligned}\mathcal{I}_{s+2^j}(X) = \int_0^1 |H(\alpha)|^{s+2^j} d\alpha &\ll X^{2^{j+1}-j-2} \int_0^1 \mathcal{G}(\alpha)|H(\alpha)|^s d\alpha \\ &\ll X^{2^{j+1}-1} \int_0^1 |H(\alpha)|^s d\alpha + X^{2^{j+1}-j-2} \int_0^1 |\mathcal{G}_1(\alpha)H(\alpha)^s| d\alpha.\end{aligned}$$

Let $\mathcal{T}$ denote the mean value

$$\mathcal{T}(X) = \int_0^1 |\mathcal{G}_1(\alpha)|^2 d\alpha. \tag{4.10}$$

Then an application of Schwarz's inequality leads us to the estimate

$$\mathcal{I}_{s+2^j}(X) \ll X^{2^{j+1}-1}\mathcal{I}_s(X) + X^{2^{j+1}-j-2}(\mathcal{T}(X))^{1/2}(\mathcal{I}_{2s}(X))^{1/2}. \tag{4.11}$$

Next observe that, in view of (4.8) and (4.9), and on considering the diophantine equation underlying (4.10), the mean value $\mathcal{T}(X)$ is bounded above by $\mathcal{K}(X)$, where $\mathcal{K}(X)$ denotes the number of integral solutions of the equation

$$p(x_1; y_1; \mathbf{g}_1; \mathbf{h}_1) = p(x_2; y_2; \mathbf{g}_2; \mathbf{h}_2), \tag{4.12}$$

with $\mathbf{m}_i \in [-2X, 2X]^j$, in the sense of (4.7), also with $|x_i|, |y_i| \leqslant X$ $(i = 1, 2)$, and subject to the conditions $p(x_i, y_i; \mathbf{g}_i; \mathbf{h}_i) \neq 0$ $(i = 1, 2)$. A comparison between (4.11) and the estimate claimed in the statement of the lemma therefore reveals that the desired conclusion follows immediately from the upper bound

$$\mathcal{K}(X) \ll X^{j+2+\delta+\varepsilon}. \tag{4.13}$$

We henceforth concentrate our efforts on establishing (4.13).

On recalling (4.4), a modicum of computation reveals that

$$p(x; y; \mathbf{g}; \mathbf{h}) = m_1 \dots m_j F(x, y; \mathbf{m}), \tag{4.14}$$

where

$$F(x, y; \mathbf{m}) = \sum_{i=t}^{k} D_i \phi_{k-i}(y; \mathbf{m}) \psi_i(x; \mathbf{m}), \tag{4.15}$$

in which D_i is an integer for $t \leqslant i \leqslant k$, and $D_t \neq 0$, and in which each $\psi_i(x; \mathbf{m})$ is a polynomial with integral coefficients of degree $i - w$ with respect to x, and each $\phi_{k-i}(y; \mathbf{m})$ is a polynomial with integral coefficients of degree $k - i - j + w$ with respect to y. In view of (4.1), one has $2 \leqslant k - t - j + w \leqslant k - j - 1$ and $t - w \geqslant 1$. Thus $F(x, y; \mathbf{m})$ has degree at least 1 with respect to x, and degree at least 2 and at most $k - j - 1$ with respect to y. We note also for future reference that when $w = 1$ and $j = 1$, one may take

$$\phi_{k-t}(y; m) = y^{k-t} \quad \text{and} \quad \psi_t(x; m) = m^{-1}((x + m)^t - x^t). \tag{4.16}$$

Finally, we observe that the argument surrounding equations (6.17) and (6.18) of [22] easily establishes that when $m_l \neq 0$ $(1 \leqslant l \leqslant j)$, then one has that the polynomial $F(x, y; \mathbf{m})$ is non-degenerate with respect to x and y.

When $\mathbf{m} \in [-2X, 2X]^j$, let $\rho(n; \mathbf{m})$ denote the number of integral solutions of the equation $p(x; y; \mathbf{g}; \mathbf{h}) = n$, with $|x|, |y| \leqslant X$. Then it follows from (4.14) and (4.12) that

$$\mathcal{K}(X) = \sum_{n \in \mathbb{Z} \setminus \{0\}} \Bigg(\sum_{\substack{|m_1| \leqslant 2X \\ m_1 | n}} \cdots \sum_{\substack{|m_j| \leqslant 2X \\ m_j | n}} \rho(n; \mathbf{m}) \Bigg)^2.$$

