BURGESS BOUNDS FOR SHORT CHARACTER SUMS EVALUATED AT
FORMS II: THE MIXED CASE

LILLIAN B. PIERCE

ABSTRACT. This work proves a Burgess bound for short mixed character sums in \( n \) dimensions. The non-principal multiplicative character of prime conductor \( q \) may be evaluated at any “admissible” form, and the additive character may be evaluated at any real-valued polynomial. The resulting upper bound for the mixed character sum is nontrivial when the length of the sum is at least \( q^\beta \) with \( \beta > 1/2 - 1/(2(n+1)) \) in each coordinate. This work capitalizes on the recent stratification of multiplicative character sums due to Xu, and the resolution of the Vinogradov Mean Value Theorem in arbitrary dimensions.

1. INTRODUCTION

Let \( \chi \) be a non-principal multiplicative Dirichlet character modulo a prime \( q \). Let \( g \in \mathbb{R}[x_1, \ldots, x_n] \) be a polynomial of total degree \( d \geq 1 \), and let \( F \in \mathbb{Z}[x_1, \ldots, x_n] \) be a form of degree \( D \geq 1 \). Define

\[
S(F, g; N, H) = \sum_{x \in (N, N+H]} e(g(x))\chi(F(x)),
\]

where \( N = (N_1, \ldots, N_n) \), \( H = (H_1, \ldots, H_n) \) and \( x \in (N, N+H] \) denotes those tuples \( x \in \mathbb{Z}^n \) such that \( x_i \in (N_i, N_i + H_i) \) for each \( 1 \leq i \leq n \). Such character sums are the building blocks of many methods in analytic number theory. The trivial bound is \( |S(F, g; N, H)| \leq H_1 \cdots H_n \), and bounds that improve on this have many applications. Conjecturally one could expect square-root cancellation to hold, for appropriate functions \( F \) and \( g \). In the particular case of short sums, namely those in which \( H_i \leq q^{1/2} \), this remains out of reach, and a central goal is to provide any nontrivial upper bound, that is \( |S(F, g; N, H)| = o(H_1 \cdots H_n) \), valid for general choices of \( F, g \).

Historically the most fundamental case has been that of a one-dimensional multiplicative character sum, in which case Burgess’ work set the gold standard, also establishing a long-standing subconvexity result for Dirichlet \( L \)-functions; see e.g. [Bur57, Bur63]. (This subconvexity bound has only now been improved, in [PY19].) Burgess’s method of proof has been resistant to substantial improvement, but recent work has begun to generalize the method to new settings. For a survey of Burgess bounds, in particular in the case of purely multiplicative sums, we refer to the overview given in [PX19].

In this paper we prove a Burgess bound for mixed sums of the form \( S(F, g; N, H) \), for the largest class of forms \( F \) (acting nontrivially on all variables) for which one would anticipate a nontrivial bound could be obtained. We formally define this set of “admissible forms,” before stating our main result.

**Condition 1.1** \((\Delta, q)\)-admissible. Fix a prime \( q \) and an integer \( \Delta \geq 1 \). A polynomial \( f \in \mathbb{F}_q[x_1, \ldots, x_n] \) is \((\Delta, q)\)-admissible if the following holds. Upon writing \( f = g^\Delta h \) where \( g, h \in \mathbb{F}_q[x_1, \ldots, x_n] \) and \( h \) is \( \Delta \)-th power-free over \( \mathbb{F}_q \), then \( h \) cannot be made independent of a variable after a linear transformation, i.e. there exists no \( A \in \text{GL}_n(\mathbb{F}_q) \) such that \( h(xA) \in \mathbb{F}_q[x_2, \ldots, x_n] \).

We recall that a polynomial \( h \) is said to be \( \Delta \)-th power-free over \( \mathbb{F}_q \) if when \( h \) is factored over \( \mathbb{F}_q \) into irreducible pairwise non-associate factors \( h_i \), each \( h_i \) appears to a power strictly smaller than \( \Delta \). If a form \( F = G^\Delta H \) with \( G, H \in \mathbb{Z}[x_1, \ldots, x_n] \) where \( H \) is \( \Delta \)-th power-free and \( H \) cannot be made independent of a variable after a \( GL_n(\mathbb{Z}) \) change of variables, then \( F \) has \((\Delta, q)\)-admissible
reduction modulo \( q \) for all but finitely many primes \( q \). Such forms are generic amongst the set of all forms in \( \mathbb{Z}[x_1, \ldots, x_n] \) of degree \( D \). Moreover, an example of such a form is \( x_1^D + \cdots + x_n^D \). See [PX19 §3.1] for further details on these facts.

Our main result is the following theorem.

**Theorem 1.2.** Fix \( n \geq 2 \) and \( d, D \geq 1 \). Let \( q \) be a fixed prime, and let \( \chi \) be a non-principal character of conductor \( q \) and order \( \Delta \). Let \( F \in \mathbb{Z}[x_1, \ldots, x_n] \) be a form of degree \( D \) such that its reduction modulo \( q \) is \((\Delta, q)\)-admissible. Let \( g \in \mathbb{R}[x_1, \ldots, x_n] \) be a polynomial of total degree \( d \geq 1 \). Define for each integer \( r \),

\[
\Theta = \Theta_{n,r} = \left\lfloor \frac{r - 1}{n - 1} \right\rfloor, \quad M = M_{d,n} = d \left( \frac{n + d}{n} \right) \frac{n}{n + 1}.
\]

Let \( H = (H, \ldots, H) \). Then for every integer \( r \geq 1 \) such that \( \Theta = \Theta_{n,r} > M \) and \( H < q^{\frac{1}{2} + \frac{1}{8\Theta - M}} \),

\[
|S(F, g; \mathbf{N}, \mathbf{H})| \ll H^{n+\frac{1}{2}} q^{\frac{n(\Theta - M)+1}{4\Theta - M}} q^\varepsilon,
\]

for every \( \varepsilon > 0 \); the implied constant may depend on \( n, d, D, \Delta, r, \varepsilon \) but is independent of \( g, F \).

In dimension \( n = 1 \), an inequality analogous to Theorem 1.2 with \( \Theta = \Theta_{1,r} = r \) and \( F(x) = x \), was proved by Heath-Brown and the author [HBP15]. (At the time it was conditional on the Main Conjecture in the Vinogradov Mean Value Method, which now has been proved [Woo16, BDG16], making [HBP15] unconditional.) Upon setting \( \Theta_{1,r} = r \) in each instance where \( \Theta = \Theta_{n,r} \) appears in this paper, the method of the present paper also recovers this case, but we focus on \( n \geq 2 \).

Theorem 1.2 is the first Burgess bound for mixed sums in dimensions \( n \geq 2 \) in which \( F \) is allowed to be any admissible form. In this sense it is a natural sequel to the work of the author with Xu [PX19], which introduced the \((\Delta, q)\)-admissible class of forms, in the setting of purely multiplicative sums (namely the case \( S(F, 0; \mathbf{N}, \mathbf{H}) \)). Theorem 1.2 is of comparable strength to the purely multiplicative case considered in [PX19]. Precisely, define

\[
\beta_n = \frac{1}{2} - \frac{1}{2(n + 1)}.
\]

Theorem 1.2 provides a nontrivial bound of the form \( |S(F, g; \mathbf{N}, \mathbf{H})| \ll H^n q^{-\delta} \) when \( H > q^{\beta_n + \kappa} \) for some sufficiently small \( \kappa > 0 \), and the savings is of the strength

\[
\delta \approx \frac{(n + 1)^2}{4(n - 1)} \kappa^2
\]

as \( \kappa \to 0 \); see [6.1] for details. In particular, note that this savings is independent of the degree \( D \) of the form \( F \) and the degree \( d \) of the polynomial \( g \); this is achieved by an application of the sharp upper bound in the multi-dimensional Vinogradov Mean Value Theorem, due to [PP13] in many cases and [GZ19] in complete generality.

Earlier work on two special types of mixed sums in dimensions \( n \geq 2 \) appeared in two recent papers. In the special case \( F(x) = x_1 \cdots x_n \), the author proved nontrivial bounds for \( |S(F, g; \mathbf{N}, \mathbf{H})| \) as long as \( H_i > q^{1/4 + \kappa} \) for some small \( \kappa > 0 \) [Pie16]. See also the preprint of Kerr [Ker14] in the case that \( F(x) = \prod_{i=1}^n L_i(x) \) is the product of \( n \) linear forms \( L_i \) that are linearly independent over \( \mathbb{F}_q \). In each of these special settings, additional structure allowed the argument to achieve the Burgess threshold \( q^{1/4 + \kappa} \) for any \( \kappa > 0 \), in any dimension.

### 1.1. Method of proof.

The proof of Theorem 1.2 capitalizes upon recent foundational work of two kinds:

1. Xu’s stratification of multiplicative character sums [Xu18];
In particular, the proof of Theorem 1.2 requires a sharp upper bound for the number of solutions to the generalized Vinogradov system
\[ x_1^\beta + \cdots + x_n^\beta = x_{r+1}^\beta + \cdots + x_{2r}^\beta, \quad 1 \leq |\beta| \leq d \]
with \( x_j \in \mathbb{Z}^n \) and \( 1 \leq x_{j,i} \leq X \) for \( 1 \leq j \leq 2r, 1 \leq i \leq n \); here \( \beta = (\beta_1, \ldots, \beta_n) \) is a multi-index with \( |\beta| = \beta_1 + \cdots + \beta_n \). In fact our work also applies to more general translation-dilation invariant systems (see Theorem 1.1), and as our method naturally uses the properties of such systems, we introduce the relevant terminology in the following section.

2. INTRODUCING THE ASSOCIATED VINogradov SYSTEM

To prove Theorem 1.2 for a fixed (admissible) choice of \( F \in \mathbb{Z}[x_1, \ldots, x_n] \) and \( g \in \mathbb{R}[x_1, \ldots, x_n] \), our primary object of focus will be
\[ T(F, G; N, H) = \sup_{g \in \mathcal{F}_0(G)} \sup_{K \leq H} \left| \sum_{x \in (N, N+K]} e(g(x)) \chi(F(x)) \right|, \]
in which \( \mathcal{F}_0(G) \) is a certain set of polynomials including the fixed polynomial \( g \) of our choice. Our main bound for \( T(F, G; N, H) \) involves counting the number of solutions to a system of Diophantine equations, which we now introduce precisely.

To prove Theorem 1.2, we may take \( G \) to be the set of all non-constant monomials in \( n \) variables of total degree at most \( d \), that is
\[ G = \{ x_1^{\beta_1}, \ldots, x_n^{\beta_n} : \beta = (\beta_1, \ldots, \beta_n), 1 \leq |\beta| \leq d \}, \]
in which \( |\beta| = \beta_1 + \cdots + \beta_n \). (Momentarily, we will also consider other sets of monomials.) Given any set \( G \) of monomials, we define \( \mathcal{F}_0(G) \) to be the set of all real-variable polynomials that are linear combinations of the elements in \( G \cup \{1\} \) (that is, including constant terms). We will call the set \( G \) defined in (2.1) the standard system of monomials in \( n \) variables of degree at most \( d \). With this choice of \( G \), given any polynomial \( g \in \mathbb{R}[x_1, \ldots, x_n] \) of degree \( d \), we can embed it in \( \mathcal{F}_0(G) \). In particular, \( |S(F, g; N, H)| \leq T(F, G; N, H) \).

The main outcome of the Burgess argument we develop is an upper bound for \( T(F, G; N, H) \) in terms of (i) a complete multiplicative character sum and (ii) a complete additive character sum. We apply Xu’s stratification [Xu18] to bound the complete multiplicative character sum. We evaluate the complete additive character sum precisely, and then dominate the outcome by the number of integral solutions to the (multi-dimensional) Vinogradov system of Diophantine equations associated to the system \( G \) given in (2.1) (sometimes also called a Parsell-Vinogradov system of equations when \( n \geq 2 \)). Let \( J_r(G, X) \) denote the number of integral solutions to the system (2.4) with \( 1 \leq x_{j,i} \leq X \) for \( 1 \leq j \leq 2r, 1 \leq i \leq n \). We also let \( M = M(G) \) denote the sum of the total degree of all multi-indices in \( \{ \beta \in \mathbb{Z}_{\geq 0}^n : 1 \leq |\beta| \leq d \} \).

Our main result in the context of Theorem 1.2 is as follows. For any \( n \geq 2 \) and \( r \geq 1 \), define \( \Theta = \Theta_{n,r} \) as in Theorem 1.2. For any \( r \geq 1 \), and any \( H = (H, \ldots, H) \) with \( H < q^{1/2+1/4(\Theta-M)} \), for any integer \( P \) such that \( 1 \leq P \leq Hq^{-1/2\Theta} \),
\[ T(F, G; N, H) \ll (H/P)^{M/2r} H^{-n/2r} P^{n-1/2r} q^{d/4} (\log q)^{n+1} \{ J_r(G, 2H/P)^{1/2r} + q^{1/4r} (H/P)^{\Theta/2r} \}. \]
Theorem 1.2 then follows from an appropriate bound for \( J_r(G, X) \) provided by the multi-dimensional Vinogradov Mean Value Theorem, and an optimal choice for \( P \) in terms of \( H, q \).
2.1. Remark on more general systems $G$. Without any additional difficulty, our main arguments can replace the standard system $G$ specified in (2.1) by any reduced monomial translation-dilation invariant system. This terminology was introduced in [PPW13], and we briefly recall the definitions. A given collection $G = \{g_1, \ldots, g_R\}$ of $R$ non-constant monomials in $\mathbb{Z}[x_1, \ldots, x_n]$ is said to be translation-dilation invariant if there exist polynomials $c_{m,t} \in \mathbb{Z}[\xi_1, \ldots, \xi_n]$ for $1 \leq m \leq R$, $0 \leq t \leq m$ with $c_{mm} = 1$ for $1 \leq m \leq R$ and such that for any $\xi \in \mathbb{Z}^n$,

$$g_m(x + \xi) = c_{j_0}(\xi) + \sum_{t=1}^{m} c_{m,t}(\xi)g_t(x), \quad 1 \leq m \leq R.$$ (See [PPW13 Eqn (2.3)]) for an explanation of why such systems are called translation-dilation invariant.) The system $G$ is said to be reduced if the set $\{g_1, \ldots, g_R\}$ is linearly independent over $\mathbb{R}$. To avoid degenerate cases, we will only work with $G$ that include all variables nontrivially, and in particular include linear monomials in each variable.

For either the standard system (2.1) or for any reduced monomial translation-dilation invariant system $G$, the following quantities will arise in our proof. Given $G$ as above, we say it has dimension $n$ and rank $R$. Let $\Lambda(G)$ be the associated set of multi-indices, so that $G = \{x^\beta : \beta \in \Lambda(G)\}$. We set the degree $d(G) = \max\{|\beta| : \beta \in \Lambda(G)\}$ to be the highest total degree appearing in a monomial in $G$. The rank is $R(G) = |\Lambda(G)|$ and we define the weight $M(G)$ (or homogeneous dimension) by

$$M(G) = \sum_{\beta \in \Lambda(G)} |\beta|.$$ (2.2)

For the standard system $G$ in (2.1) of monomials in $n$ variables of total degree at most $d$,

$$R = R(G) = \left(\frac{n + d}{n}\right) - 1, \quad M = M(G) = d\left(\frac{n + d}{n}\right)\frac{n}{n + 1}.$$ (2.3)

We define the associated Vinogradov system of $R(G)$ equations in $2r$ variables by

$$x_1^\beta + \cdots + x_i^\beta = x_{r+1}^\beta + \cdots + x_{2r}^\beta, \quad \beta \in \Lambda(G).$$ (2.4)

We let $J_r(G, X)$ denote the number of integral solutions to the system (2.4) with $1 \leq x_{j,i} \leq X$ for $1 \leq j \leq 2r$, $1 \leq i \leq n$. In full generality, our methods prove that $T(F, G; N, H)$ can be controlled by the number of solutions $J_r(G, X)$.