Consequently, on applying Cauchy's inequality in combination with an elementary estimate for the divisor function, one obtains

$$\begin{aligned}\mathcal{K}(X) &\ll X^{\varepsilon} \sum_{n\in\mathbb{Z}\setminus\{0\}} \sum_{\substack{\mathbf{m}\in[-2X,2X]^j\\ m_i\neq 0\ (1\leqslant i\leqslant j)}} \rho(n;\mathbf{m})^2\\ &\ll X^{j+\varepsilon} \max_{\substack{\mathbf{m}\in[-2X,2X]^j\\ m_i\neq 0\ (1\leqslant i\leqslant j)}} \sum_{n\in\mathbb{Z}\setminus\{0\}} \rho(n;\mathbf{m})^2\\ &= X^{j+\varepsilon} \max_{\substack{\mathbf{m}\in[-2X,2X]^j\\ m_i\neq 0\ (1\leqslant i\leqslant j)}} \mathcal{M}(X;\mathbf{m}),\end{aligned} \tag{4.17}$$

where $\mathcal{M}(X;\mathbf{m})$ denotes the number of solutions of the equation

$$F(x_1,y_1;\mathbf{m}) = F(x_2,y_2;\mathbf{m}), \tag{4.18}$$

with $|x_i|, |y_i| \leqslant X$ $(i=1,2)$.

We recall at this point that, by hypothesis, one has either

$$k=5, \quad t=2 \text{ or } 3 \quad \text{and} \quad j=1 \text{ or } 2,$$

or else

$$6\leqslant k\leqslant 10, \quad 2\leqslant t\leqslant k-3 \quad \text{and} \quad 1\leqslant j\leqslant k-3.$$

We now divide our argument into a number of cases, our aim being to make a choice of w, satisfying the condition (4.1), for which the estimates of §3 prove effective.

(a) (k,t,j) satisfies $j=k-3$. We take $w=t-1$. In this situation it is apparent that

$$k-t-j+w=2, \tag{4.19}$$

and also that

$$(k-j,t-w)=1. \tag{4.20}$$

Consider the shape of the polynomial $F(x,y;\mathbf{m})$ when the conditions (4.19) and (4.20) hold. We isolate the monomial of highest degree with respect to y that has highest degree with respect to x. In view of (4.15) and the associated discussion, this monomial has the shape

$$D_t x^{t-w} y^{k-t-j+w}. \tag{4.21}$$

Suppose, if possible, that there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a) and (b) of the statement of Lemma 3.2, and also satisfying the condition that $F(x,y)=g(G(x,y))$. Then the monomial of highest degree with respect to y that has highest degree with respect to x in the polynomial $g(G(x,y))$, must necessarily have the shape $Cx^{hl}y^{hm}$, where h is the degree of $g(t)$. But the condition (4.20) implies that

$$(k-t-j+w,t-w)=1, \tag{4.22}$$

and so the latter conclusion contradicts (4.21). It follows, in particular, that for no rational numbers λ and μ is it true that there exists a polynomial $f(x,y) \in \mathbb{Z}[x,y]$ for which the equation

$$F(x,y;\mathbf{m}) = \lambda f(x,y)^2 + \mu$$

is satisfied identically in x and y. But in view of (4.21) and (4.19), the polynomial $F(x,y;\mathbf{m})$ has degree precisely 2 with respect to y. Since, moreover, the coefficients of F are each bounded in absolute value by a fixed power of X, it follows from Lemma 3.4 that in this case one has

$$\mathcal{M}(X;\mathbf{m}) \ll X^{2+\varepsilon}. \tag{4.23}$$

On recalling (4.17), we find that (4.13) holds with $\delta = 0$, and thus the proof of the lemma in the case $j = k-3$ is complete.

(b) (k,t,j) satisfies $t-1 \leqslant j < k-3$. We take $w = t-1$, and find that (4.20), and hence also (4.22), remain true. In view of the discussion in case (a) above, it follows that there can exist no polynomials g and G that satisfy the hypotheses (a), (b), (c) of Lemma 3.2(i). We therefore deduce from Lemma 3.2(ii) that in this case one has

$$\mathcal{M}(X;\mathbf{m}) \ll X^{2+1/(k-j)+\varepsilon}. \tag{4.24}$$

Recalling (4.17) again, we now find that (4.13) holds with $\delta = 1/(k-j)$, and hence the proof of the lemma follows in the case currently under consideration.