Proposition 2.1. Let $n \geq 2$. Let $q$ be a fixed prime and let $\chi$ be a Dirichlet character of order $\Delta$ modulo $q$. Let $F \in \mathbb{Z}[x_1, \ldots, x_n]$ be a form of degree $D$, with $(\Delta, q)$-admissible reduction modulo $q$. Let $G$ be a reduced monomial translation-dilation invariant system with weight $M(G)$ (containing linear monomials in each variable). For each $r \geq 1$ define $\Theta = \Theta_{n,r} = \frac{\Delta - 1}{\Delta n - 1}$. Then for all integers $r \geq 1$, for all $H = (H, \ldots, H)$ with $H < q^{1/2 + 1/4(\Theta - M(G))}$, as long as $P \leq Hq^{-1/2} \Theta$,

$$T(F, G; N, H) \ll (H/P)^{M(G)/2r} H^{-n/2r} P^{n-1/2r} q^{n/4r} (\log q)^{n+1} \{J_r(G, 2H/P)^{1/2r} + q^{1/4r} (H/P)^{n-\Theta/2r}\}.$$ (2.5)

Thus for any translation-dilation invariant system $G$ for which a suitable bound is known for $J_r(G, X)$, we can deduce a Burgess bound for $|S(F, g_t; N, H)|$, for any polynomial $g$ in the span of $G$. (Given $g$, one could optimize the choice of $G$ as in [Pie16], but we do not pursue this here.)

2.2. Key results for Vinogradov Mean Value Theorems in multi-dimensional settings. Once we have proved Proposition 2.1, it is clear that the key remaining step to prove Theorem 1.2 is to bound $J_r(G, X)$. In dimensions $n \geq 2$, Parsell, Prendiville, and Wooley [PPW13] proved that for any reduced translation-dilation invariant system $G$, for all $r > R(G)(d(G) + 1)$, the sharp upper bound for $J_r(G, X)$ holds. For this range of $r$, the sharp upper bound is $J_r(G, X) \ll_{r,n,d,\varepsilon} X^{2nr - M(G) + \varepsilon}$. Recently, Guo and Zhang [GZ19] have proved the sharp upper bound for $J_r(G, X)$ for the standard system (2.1), for all $n \geq 2, d \geq 1$ and all $r \geq 1$; the exact form of the sharp
and taking the supremum over \( g \). In particular, we note that \( K \chi \) as defined above; to avoid degenerate situations, we assume that \( G \) contains linear monomials in each of the \( n \) variables. In particular, for Theorem 1.2, we can take \( G \) as in \([21]\). We then define

\[
T(F; G; N, H) = \sup_{g \in \mathcal{F}_0(G)} \sup_{K \leq H} \left| \sum_{x \in (N, N + K]} e(g(x)) \chi(F(x)) \right|.
\]

The construction \( T(F; G; N, H) \) (which also appeared in \([15]\)) has several advantageous properties in comparison to a sum \( S(F, g; N, H) \) with a fixed polynomial \( g \). First, \( T(F; G; N, H) \) is periodic under any shift of \( N \) by multiples of \( q \), and thus we will assume from now on that \( 0 \leq N_i < q \) for \( i = 1, \ldots, n \). Second, the fact that \( T(F; G; N, H) \) includes a supremum over polynomials in \( \mathcal{F}_0(G) \) will allow us to replace a supremum over ranges of summation by a supremum over linear phases via Fourier inversion (Lemma 3.2), which in turn is subsumed in the supremum over polynomials in \( \mathcal{F}_0(G) \). Finally, and most crucially, the supremum over \( g \in \mathcal{F}_0(G) \) will allow us to run the Burgess argument including the factor \( e(g(x)) \), as we now demonstrate.

To begin the Burgess argument, we suppose \( H = (H_1, \ldots, H) \) is fixed with \( H < q \). We consider any \( K \leq H \), by which we mean \( K_i \leq H \) for \( i = 1, \ldots, n \). For any tuple \( K \) we denote \( ||K|| = K_1 \cdots K_n \).

For a parameter \( P \) assumed to satisfy \( 1 \leq P \leq H \) we define the set \( \mathcal{P} \) of auxiliary primes by

\[
\mathcal{P} = \{ P < p \leq 2P : p \nmid q \},
\]

so that \( |\mathcal{P}| \gg P / \log P \). We write each \( x \in (N, N + K] \) according to its residue class modulo \( p \), as

\[
x = aq + mp,
\]

where \( a = (a_1, \ldots, a_n) \) with \( 0 \leq a_i < p \) and \( m \in (N^{a,p}, N^{a,p} + K^p) \), with the definitions

\[
N^{a,p} = N/p - aq/p, \quad K^p = K/p.
\]

Then by applying the periodicity and multiplicativity of \( \chi \) and the homogeneity of \( F \), for any \( g \in \mathcal{F}_0(G) \),

\[
\sum_{x \in (N, N + K]} e(g(x)) \chi(F(x)) = \sum_a \sum_{0 \leq a_i < p} \sum_{m \in (N^{a,p}, N^{a,p} + K^p)} e(g(aq + mp)) \chi(F(aq + mp))
\]

\[
= \chi(p^D) \sum_a \sum_{0 \leq a_i < p} \sum_{m \in (N^{a,p}, N^{a,p} + K^p)} e(g(aq + mp)) \chi(F(m)).
\]

In particular, we note that \( K^p \leq K/p \leq H/p \leq H/P \). Consequently, after taking absolute values and taking the supremum over \( g \in \mathcal{F}_0(G) \) and \( K \leq H \), we have

\[
T(F; G; N, H) \leq \sum_a T(F; G; N^{a,p}, H/P).
\]
Finally, we average this inequality over all \( p \in \mathcal{P} \), so that
\[
T(F, G; N, H) \leq |\mathcal{P}|^{-1} \sum_{p \in \mathcal{P}} \sum_{0 \leq a_i < p} T(F, G; N^{a,p}, H/P).
\]

We remark that in our earlier work [Pie16], we restricted to the special case \( F(x_1, \ldots, x_n) = x_1 \cdots x_n \), and we could freely average over distinct sets of primes in each coordinate. Due to averaging over a larger set, we could recover a nontrivial bound for \( H_i \) as small as \( q^{1/4 + \kappa} \) for \( \kappa > 0 \). In our present setting, we can see from the argument above that to exploit the homogeneity of \( F \) we must use the same prime \( p \) for each coordinate, leading to a smaller set to average over. Nevertheless, many of the arguments of [Pie16] may be adapted, and thus we will be efficient in our presentation.

Our next step is to introduce further averaging so that we may free the starting points \( N^{a,p} \) from the dependence on \( a, p \), enabling us to later interchange the order of summation, and apply H"older’s inequality.

**Lemma 3.1.** Fix \( U \in \mathbb{R}^n \) and \( L \in \mathbb{R}_{\geq 1}^n \). For any \( K \leq L \),
\[
T(F, G; U, K) \leq 2^n\|L\|^{-1} \sum_{U - L < m \leq U} T(F, G; m, 2L).
\]

This follows verbatim from the inclusion-exclusion proof given in [Pie16] Lemma 3.1, with each instance of \( \chi_1(x_1) \cdots \chi_n(x_n) \) replaced by \( \chi(F(x)) \), so we do not repeat the proof here. (See also [PX19] Lemma 5.1 for more details on the inclusion-exclusion.)

We apply the lemma to (3.1) with \( L = H/P \) (recalling \( H/P \geq 1 \)) to obtain
\[
T(F, G; N, H) \ll \|H/P\|^{-1}|\mathcal{P}|^{-1} \sum_{p \in \mathcal{P}} \sum_{0 \leq a_i < p} \sum_{N^{a,p} - H/P < m \leq N^{a,p}} T(F, G; m, 2H/P).
\]

Now for each \( m \) we define
\[
\mathcal{A}(m) = \{ p \in \mathcal{P}, a, 0 \leq a_i < p : N^{a,p} - H/P < m \leq N^{a,p} \}.
\]

By Lemma 5.2 of [PX19], \( \mathcal{A}(m) \) vanishes unless \( |m_i| \leq 2q \) for each \( i \), and moreover as long as
\[
HP < q,
\]
then
\[
\sum_m \mathcal{A}(m) \ll \sum_m \mathcal{A}(m)^2 \ll P\|H\|.
\]

Applying H"older’s inequality twice to (3.2) then shows that
\[
T(F, G; N, H) \ll \|H/P\|^{-1}|\mathcal{P}|^{-1} \left( \sum \mathcal{A}(m) \right)^{-1/2r} \left( \sum \mathcal{A}(m)^2 \right)^{1/2r} \left( \sum_{m, |m_i| < 2q} T(F, G; m, 2H/P) \right)^{1/2r},
\]
from which we see after simplification (recalling the periodicity of \( T(F, G; m, H) \) under shifts of \( m \) modulo \( q \)) that
\[
T(F, G; N, H) \ll \|H\|^{-1/2r} P^{n-1/2r} \log P \left( \sum_{m \text{ (mod } q)} T(F, G; m, 2H/P) \right)^{1/2r}.
\]

### 3.1. Strategy to remove the suprema

We required the definition of \( T(F, G; N, H) \) to include two suprema in order to complete the various averaging arguments in the opening steps of the Burgess method, described above. Now we work to remove these suprema, in two steps. We define
\[
T_0(F, G; m, K) = \sup_{g \in \mathcal{F}_0(G)} \left| \sum_{x \in [m, m+K]} e(g(x)) \chi(F(x)) \right| = \sup_{g \in \mathcal{F}_0(G)} \left| \sum_{x \in [0, K]} e(g(x)) \chi(F(x + m)) \right|.
\]
Next, we suppose that we have indexed a finite set of polynomials \( \theta_{\alpha} \in \mathbb{R}[x_1, \ldots, x_n] \) according to a finite set of indices \( \alpha \), and for each such polynomial we define

\[
T_1(F, \theta_{\alpha}; m, K) = \left| \sum_{0 \leq \gamma \leq K} e(\theta_{\alpha}(\mathbf{x})) \chi(F(\mathbf{x} + m)) \right|.
\]

(Note that this is equal to \( S(F, g; m, K) \) with \( g(\cdot) = \theta_{\alpha}(\cdot - m) \), but for technical reasons it is easier to work with the notation \( T_1 \), which builds the shift by \( m \) into the argument of the multiplicative character.) We will pass from expressions involving \( T \) to expressions involving \( T_0 \), and then to expressions involving \( T_1 \). Then we will be ready to evaluate the contribution of the multiplicative character exactly, and to apply Xu’s stratification to bound the contribution of the multiplicative character.

3.2. Approximations of the additive character contribution. We first pass from \( T \) to \( T_0 \) inside (3.4), by applying Lemma 3.3 of [Pie16] (an \( n \)-dimensional version of [BIS86] Lemma 2), which we recall here:

**Lemma 3.2.** Let \( a(n) \) be a sequence of complex numbers indexed by integral tuples \( n \) supported on the set \( n \in (A, A + B) \subset \mathbb{Z}^n \). Let \( I = (C, C + D) \) be any product of intervals with \( I \subseteq (A, A + B) \). Then

\[
\sum_{n \in I} a(n) \ll (\prod_{i=1}^{n} \log(B_i + 2)) \sup_{\theta \in \mathbb{R}^n} \left| \sum_{n \in (A, A + B)} a(n) e(\theta \cdot n) \right|.
\]

This lemma shows that for any \( m \),

\[
T(F, G; m, 2H/P) \ll (\prod_{i=1}^{n} \log(2H_i/P + 2))T_0(F, G; m, 2H/P) \ll (\log q)^n T_0(F, G; m, 2H/P),
\]

since \( 2H_i/P < 2q \). Note that here we use the fact that \( G \) contains linear monomials in each variable. In the setting of Theorem 1.2, we are using the hypothesis that the degree \( d \) of the polynomial \( q \) is at least 1, so that we consider the standard system \( G \) of all monomials in \( n \) variables of total degree at most \( d \).

Applying this in (3.4) proves

\[
(3.5) \quad T(F, G; N, H) \ll \|H\|^{-1/2r} P^{n-1/2r} (\log q)^{n+1} \sum_{m \pmod q} T_0(F, G; m, 2H/P)^{2r} 1/2r.
\]

In order to pass from \( T_0 \) to expressions involving \( T_1 \), we must fix a set of representative polynomials \( \theta_{\alpha} \) (indexed by \( \alpha \)) with the following property: for each \( m \), one of the representative polynomials \( \theta_{\alpha} \) (depending on \( m \)) is such that \( T_1(F, \theta_{\alpha}; m, 2H/P) \) is sufficiently close in value to \( T_0(F, G; m, 2H/P) \).

We define these representative polynomials as in [Pie16] §4, according to a fixed integer \( Q \geq 1 \) (to be chosen later), and \( Q = (Q, \ldots, Q) \). (The outcome here is simplified relative to [Pie16] since all the coordinates of \( Q \) are the same in our setting; in particular we do not require the notion of the “density” \( \gamma(G) \) of the system \( G \), and hence any term of the form \( Q^\gamma \) in [Pie16] we can replace here by \( Q^M \) with \( M = M(G) \) the weight of the system \( G \).) We let \( \Lambda_0(G) = \Lambda(G) \cup \{(0, \ldots, 0)\} \) so \( |\Lambda_0(G)| = R + 1 = R(G) + 1. \) Since within the phase of an exponential sum, the coefficients of a polynomial \( g \) are regarded modulo 1, \( \mathcal{F}_0(G) \) is represented by \( [0,1]^{R+1} \), and upon ordering the \( R + 1 \) multi-indices \( \beta \in \Lambda_0(G) \) in a fixed manner as \( \beta^{(0)}, \ldots, \beta^{(R)} \) once and for all, we partition the \( \beta^{(j)} \)-th unit interval \([0, 1]\) in this product \([0, 1]^{R+1}\) into \( Q^{\beta^{(j)}} = Q^{\beta^{(j)}} \) sub-intervals of length \( Q^{-|\beta^{(j)}|} \). We recall from [Pie16] §4 the following facts about this decomposition. This partitions \([0, 1]^{R+1}\) into \( Q^M \) boxes, which we call \( B_\alpha \), according to indices \( \alpha \) that we order once and for all.
For each such box $B_\alpha$ we assign its distinguished vertex $\theta_\alpha$ to be the vertex with the least value in each coordinate, which is of the form
\[
\theta_\alpha = (\theta_{\alpha,0}, \ldots, \theta_{\alpha,R}) = (c_{\beta(0)} Q^{-|\beta(0)|}, \ldots, c_{\beta(R)} Q^{-|\beta(R)|}),
\]
where $c_{\beta(0)} = 0$ and for each $j = 1, \ldots, R$, $c_{\beta(j)}$ is an integer with $0 \leq c_{\beta(j)} \leq Q^{j(R)} - 1$. Now for any point $\theta \in [0, 1]^{R+1}$, we define an associated real-valued polynomial in $\mathbb{R}[X_1, \ldots, X_n]$ by
\[
\theta(X) := \sum_{\beta \in \Lambda_\alpha} \theta_{\beta} X^\beta.
\]
In particular, for each box $B_\alpha$ with distinguished vertex $\theta_\alpha \in [0, 1]^{R+1}$, we define the associated polynomial $\theta_\alpha(X)$. Finally, we define
\[
S_F(k) := \sum_{\alpha} \sum_{m \equiv \alpha \mod q} T_1(F, \theta_\alpha; m, k)^{2r}.
\]
The following lemma records an upper bound for $T(F, G; N, H)$ in terms of $S_F(k)$, according to this decomposition.