(c) $(k,t,j) = (5,3,1)$. We take $w = 1$, and find that (4.19) holds. Then it follows from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^2y^2. It is possible that Lemma 3.4 succeeds in supplying the bound (4.23). If such is not the case, then there exists a polynomial $f(x,y) \in \mathbb{Z}[x,y]$, and rational numbers λ and μ, for which $F(x,y;\mathbf{m}) = \lambda f(x,y)^2 + \mu$. Moreover, it is apparent that λ must be non-zero, and our previous discussion ensures that $f(x,y)$ must be non-degenerate of total degree 2, with degree precisely one in terms of y. Thus we deduce that

$$\mathcal{M}(X;\mathbf{m}) \leqslant \sum_{n\in\mathbb{Z}} r_f(n;X)^2,$$

whence by Lemma 3.5 one again obtains the conclusion (4.23). Recalling (4.17), we now find that (4.13) holds with $\delta = 0$, and thus the proof of the lemma again follows.

By combining the conclusions of cases (a) and (b) above, one finds that when $k = 5$ and $j = 1$ or 2, our hypothesis that $t = 2$ or 3 leaves only the situation in which $(k,t,j) = (5,3,1)$ to consider. But the latter case is resolved in case (c) above, and so henceforth we may suppose that $6 \leqslant k \leqslant 10$. In the latter circumstances, cases (a) and (b) also dispose of all cases in which $t = 2$, and also all cases wherein $j \geqslant t-1$. Thus we may suppose henceforth that

$$6 \leqslant k \leqslant 10, \quad 3 \leqslant t \leqslant k-3 \quad \text{and} \quad 1 \leqslant j \leqslant t-2.$$

We treat the remaining allowable cases by hand.

(d) $(k,t,j)=(6,3,1)$, $(7,4,2)$. We take $w=j$, and find that (4.20) holds, since $(5,2)=1$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(e) $(k,t,j)=(8,3,1)$, $(8,4,1)$, $(8,5,1)$. We take $w=1$, and find that (4.20) holds, since $(7,t-1)=1$ for $t=3,4,5$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(f) $(k,t,j)=(8,5,3)$. We take $w=3$, and find that (4.20) holds, since $(5,2)=1$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(g) $(k,t,j)=(9,4,1)$, $(9,6,1)$. We take $w=1$, and find that (4.20) holds, since $(8,t-1)=1$ for $t=4,6$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(h) $(k,t,j)=(9,4,2)$, $(9,5,2)$, $(9,6,2)$. We take $w=2$, and find that (4.20) holds, since $(7,t-2)=1$ for $t=4,5,6$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(i) $(k,t,j)=(10,3,1)$, $(10,5,1)$, $(10,6,1)$. We take $w=1$, and find that (4.20) holds, since $(9,t-1)=1$ for $t=3,5,6$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(j) $(k,t,j)=(10,4,2)$, $(10,6,2)$. We take $w=1$, and find that (4.20) holds, since $(8,t-1)=1$ for $t=4,6$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(k) $(k,t,j)=(10,5,2)$, $(10,7,2)$. We take $w=2$, and find that (4.20) holds, since $(8,t-2)=1$ for $t=5,7$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(l) $(k,t,j)=(10,5,3)$, $(10,6,3)$, $(10,7,3)$. We take $w=3$, and find that (4.20) holds, since $(7,t-3)=1$ for $t=5,6,7$. The argument of part (b) therefore yields the bound (4.24), and hence also (4.13).

(m) $(k,t,j)=(9,6,4)$. We take $w=3$, and find that (4.19) and (4.20) both hold, since $(5,3)=1$. The argument of part (a) now establishes the bound (4.23), and hence also (4.13).

(n) $(k,t,j)=(10,7,5)$. We take $w=4$, and find that (4.19) and (4.20) both hold, since $(5,3)=1$. The argument of part (a) now establishes the bound (4.23), and hence also (4.13).

(o) $(k,t,j)=(7,4,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^3y^3. It is possible that Lemma 3.2 succeeds in supplying the bound (4.24). If such is not the case, then there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2. It is evident, moreover, that in such circumstances the degree of g must be 3, and the total degree of $G(x,y)$ must be 2, and also the degree of $G(x,y)$ with respect to y must be 1. In the latter circumstances, it follows from Lemma 3.5 that one has the estimate

$$\sum_{n\in\mathbb{Z}} r_G(n;X)^2 \ll X^{2+\varepsilon}, \tag{4.25}$$

whence by Lemma 3.2(i),

$$\begin{aligned}\mathcal{M}(X;\mathbf{m}) &\ll X^{2+1/(k-j)+\varepsilon} + X^{\varepsilon}\sum_{n\in\mathbb{Z}} r_G(n;X)^2 \\ &\ll X^{2+1/(k-j)+\varepsilon}. \end{aligned}\tag{4.26}$$

Then in any case, one has the upper bound (4.24), and hence (4.13).