**Lemma 3.3.** Fix any $Q \geq 2H/P$. Let $[0, 1]^{R+1} = \bigcup_\alpha B_\alpha$ be partitioned as described above, according to indices $\alpha$. As long as $HP < q$,
\[
(3.8) \quad T(F, G; N, H) \ll \|H\|^{-1/2r} P^{n-1/2r} (\log q)^{n+1} \sup_{k \leq 2H/P} S_F(k)^{1/2r}.
\]

The proof follows that of [Pie16, Lemma 4.3] verbatim, upon replacing each appearance of $\chi_1(x_1) \cdots \chi_n(x_n)$ by $\chi(F(x))$, and thus we do not repeat it here, although we recall the underlying reasoning. Fix $m$ and consider the corresponding term $T_0(F, G; m, 2H/P)^{2r}$ on the right-hand side of (3.5). Since the coefficients of polynomials $g \in \mathcal{F}_0(G)$ are regarded modulo 1, by compactness the supremum over $g \in \mathcal{F}_0(G)$ in $T_0(F, G; m, 2H/P)$ occurs for a particular polynomial, say $g_0$ (depending on $m$). Under the assumption that $Q \geq 2H/P$, the partition of $[0, 1]^{R+1}$ is sufficiently fine that there exists an index $\alpha$ depending on $m$ such that partial summation shows that replacing $g_0(X)$ by the particular polynomial $\theta_\alpha(X)$ makes a sufficiently small error when replacing $T_0(F, G; m, 2H/P)$ by $T_1(F, \theta_\alpha; m, 2H/P)$. Then, since we do not know which index $\alpha$ was chosen to approximate $g_0(X)$ by $\theta_\alpha(X)$, we sum over all $\alpha$, so that our conclusion holds uniformly in $m$. This leads to the definition of $S_F(k)$. While this seems wasteful, the key observation is that if the partition of $[0, 1]^{R+1}$ is chosen in an arithmetically meaningful way, this sum over $\alpha$ can later be precisely evaluated. This observation occurred first in [HP15] in the case of dimension $n = 1$, and then in [Pie16] in arbitrary dimensions. The precise evaluation of the sum over $\alpha$, which we carry out in the next section, introduces bounds for the number of solutions to a system of Diophantine equations associated to $G$, and the corresponding analogue of the Vinogradov Mean Value Theorem. This leads to a savings that precisely compensates for the loss incurred by summing over all $\alpha$ in the definition of $S_F(k)$.

Finally, we remark on the fact that the right-hand side of (3.8) still contains a supremum, while we claimed our maneuvers aimed to remove the suprema from the objects we were considering. The point is that we will bound $S_F(k)$ by a non-negative function that is increasing in the coordinates of $k$, so that this supremum may be handled quite simply at a later step (see (5.7)).

4. Evaluation of the additive component

We now turn to studying $S_F(k)$ for a fixed $k \leq 2H/P$, beginning by expanding the $2r$-th power. It is convenient to define the following notation. Given $2r$ tuples $x^{(1)}, \ldots, x^{(2r)} \in \mathbb{Z}^n$, we will represent this collection by $\{x\}$. For each $j = 1, \ldots, 2r$, let $\varepsilon(j) = (-1)^{j+1}$ and set $\delta(j) = +1$ if $j$
is odd and $\Delta - 1$ if $j$ is even, where $\Delta$ is the order of $\chi$ modulo $q$. Given such a collection $\{x\}$, we then define

$$\Sigma_{\text{add}}(\{x\}) := \sum_{\alpha} e \left( \sum_{j=1}^{2r} \varepsilon(j) \theta_{\alpha}(x^{(j)}) \right)$$

and

$$\Sigma_{\text{mult}}^{F}(\{x\}) := \sum_{m \pmod{q}} \chi(F_{\{x\}}(m)),$$

in which

$$(4.1)\quad F_{\{x\}}(X) = \prod_{j=1}^{2r} F(X + x^{(j)})^{\delta(j)}.$$

Define $\Xi(G; \{x\})$ to be the indicator function for the set

$$V_{r}(G) := \{x^{(1)}, \ldots, x^{(2r)} \in \mathbb{Z}^n : \sum_{j=1}^{2r} \varepsilon(j)(x^{(j)})^\beta = 0, \forall \beta \in \Lambda(G)\}.$$ 

Later we will use the fact that

$$|V_{r}(G) \cap (0, k)_{2r}| \leq J_r(G, k_{\text{max}}),$$

where $k_{\text{max}} = \max\{k_1, \ldots, k_n\}$, and we recall the notation $J_r(G, X)$ from [2.1].

Evaluating the sum over $\alpha$ in $\Sigma_{\text{add}}(\{x\})$ leads to the following identity, which we will apply with the choice $K = 2H/P$:

**Lemma 4.1.** Let $K = (K_1, \ldots, K)$. As long as $Q = \lfloor 2r K \rfloor$, then for each $k \leq K$,

$$S_F(k) = Q^M \sum_{x^{(1)}, \ldots, x^{(2r)} \in \mathbb{Z}^n : 0 < x^{(j)} < k} \Xi(G; \{x\})\Sigma_{\text{mult}}^{F}(\{x\}).$$

To prove the lemma, expand the $2r$-th power in the definition of $S_F(k)$ to show that

$$(4.3)\quad S_F(k) = \sum_{x^{(1)}, \ldots, x^{(2r)} \in \mathbb{Z}^n : 0 < x^{(j)} < k} \Sigma_{\text{add}}(\{x\})\Sigma_{\text{mult}}^{F}(\{x\}).$$

Now we recall the partition $[0, 1]^{R+1} = \bigcup_0 B_\alpha$ and the distinguished vertices $\theta_\alpha$, which allow us to evaluate precisely the sum $\Sigma_{\text{add}}(\{x\})$ for each fixed collection $\{x\}$. Briefly (as also described in [Pie16 56]), by definition of the distinguished vertices $\theta_\alpha$ in (3.6) and their associated polynomials,

$$\Sigma_{\text{add}}(\{x\}) = \sum_{\alpha} e \left( \sum_{j=1}^{2r} \varepsilon(j) \theta_{\alpha}(x^{(j)}) \right) = \sum_{\varepsilon_{\beta(0)} \ldots , \varepsilon_{\beta(R)}} e \left( \sum_{\beta = \varepsilon_{\beta(0)} \ldots , \varepsilon_{\beta(R)}} c_{\beta} Q^{-\delta} \left( \sum_{j=1}^{2r} \varepsilon(j)(x^{(j)})^\beta \right) \right),$$

where the sum over $c_{\beta(0)}, \ldots, c_{\beta(R)}$ indicates summing for each $i = 0, \ldots, R$ the parameter $c_{\beta(i)}$ over integers $0 \leq c_{\beta(i)} \leq Q^{\beta(i)} - 1$. This can be re-written as

$$\Sigma_{\text{add}}(\{x\}) = \prod_{\beta = \varepsilon_{\beta(0)} \ldots , \varepsilon_{\beta(R)}} \left\{ \sum_{c_{\beta}(\text{mod } Q^\beta)} e \left( c_{\beta} Q^{-\delta} \left( \sum_{j=1}^{2r} \varepsilon(j)(x^{(j)})^\beta \right) \right) \right\},$$

so that by orthogonality of characters, for each multi-index $\beta$ we get a nonzero contribution of $Q^\beta = Q^{\beta}$ if and only if an appropriate congruence holds for $\sum_{j=1}^{2r} \varepsilon(j)(x^{(j)})^\beta$ modulo $Q^\beta$. Precisely,

$$\Sigma_{\text{add}}(\{x\}) = Q^M \Xi_Q(G; \{x\}),$$
where we recall the definition of \( M = M(G) \) from (2.2) and we define \( \Xi_Q(G;\{x\}) \) to be the indicator function for the set

\[
(4.4) \quad \{x^{(1)}, \ldots, x^{(2r)} \in \mathbb{Z}^n : \sum_{j=1}^{2r} \varepsilon(j)(x^{(j)})^\beta \equiv 0 \pmod{Q^\beta}, \forall \beta \in \Lambda(G) \}.
\]

(Note that we do not need to consider a congruence condition for \( Q^\beta \) when \( \beta = (0, \ldots, 0) \), so we can write \( \Lambda(G) \) instead of \( \Lambda_0(G) \) in the definition of this set.) If we choose \( Q \) sufficiently large, we may force the congruences in the definition of this set to be identities in \( \mathbb{Z} \), for every collection \( \{x\} \) such that \( x^{(j)} \in \{0, k\} \), with \( k \leq K \). It suffices to choose

\[
(4.5) \quad Q = \lceil 2rK \rceil.
\]

Indeed, with this choice of \( Q \), we see that for any fixed \( k \leq K \), each congruence in (4.4) can only hold with \( x^{(j)} \in \{0, k\} \) if it holds as an identity in \( \mathbb{Z} \). We conclude that

\[
\Sigma_{\text{add}}(\{x\}) = Q^M \Xi(G;\{x\}),
\]

and this proves Lemma 4.1.

5. Stratification of the multiplicative component

The next step is to count the number of collections \( \{x\} \) for which \( \Sigma_{\text{mult}}^r(\{x\}) \) satisfies certain upper bounds. To do so, we will apply the stratification of Xu [Xu18], in the format of [PX19, Theorem 4.4]. The key result we prove in this section is:

**Proposition 5.1.** Suppose that \( K = (K, \ldots, K) \) with \( K \geq 1 \) and \( Q = \lceil 2rK \rceil \). Define \( \Theta = \Theta_{n,r} = \lfloor (r-1)/(n-1) \rfloor \). Under the assumption \( q^{1/2}K^{-\Theta} \leq 1 \),

\[
\sup_{k \leq K} S_F(k) \ll Q^M \{ J_r(G, K)q^{n/2} + K^{2nr-\Theta}q^ {n/2+1/2} \}.
\]

We remark that the first term in braces may be seen as the contribution of “good” collections \( \{x\} \), namely those that lead to a complete character sum \( \Sigma_{\text{mult}}^r(\{x\}) \) with square-root cancellation; a key result of Xu’s work is that such “good” collections are generic among all tuples in \( \mathbb{Z}^{2nr} \cap \{0, k\}^{2r} \), once \( r \) is sufficiently large that \( \Theta_{n,r} \geq 1 \). This term includes a factor \( J_r(G, K) \) instead of \( K^{2nr} \) because of the advantageous evaluation of the additive character sum, leading to the presence of the indicator function \( \Xi(G;\{x\}) \) in (1.2), which imposes that the collection \( \{x\} \) must lie in the set \( V_r(G) \). The savings here will compensate for the factor \( Q^M \) in front, which we lost by summing over all representatives \( \alpha \) when defining \( S_F(k) \).

The second term in braces is the contribution of the “bad” collections \( \{x\} \). The “bad” collections lead to character sums with bounds ranging from \( O(q^{(n+1)/2}) \) to \( O(q^n) \); Xu’s stratification helpfully shows that “bad” collections have positive codimension in \( \mathbb{Z}^{2nr} \), and the collections that yield progressively worse bounds for the character sum have progressively higher codimension.

Before we prove Proposition 5.1, let us see how it implies Proposition 2.1. We apply the bound from Proposition 5.1 with the choice \( K = 2H/P \) in (3.8), so that \( Q = \lfloor 4rH/P \rfloor \) and so \( Q^M \ll_r (H/P)^M \). We conclude that

\[
T(F, G; N, H) \ll (H/P)^{M/2r} H^{-n/2r} P^{n-1/2r} q^{n/4r} (\log q)^{n+1} \{ J_r(G, 2H/P)^{1/2r} + q^{1/4r} (H/P)^{n-\Theta/2r} \},
\]

as claimed in Proposition 2.1.
5.1. Proof of Proposition 5.1. We first define some notation. Fix \( n \geq 2, r \geq 1 \). Recall that \( \Theta = \Theta_{n,r} = \lceil \frac{r}{n-1} \rceil \). For any \( 1 \leq j \leq n \), and for a tuple \( \mathbf{k} = (k_1, \ldots, k_n) \) with \( k_1 \leq \cdots \leq k_n \), we define

\[
B_{n,r}(j; \mathbf{k}) = B_{n,r}(j; k_1, \ldots, k_n) = \begin{cases} 
1 & j = 0 \\
\frac{k_1^j}{k_1^{j-1}} & j = 1, \ldots, n-2 \\
(k_1 \cdots k_{n/2})^{2r} & j = n, n \text{ even} \\
(k_1 \cdots k_{(n-1)/2})^{2r} k_{(n+1)/2}^j & j = n, n \text{ odd}.
\end{cases}
\]

We now recall [PX19 Thm. 4.4], which is essentially the result of [Xu18], specialized to our setting.

**Theorem A.** Let integers \( n, r, \Delta, D \geq 1 \) be fixed. Then there exist constants \( C = C(n, r, D) \geq 1 \) and \( C'' = C''(n, r, \Delta, D) \geq 1 \) such that the following holds.

Fix a prime \( q \), let \( \chi \) be a multiplicative Dirichlet character of order \( \Delta \) modulo \( q \) and let \( F \in \mathbb{Z}[x_1, \ldots, x_n] \) be a form of degree \( D \) with \( (\Delta, q) \)-admissible reduction modulo \( q \). Define \( F_{(x)}(\mathbf{x}) \) for each collection \( \{x\} \in \mathbb{Z}^{2nr} \) as in (4.4). Then for every \( 1 \leq j \leq n \), for every tuple \( \mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{Z}^n \) with \( 1 \leq k_1 \leq k_2 \leq \cdots \leq k_n \leq q \),

\[
\# \left\{ (x^{(1)}, \ldots, x^{(2r)}) \in (0, \mathbf{k})^{2r} : \sum_{\mathbf{m} \pmod{q}} \chi(F_{(x)}(\mathbf{m})) > Cq^{(n+j-1)/2} \right\} \leq C'' \|\mathbf{k}\|^{2r} B_{n,r}(j; \mathbf{k})^{-1}.
\]

**Remark 5.2.** For more information on the stratification theorem of Xu in this setting, see [PX19 §2, §4], as well as the original work [Xu18]. Roughly speaking, the exponent \( \Theta = \Theta_{n,r} \) arises from a lower bound on the codimension of a subscheme of those collections \( \{x\} = (x^{(1)}, \ldots, x^{(2r)}) \in \mathbb{Z}^{2nr} \) for which square-root cancellation could fail to hold for the complete character sum. The number of collections with a corresponding complete character sum that exceeds square-root cancellation, i.e. \( Cq^{n/2} \), is bounded above by \( C'' \|\mathbf{k}\|^{2r} k_1^{-\Theta} \), the number of collections with a corresponding complete character sum that exceeds \( Cq^{(n+1)/2} \) is bounded above by the smaller quantity \( C'' \|\mathbf{k}\|^{2r} k_1^{-2\Theta} \), and so on. This motivates the definition of the functions \( B_{n,r}(j; \mathbf{k}) \).