(p) $(k,t,j)=(8,5,2)$. We take $w=2$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^3y^3. The desired bound (4.24), and hence (4.13), now follows by the argument of case (o).

(q) $(k,t,j)=(9,5,3)$. We take $w=2$, and find from (4.21) that we may again apply the argument of case (p).

(r) $(k,t,j)=(9,6,3)$. We take $w=3$, and find from (4.21) that we may again apply the argument of case (p).

(s) $(k,t,j)=(10,6,4)$. We take $w=3$, and find from (4.21) that we may again apply the argument of case (p).

(t) $(k,t,j)=(10,7,4)$. We take $w=4$, and find from (4.21) that we may again apply the argument of case (p).

(u) $(k,t,j)=(7,3,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^2y^4. It follows that if there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2, then g must have degree 2. Moreover, in the polynomial $G(x,y)$, the monomial of highest degree with respect to y, that has highest degree with respect to x, must have the shape $C'xy^2$. In such circumstances, one may apply the argument of case (a) above to obtain the bound (4.25). Then the estimate (4.24) follows in all circumstances from Lemma 3.2, and this suffices to establish (4.13).

(v) $(k,t,j)=(8,4,2)$. We take $w=2$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^2y^4. The desired bound (4.24), and hence (4.13), now follows by the argument of case (u).

(w) $(k,t,j)=(9,5,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^4y^4. It follows that if there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2, then g must have degree either 2 or 4. When the degree of g is 4, the polynomial $G(x,y)$ must have total degree 2, and the degree of $G(x,y)$ with respect to y must be 1. In these circumstances, Lemma 3.5 establishes the estimate (4.25). When the degree of g is 2, meanwhile, we may suppose without loss of generality that there are no rational numbers λ and μ for which a polynomial $f(x,y)\in\mathbb{Z}[x,y]$ exists satisfying $G(x,y)=\lambda f(x,y)^2+\mu$ (if such a polynomial f were to exist, then we would be in the situation already considered wherein the degree of g was presumed to be 4). But then, in the polynomial

$G(x,y)$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape $C'x^2y^2$. The hypotheses of Lemma 3.4 are therefore satisfied with F replaced by G, and the upper bound (4.25) again follows. We may therefore conclude from Lemma 3.2(i) that in either case, one has the estimate (4.26). If no such polynomials G and g exist, on the other hand, then the estimate (4.26) is immediate from Lemma 3.2(ii).

(x) $(k,t,j)=(10,4,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^3y^6. It follows that if there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2, then g must have degree 3. Moreover, in the polynomial $G(x,y)$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape $C'xy^2$. Thus we may proceed as in case (u) to obtain the desired estimate (4.13).

(y) $(k,t,j)=(9,3,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^2y^6. It follows that if there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2, then g must have degree 2, and $G(x,y)$ must have the shape

$$G(x,y)=\alpha xy^3+\beta y^3+H(x,y), \tag{4.27}$$

with $H(x,y)$ of degree at most 2 with respect to y. In view of (4.16), one finds that with a suitable non-zero constant K, one has that

$$F(x,y;m)=Ky^6(3x^2+3xm+m^2)+I(x,y;m), \tag{4.28}$$

where $I(x,y;m)$ is a polynomial of degree at most 5 with respect to y. Since we may suppose that g has degree 2, it follows from (4.27) and (4.28) that there is a non-zero number a with

$$a(\alpha xy^3+\beta y^3)^2=Ky^6(3x^2+3xm+m^2).$$

On equating coefficients of powers of x, we find that

$$a\alpha^2=3K,\quad 2a\alpha\beta=3Km,\quad a\beta^2=Km^2,$$

whence

$$9K^2m^2=4(a\alpha^2)(a\beta^2)=12K^2m^2.$$

This yields a contradiction whenever $m\neq 0$, as we may suppose. In this way we find that no such polynomials G, g exist, and hence Lemma 3.2(ii) establishes that the estimate (4.26) holds.

(z) $(k,t,j)=(10,7,1)$. We take $w=1$, and find from (4.21) that in the polynomial $F(x,y;\mathbf{m})$, the monomial of highest degree with respect to y, that has highest degree with respect to x, has the shape Cx^6y^3. It follows that if there exist polynomials $G(x,y)\in\mathbb{Z}[x,y]$ and $g(t)\in\mathbb{Q}[t]$ satisfying the conditions (a), (b), (c) of Lemma 3.2, then g must have degree 3, and $G(x,y)$ must have the shape

$$G(x,y)=\alpha yx^2+\beta yx+\gamma y+H(x), \tag{4.29}$$

with $H(x)$ a polynomial independent of y. In view of (4.16), one finds that with a suitable non-zero constant K, one has that

$$\begin{aligned}F(x,y;m)=&Ky^3(7x^6+21x^5m+35x^4m^2+35x^3m^3+21x^2m^4+7xm^5+m^6)\\&+I(x,y;m),\end{aligned} \tag{4.30}$$

where $I(x,y;m)$ is a polynomial of degree at most 2 with respect to y. Since we may suppose that g has degree 3, it follows from (4.29) and (4.30) that there is a non-zero number a with

$$\begin{aligned}a(\alpha x^2y+&\beta xy+\gamma y)^3\\&=Ky^3(7x^6+21x^5m+35x^4m^2+35x^3m^3+21x^2m^4+7xm^5+m^6).\end{aligned}$$

On equating coefficients of powers of x, we find that

$$a\alpha^3=7K,\quad 3a\alpha^2\beta=21Km,\quad a(3\alpha^2\gamma+3\alpha\beta^2)=35Km^2,\quad a\gamma^3=Km^6.$$

Thus we deduce that

$$27\alpha^6a^2Km^6=a^3(3\alpha^2\gamma)^3=(35Km^2-3a\alpha\beta^2)^3,$$

whence

$$\begin{aligned}3^37^5(Km)^6&=(35Km^2(a\alpha^3)-3(a\alpha^2\beta)^2)^3\\&=(245K^2m^2-147K^2m^2)^3=(98K^2m^2)^3.\end{aligned}$$

Since $3^37^5\neq 98^3$, we obtain a contradiction whenever $m\neq 0$, as we may suppose. In this way, we find that no such polynomials G, g exist, and hence Lemma 3.2(ii) establishes that the estimate (4.26) holds.

On collecting together the conclusions of cases (a)–(z), we find that the estimate (4.13) holds in all circumstances under consideration. The conclusion of the lemma now follows immediately from (4.11).

5. The completion of the proof of Theorem 1. We are now prepared to complete the proof of Theorem 1. We begin with an induction based on the use of Lemma 4.1 for the small moments.

Lemma 5.1. *Let $\Psi(x,y) \in \mathbb{Z}[x,y]$ be a non-degenerate form of degree k, with $5 \leqslant k \leqslant 10$, of the shape discussed in the opening paragraph of* §4*. Then for each j with $3 \leqslant j \leqslant k-2$, and for each positive number ε, one has*

$$\int_0^1 |H_\Psi(\alpha;X)|^{2^{j-1}} d\alpha \ll X^{2^j-j+1/(k-j+2)+\varepsilon}.$$

Proof. The conclusion of the lemma is either immediate from part (i) of Lemma 4.1, or else we may apply part (ii) of that lemma. By part (i) of Theorem 1, which we have already established in Lemma 2.2, one has

$$\int_0^1 |H_\Psi(\alpha;X)|^2 d\alpha \ll X^{2+\varepsilon}.$$

Thus we have

$$\mathcal{I}_2(X) \ll X^{2+\varepsilon}.$$

Suppose that, in fact, one has the estimate

$$\mathcal{I}_{2^{j-1}}(X) \ll X^{2^j-j+1/(k-j+2)+\varepsilon}, \tag{5.1}$$

for $2 \leqslant j \leqslant J$, where J is an integer with $2 \leqslant J \leqslant k-3$. We apply part (ii) of Lemma 4.1 with $j = J-1$ and $s = 2^{J-1}$ in order to obtain

$$\mathcal{I}_{s+2^{J-1}}(X) \ll X^{2^J-1}\mathcal{I}_s(X) + X^{2^J-\frac{1}{2}(J+1-\delta)+\varepsilon}(\mathcal{I}_{2s}(X))^{1/2},$$

with $\delta = 1/(k-J+1)$. On employing the inductive hypothesis (5.1), we obtain

$$\mathcal{I}_{2^J}(X) \ll X^{2^{J+1}-J-1+1/(k-J+2)+\varepsilon} + X^{2^J-\frac{1}{2}(J+1-\delta)+\varepsilon}(\mathcal{I}_{2^J}(X))^{1/2},$$

whence

$$\mathcal{I}_{2^J}(X) \ll X^{2^{J+1}-J-1+1/(k-J+1)+\varepsilon}.$$

This establishes the inductive hypothesis for $j = J+1$, and thus the conclusion of the lemma follows by induction.