Now we prepare to apply the stratification result of Theorem A to \( S_F(\mathbf{k}) \). Let \( C = C(n, r, D) \) be the constant provided by Theorem A. Let us fix \( \mathbf{k} \leq \mathbf{K} \), not yet assuming the ordering \( k_1 \leq \cdots \leq k_n \). For each \( 1 \leq j \leq n \), define

\[
Y_j = Y_j^F(\mathbf{k}) := \left\{ \{x\} \in (0, \mathbf{k})^{2r} : \sum_{\mathbf{m} \pmod{q}} \chi(F_{(x)}(\mathbf{m})) > Cq^{(n+j-1)/2} \right\},
\]

Then \( (0, \mathbf{k})^{2r} =: Y_0 \supseteq Y_1 \supseteq Y_2 \supseteq \cdots \supseteq Y_n \supseteq Y_{n+1} := \emptyset \). Upon employing the disjoint dissection \( (0, \mathbf{k})^{2r} = \bigsqcup_{j=0}^n Y_j \setminus Y_{j+1} \) in (4.2), we now see that for this fixed \( \mathbf{k} \),

\[
S_F(\mathbf{k}) \leq QM \sum_{j=0}^n \sum_{\{x\} \in (Y_j \setminus Y_{j+1}) \cap V_r(\mathbf{G})} |\Sigma_{\text{mult}}^F(\{x\})| \\
\leq QM \sum_{j=0}^n \#(Y_j \cap V_r(\mathbf{G})) Cq^{(n+j-1)/2}. 
\]

Given that the method leading to Theorem A (see [Xu18]) can only compute upper bounds for \#\( Y_j \) in terms of the dimension of \( Y_j \), it is difficult to obtain a nontrivial upper bound for the intersection \( Y_j \cap V_r(\mathbf{G}) \), except in the case of \( Y_0 = (0, \mathbf{k})^{2r} \), in which case

\[
\#(Y_0 \cap V_r(\mathbf{G})) = \#(V_r(\mathbf{G}) \cap (0, \mathbf{k})^{2r}) \leq J_r(\mathbf{G}, k_{\text{max}}).
\]
Thus we obtain

\[ S_F(k) \leq CQ^M J_r(G, k_{\text{max}})q^{n/2} + CQ^n \sum_{j=1}^n \#(Y_j^F(k))q^{(n+j+1)\ell-1}/2. \]

At this point, if in particular \( k_1 \leq k_2 \leq \ldots \leq k_n \leq q \) then we can apply Theorem A in the form of the upper bound \( \#Y_j^F(k) \leq C''\|k\|^{2r}B_{n,r}(j; k)^{-1} \) for each \( 1 \leq j \leq n \), uniformly in \( F \). Consequently in this case we have

\[ S_F(k) \leq CQ^M J_r(G, k_{\text{max}})q^{n/2} + C C''Q^n \sum_{j=1}^n q^{(n+j)\ell/2}k^\ell B_{n,r}(j; k)^{-1}. \]

More generally, given any fixed \( k \), we will re-order the variables \( x_1, \ldots, x_n \) in \( F(x_1, \ldots, x_n) \) so that the correct ordering does hold for the entries in \( k \), and then we will apply Theorem A to a form defined according to this re-ordering. This will use the uniformity of the bound in Theorem A with respect to the form \( F \).

We now give the precise argument. Given any permutation \( \pi \) on \( \{1, \ldots, n\} \), define the form

\[ F^\pi(x_1, \ldots, x_n) = F(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}); \]

\( F^\pi \) has a \((\Delta, q)\)-admissible reduction modulo \( q \) if and only if \( F \) does. Given any \( k \leq K \), let \( \sigma \) be a permutation on the indices \( \{1, \ldots, n\} \) such that

\[ k_{\sigma(1)} \leq k_{\sigma(2)} \leq \cdots \leq k_{\sigma(n)}. \]

Given any tuple \( x \in \mathbb{Z}^n \) let \( x_{\sigma} = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \) and similarly let \( \{x\}_{\sigma} \) denote the collection \( x_{\sigma(1)}, \ldots, x_{\sigma(2r)} \). Since \( F(x_1, \ldots, x_n) = F^{\sigma^{-1}}(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \) we see that for any fixed collection \( \{x\} \),

\[ \sum_{m \pmod{q}} \chi(F^{\sigma^{-1}}_{\{x\}}(m)) = \sum_{m_\sigma \pmod{q}} \chi(F^{\sigma^{-1}}_{\{x\}_\sigma}(m)), \]

Of course, the set of all \( \{x\}_{\sigma} \in (0, k_{\sigma})^{2r} \) is the same as the set of all \( \{x\} \in (0, k)^{2r} \). In particular, if we recall that \( Y_j^F(k) \) denotes the set in \((5.2)\) and let

\[ Y_j^{F^{\sigma^{-1}}}(k_{\sigma}) := \left\{ \{x\}_{\sigma} \in (0, k_{\sigma})^{2r} : \sum_{m_\sigma \pmod{q}} \chi(F^{\sigma^{-1}}_{\{x\}_\sigma}(m)) > C q^{(n+j-1)/2} \right\}, \]

we see that \( Y_j^F(k) = Y_j^{F^{\sigma^{-1}}}(k_{\sigma}) \) for each \( j = 1, \ldots, n \). We will apply this inside each term in \((5.3)\) (which we recall holds for \( k \) without assuming an ordering on the coordinates of \( k \)). We thus obtain from \((5.3)\) that

\[ S_F(k) \leq CQ^M J_r(G, k_{\text{max}})q^{n/2} + CQ^n \sum_{j=1}^n \#(Y_j^{F^{\sigma^{-1}}}(k_{\sigma}))q^{(n+j+1)\ell-1}/2. \]

Now \( k_{\sigma} \) satisfies the ordering \((5.5)\) and thus we may apply Theorem A to bound the cardinality of the sets \( Y_j^{F^{\sigma^{-1}}}(k_{\sigma}) \) and conclude that

\[ S_F(k) \leq CQ^M J_r(G, k_{\text{max}})q^{n/2} + C C''Q^n \sum_{j=1}^n q^{(n+j)\ell/2}k^\ell B_{n,r}(j; k_{\sigma})^{-1}. \]

We observe that in terms of a variable \( k \in \mathbb{R}^n \), \( \|k\|^{2r}B_{n,r}(j; k)^{-1} \) is a non-decreasing function in each coordinate of \( k \); that is, for fixed \( r \geq 1 \), for each \( 1 \leq j \leq n \), there exist exponents \( \alpha_{j,1}, \ldots, \alpha_{j,n} \geq 0 \) (also depending on \( r \)) such that

\[ \|k\|^{2r}B_{n,r}(j; k)^{-1} = k_1^{\alpha_{j,1}} \cdots k_n^{\alpha_{j,n}}. \]
for all tuples \( k \) with \( k_1 \leq \cdots \leq k_n \). This is an immediate consequence of the definition of the functions \( B_{n,r}(j; \cdot) \). In particular, for any \( k \leq K \) with \( k_1 \leq \cdots \leq k_n \), where \( K = (K, \ldots, K) \), we obtain that

\[
\|k\|^{2r} B_{n,r}(j; k)^{-1} \leq K_{\alpha_1}^{n} \cdots K_{\alpha_n}^{n} = \|K\|^{2r} B_{n,r}(j; K)^{-1}.
\]

We apply this to each term in (5.7), and we conclude that

\[
\text{Proposition 5.1 will now follow, after we verify a lemma about sums of the functions } B_{n,r}(j; \cdot), \text{ which we prove in a general setting.}
\]