Lemma 5.2. *With the hypotheses of the statement of Lemma* 5.1*, one has*

$$\int_0^1 |H_\Psi(\alpha;X)|^{\frac{9}{32}2^k} d\alpha \ll X^{\frac{9}{16}2^k-k+1+\varepsilon}.$$

Proof. As in the proof of the previous lemma, the desired conclusion is either immediate from part (i) of Lemma 4.1, or else we may apply part (ii) of that lemma. By the conclusion of Lemma 5.1 with $j = k-2$, one has

$$\mathcal{I}_{2^{k-3}}(X) = \int_0^1 |H_\Psi(\alpha;X)|^{2^{k-3}} d\alpha \ll X^{2^{k-2}-k+9/4+\varepsilon}. \tag{5.2}$$

On applying part (ii) of Lemma 4.1 with $s = 2^{k-3}$ and $j = k-3$, one obtains

$$\mathcal{I}_{2^{k-2}}(X) \ll X^{2^{k-2}-1}\mathcal{I}_{2^{k-3}}(X) + X^{2^{k-2}-(k-1)/2+\varepsilon}(\mathcal{I}_{2^{k-2}}(X))^{1/2},$$

whence by (5.2),

$$\mathcal{I}_{2^{k-2}}(X) \ll X^{2^{k-1}-k+5/4+\varepsilon}. \tag{5.3}$$

An application of Hölder's inequality establishes the upper bound

$$\begin{aligned}\mathcal{I}_{\frac{5}{32}2^k}(X) &= \int_0^1 |H_\Psi(\alpha;X)|^{\frac{5}{32}2^k}\,d\alpha \\ &\leqslant \Big(\int_0^1 |H_\Psi(\alpha;X)|^{2^{k-2}}d\alpha\Big)^{1/4}\Big(\int_0^1 |H_\Psi(\alpha;X)|^{2^{k-3}}d\alpha\Big)^{3/4}.\end{aligned}$$

Thus, by (5.2) and (5.3) we deduce that

$$\mathcal{I}_{\frac{5}{32}2^k}(X) \ll X^{\frac{5}{16}2^k-k+2+\varepsilon}. \tag{5.4}$$

A second application of part (ii) of Lemma 4.1, now with $s = \frac{5}{32}2^k$ and $j = k-3$, gives the estimate

$$\mathcal{I}_{\frac{9}{32}2^k}(X) \ll X^{2^{k-2}-1}\mathcal{I}_{\frac{5}{32}2^k}(X) + X^{2^{k-2}-(k-1)/2+\varepsilon}\left(\mathcal{I}_{\frac{5}{16}2^k}(X)\right)^{1/2}.$$

But a trivial estimate for $H_\Psi(\alpha;X)$ demonstrates that

$$\mathcal{I}_{\frac{5}{16}2^k}(X) \ll X^{2^{k-4}}\int_0^1 |H_\Psi(\alpha;X)|^{\frac{9}{32}2^k}\,d\alpha = X^{2^{k-4}}\mathcal{I}_{\frac{9}{32}2^k}(X).$$

In view of (5.4), therefore, we obtain

$$\mathcal{I}_{\frac{9}{32}2^k}(X) \ll X^{\frac{9}{16}2^k-k+1+\varepsilon} + X^{\frac{9}{32}2^k-(k-1)/2+\varepsilon}\left(\mathcal{I}_{\frac{9}{32}2^k}(X)\right)^{1/2},$$

and the conclusion of the lemma follows immediately.

The large moments are estimated via the Hardy-Littlewood method by means of a treatment contained, in all essentials, within the proof of Lemma 7.4 of [22]. We include an account of the proof for the sake of completeness. We first require a major arc estimate stemming from our version of Weyl's inequality.

Lemma 5.3. *Suppose that $\Phi(x,y) \in \mathbb{Z}[x,y]$ is a non-degenerate form of degree $d \geqslant 3$, and let $\alpha \in \mathbb{R}$.*

(i) Suppose that there exist $r \in \mathbb{Z}$ and $q \in \mathbb{N}$ with $(r,q)=1$ and $|\alpha - r/q| \leqslant q^{-2}$. Then for each $\varepsilon > 0$, one has

$$\sum_{1\leqslant x\leqslant X}\sum_{1\leqslant y\leqslant X} e(\alpha\Phi(x,y)) \ll X^{2+\varepsilon}\left(q^{-1}+X^{-1}+qX^{-d}\right)^{2^{2-d}}.$$

(ii) Whenever $r \in \mathbb{Z}$ and $q \in \mathbb{N}$ satisfy $1 \leqslant q \leqslant X$ and $|q\alpha - r| \leqslant X^{1-d}$, one has

$$\sum_{1\leqslant x\leqslant X}\sum_{1\leqslant y\leqslant X} e(\alpha\Phi(x,y)) \ll X^{2+\varepsilon}(q+X^d|q\alpha-r|)^{-2^{2-d}}.$$

Proof. The first conclusion is immediate from Theorem 1 of [22], and the second conclusion is Lemma 7.3 of [22].

Lemma 5.4. *With the hypotheses of the statement of Lemma* 5.1, *one has*

$$\int_0^1 |H_\Psi(\alpha;X)|^{\frac{17}{32}2^k}\,d\alpha \ll X^{\frac{17}{16}2^k-k+\varepsilon}.$$

Proof. For the sake of concision, we abbreviate $H_\Psi(\alpha;X)$ to $H(\alpha)$. When $r\in\mathbb{Z}$ and $q\in\mathbb{N}$, write

$$\mathfrak{M}(q,r)=\left\{\alpha\in[0,1)\ :\ |q\alpha-r|\leqslant X^{1-k}\right\}.$$

Take $\mathfrak{M}$ to be the union of the intervals $\mathfrak{M}(q,r)$ with $0\leqslant r\leqslant q\leqslant X$ and $(r,q)=1$. Note that the intervals occurring in the latter union are disjoint. Also, write $\mathfrak{m}=[0,1)\setminus\mathfrak{M}$. Since Lemma 5.3(i) yields the estimate

$$\sup_{\alpha\in\mathfrak{m}}|H(\alpha)|\ll X^{2-2^{2-k}+\varepsilon},$$

and Lemma 5.2 establishes that

$$\int_0^1 |H(\alpha)|^{\frac{9}{32}2^k}\,d\alpha \ll X^{\frac{9}{16}2^k-k+1+\varepsilon},$$

we deduce that

$$\begin{aligned}\int_{\mathfrak{m}} |H(\alpha)|^{\frac{17}{32}2^k}\,d\alpha &\ll \Bigl(\sup_{\alpha\in\mathfrak{m}}|H(\alpha)|\Bigr)^{2^{k-2}}\int_0^1 |H(\alpha)|^{\frac{9}{32}2^k}\,d\alpha\\ &\ll X^{\frac{17}{16}2^k-k+\varepsilon}.\end{aligned}\tag{5.5}$$

On making use of Lemma 5.3(ii) and the definition of $\mathfrak{M}$, on the other hand, we obtain

$$\begin{aligned}\int_{\mathfrak{M}} |H(\alpha)|^{\frac{17}{32}2^k}\,d\alpha &\ll X^{\frac{17}{16}2^k+\varepsilon}\sum_{1\leqslant q\leqslant X}\sum_{\substack{a=1\\(a,q)=1}}^{q}\int_{|\beta|\leqslant(qX^{k-1})^{-1}}(q+X^kq|\beta|)^{-2}\,d\beta\\ &\ll X^{\frac{17}{16}2^k-k+\varepsilon}\sum_{1\leqslant q\leqslant X}\sum_{\substack{a=1\\(a,q)=1}}^{q} q^{-2}\\ &\ll X^{\frac{17}{16}2^k-k+2\varepsilon}.\end{aligned}\tag{5.6}$$

Consequently, on combining the estimates (5.5) and (5.6), we arrive at the upper bound

$$\begin{aligned}\int_0^1 |H(\alpha)|^{\frac{17}{32}2^k}\,d\alpha &= \int_{\mathfrak{M}} |H(\alpha)|^{\frac{17}{32}2^k}\,d\alpha+\int_{\mathfrak{m}} |H(\alpha)|^{\frac{17}{32}2^k}\,d\alpha\\ &\ll X^{\frac{17}{16}2^k-k+2\varepsilon},\end{aligned}$$

and so the conclusion of the lemma follows immediately.

On recalling the conclusion of Lemma 2.1, and noting that all of the moments occurring in the statement of Theorem 1(iii) are even, one finds that the upper bounds provided in Theorem 1(iii) are immediate from Lemmata 5.1, 5.2 and 5.4. The same is true also for the first three bounds recorded in Theorem 1(ii), but in this case, for the second two estimates, one combines Lemmata 5.1, 5.2 and 5.4 via Hölder's inequality in the respective shapes

$$\begin{aligned}\int_0^1 |f_\Phi(\alpha;P)|^8 d\alpha &\ll \int_0^1 |H_\Psi(\alpha;X)|^8 d\alpha \\ &\ll \left(\int_0^1 |H_\Psi(\alpha;X)|^4 d\alpha\right)^{1/5} \left(\int_0^1 |H_\Psi(\alpha;X)|^9 d\alpha\right)^{4/5},\end{aligned}$$

and

$$\begin{aligned}\int_0^1 |f_\Phi(\alpha;P)|^{10} d\alpha &\ll \int_0^1 |H_\Psi(\alpha;X)|^{10} d\alpha \\ &\ll \left(\int_0^1 |H_\Psi(\alpha;X)|^9 d\alpha\right)^{7/8} \left(\int_0^1 |H_\Psi(\alpha;X)|^{17} d\alpha\right)^{1/8}.\end{aligned}$$

The final estimate of Theorem 1(ii), on the other hand, may be established along the lines of the proof of Lemma 5.4, now working from the 10th moment

$$\int_0^1 |f_\Phi(\alpha;P)|^{10} d\alpha \ll P^{127/8+\varepsilon},$$

together with the minor arc bound

$$\sup_{\alpha\in\mathfrak{m}} |f_\Phi(\alpha;P)| \ll P^{15/8+\varepsilon},$$

which is immediate from Lemma 5.3.

References

1. M. A. Bennett, N. P. Dummigan and T. D. Wooley, *The representation of integers by binary additive forms*, Compositio Math. **111** (1998), 15–33.
2. B. J. Birch, *Forms in many variables*, Proc. Roy. Soc. Ser. A **265** (1961), 245–263.
3. B. J. Birch and H. Davenport, *Note on Weyl's inequality*, Acta Arith. **7** (1961/62), 273–277.
4. K. D. Boklan, *The asymptotic formula in Waring's problem*, Mathematika **41** (1994), 329–347.
5. E. Bombieri and J. Pila, *The number of integral points on arcs and ovals*, Duke Math. J. **59** (1989), 337–357.
6. J. Brüdern and T. D. Wooley, *The addition of binary cubic forms*, R. Soc. Lond. Philos. Trans. Ser. A. Math. Phys. Eng. Sci. **356** (1998), 701–737.
7. S. Chowla and H. Davenport, *On Weyl's inequality and Waring's problem for cubes*, Acta Arith. **6** (1961/62), 505–521.

8. T. Estermann, *Einige Sätze über quadratfrei Zahlen*, Math. Ann. **105** (1931), 653–662.
9. A. Granville, *Bounding the coefficients of a divisor of a given polynomial*, Monatsh. Math. **109** (1990), 271–277.
10. D. R. Heath-Brown, *Weyl's inequality, Hua's inequality, and Waring's problem*, J. London Math. Soc. (2) **38** (1988), 216–230.
11. L.-K. Hua, *On Waring's problem*, Quart. J. Math. Oxford **9** (1938), 199–202.
12. J. Pila, *Density of integer points on plane algebraic curves*, Internat. Math. Res. Notices (1996), 903–912.
13. W. M. Schmidt, *The density of integer points on homogeneous varieties*, Acta Math. **154** (1985), 243–296.
14. C. M. Skinner and T. D. Wooley, *Sums of two kth powers*, J. Reine Angew. Math. **462** (1995), 57–68.
15. C. M. Skinner and T. D. Wooley, *On the paucity of non-diagonal solutions in certain diagonal diophantine systems*, Quart. J. Math. Oxford (2) **48** (1997), 255–277.
16. W. Y. Tsui and T. D. Wooley, *The paucity problem for simultaneous quadratic and biquadratic equations*, Math. Proc. Cambridge Philos. Soc. **126** (1999), 209–221.
17. R. C. Vaughan, *On Waring's problem for cubes*, J. Reine Angew. Math. **365** (1986), 122–178.
18. R. C. Vaughan, *On Waring's problem for smaller exponents*, Proc. London Math. Soc. (3) **52** (1986), 445–463.
19. R. C. Vaughan, *The Hardy-Littlewood Method, second edition*, Cambridge University Press, 1997.
20. R. C. Vaughan and T. D. Wooley, *Further improvements in Waring's problem*, Acta Math. **174** (1995), 147–240.
21. R. C. Vaughan and T. D. Wooley, *Further improvements in Waring's problem, IV: higher powers*, Acta Arith. **94** (2000), 203–285.
22. T. D. Wooley, *On Weyl's inequality, Hua's lemma, and exponential sums over binary forms*, Duke Math. J. **100** (1999), 373–423.

Department of Mathematics, University of Michigan, East Hall, 525 East University Avenue, Ann Arbor, MI 48109-1109, U.S.A.

E-mail address: `wooley@math.lsa.umich.edu`