**Lemma 5.3.** If \( 1 \leq K_1 \leq K_2 \leq \cdots \leq K_n \) then for each \( r \geq 1 \),

\[
\sum_{j=1}^{n} q^{j/2} B_{n,r}(j; K)^{-1} \ll q^{1/2} K_1^{-\Theta}
\]

as long as

\[
q^{1/2} K_1^{-\Theta} \leq 1.
\]

By definition, the sum over \( j = 1, \ldots, n-1 \) takes the form

\[
\sum_{j=1}^{n-2} q^{j/2} K_2^{-j\Theta} + q^{(n-1)/2} K_1^{-(r-1)} \leq \sum_{j=1}^{n-1} (q^{1/2} K_1^{-\Theta})^j
\]

in which \( \Theta = \lfloor (r-1)/(n-1) \rfloor \); this used the fact that \( r-1 \geq (n-1) \lfloor (r-1)/(n-1) \rfloor = (n-1)\Theta \). The right-most expression shows that under the assumption (5.9), all terms \( j \geq 2 \) are dominated by \( j = 1 \). (On the other hand, the expression increases in size if \( q^{1/2} K_1^{-\Theta} > 1 \), and this will motivate our later choice of \( P \) so that (5.9) holds, with \( K = 2H/P \).) The last term to check is \( j = n \). For \( n \geq 2 \) even, the \( j = n \) term is

\[
q^{n/2} (K_1 \cdots K_{n/2})^{-2r} \leq q^{n/2} K_1^{-nr} \leq (q^{1/2} K_1^{-\Theta})^n,
\]

in which we have used the ordering \( K_1 \leq K_2 \leq \cdots \leq K_n \) and the fact that \( r \geq (r-1)/(n-1) \geq \Theta \) when \( n \geq 2 \). For \( n \geq 3 \) odd, similar reasoning shows the \( j = n \) term is

\[
q^{n/2} (K_1 \cdots K_{(n-1)/2})^{-2r} K_{(n+1)/2}^{-r} \leq q^{n/2} K_1^{-rn} \leq (q^{1/2} K_1^{-\Theta})^n.
\]

In either case, under the assumption (5.9) we see that the \( j = n \) term is \( \ll q^{1/2} K_1^{-\Theta} \). This completes the proof of the lemma, verifying Proposition 5.1 and hence Proposition 2.1.

6. **Concluding arguments for Theorem 1.2**

With Proposition 2.1 in hand, the final steps to prove Theorem 1.2 are to apply a bound for \( J_r(G,X) \) and to choose \( P \). In order to motivate our choice for \( P \leq H \), we recall that so far we have supposed in (5.3) and the application of Proposition 5.1 with \( K = 2H/P \) that

\[
HP < q, \quad P \leq Hq^{-1/2\Theta}.
\]

We first argue formally, in some generality. We suppose we are in a range of \( r \) where

\[
J_r(G,X) \ll X^{2r \mu + \varepsilon},
\]

for some positive integer \( \mu \) (depending on \( n, d, r \) and the system \( G \)). (As we remark below, this is known for \( \mu = M \) for all values of \( r \) that we will consider, but for later reference we first argue more generally.) Under these assumptions,

\[
J_r(G, 2H/P)^{1/2r} + (H/P)^{n-\Theta/2r} q^{1/4r} \ll q^\varepsilon ((H/P)^{n-\mu/2r} + (H/P)^{n-\Theta/2r} q^{1/4r}).
\]
Now we observe that to balance these two terms we would choose $P$ to be an integer with

$$
\frac{1}{2} Hq^{-\frac{1}{2(\Theta - \mu)}} \leq P < Hq^{-\frac{1}{2(\Theta - \mu)}}. 
$$

This supposes that $\Theta > \mu$ in order to meet the requirement that $P \leq H$; since $\Theta = |(r-1)/(n-1)|$, this is a requirement that $r$ is sufficiently large with respect to $\mu, n$. This choice for $P$ also satisfies the requirements in (6.5), as long as we assume that $H < q^{\frac{2}{\mu} + \frac{1}{4(\Theta - \mu)}}$. (This will be satisfied, by a hypothesis of the theorem, when we ultimately apply this reasoning with $\mu = M = M(G)$.) We apply this choice of $P$ in Proposition 2.1 to conclude that if (6.2) holds then

$$
T(F, G; N, H) \ll H^{n-\frac{\mu+1}{2r}} q^{\frac{n(\Theta - M) + 1 - \mu}{2r(\Theta - M)}} q^\varepsilon,
$$

for any $\varepsilon > 0$, as long as $\Theta = |(r-1)/(n-1)| > \mu$. This is a condition on $r$, namely $r > \mu + 1$ when $n = 2$, and in general it suffices to have $r > (\mu + 1)(n - 1) + 1$ when $n \geq 3$.

Now to understand $\mu$ in (6.2), we restrict our attention to $G$ being the standard system (2.1) for dimension $n$ and degree $d$. (We remark on more general systems in (7.2) in an appendix.) One can calculate that in order for $\mu, r$ to be such that (6.2) holds and simultaneously $\Theta = |(r-1)/(n-1)| > \mu$, we must be in the range of $r$ such that the savings in (6.2) is $\mu = M(G)$. (We provide the details to prove this simple observation in the appendix.) This comes from the known upper bounds in the multi-dimensional Vinogradov Mean Value Theorem, which we now recall. Precisely, for all $n \geq 2$, (6.2) with $\mu = M$ is true for all values of $r$ satisfying

$$
r > M + 1 \quad (n = 2), \quad r > (M + 1)(n - 1) + 1 \quad (n \geq 3),
$$

due to the truth of the Vinogradov Mean Value Theorem for the system $G$ defined in (2.1). To be precise, for $n \geq 3$, $J_r(G, X) \ll X^{2nr-M+\varepsilon}$ for $r$ in the range (5.5) is known from [PPW13]. (This uses the fact that for $n \geq 3$, the requirement on $r$ in (6.5) imposes that $r > R(d + 1)$, which was the requirement in the work [PPW13].) On the other hand, for $n = 2$, for each $d \geq 2$, in order to obtain this upper bound for $r$ in the range (6.5) one requires the stronger results of [GZ19], which apply for all $r \geq 1$.

Thus we now only consider the case that (6.2) holds with $\mu = M = M(G)$ and $r$ is in the range (6.5). When both these conditions are met, (6.1) shows that

$$
T(F, G; N, H) \ll H^{n-\frac{\mu+1}{2r}} q^{\frac{n(\Theta - M) + 1 - \mu}{2r(\Theta - M)}} q^\varepsilon.
$$

This suffices to complete the proof of Theorem 1.2 since $|S(F, g; N, H)| \leq T(F, G; N, H)$.

### 6.1. Quantification of the strength of Theorem 1.2

Supposing that $H = q^\beta$, then the bound for $|S(F, g; N, H)|$ provided by Theorem 1.2 is nontrivial i.e. $o(q^\beta)$ as long as

$$
\beta > \frac{1}{2} - \frac{(\Theta - M - 1)}{2(\Theta - M)(n + 1)},
$$

in which $\Theta = |(r-1)/(n-1)|$. This allows for values of $\beta$ strictly smaller than $1/2$ as long as $r$ is sufficiently large that $\Theta > M + 1$. The right-hand side in (6.7) is always $> \beta_n$ with $\beta_n$ as defined in (1.3). We thus suppose that $H = q^{\beta_n + \kappa}$ for some small $\kappa > 0$, in which case we can compute that $|S(F, g; N, H)| \ll H^{\alpha} q^{-\delta}$ with

$$
\delta = \frac{2\kappa(n + 1)(\Theta - M) - 1}{4r(\Theta - M)}.
$$

We now use the approximation of replacing $\Theta$ by $(r-1)/(n-1)$ (which in fact is exact, when $n = 2$, and not far off from the truth when $r$ grows very large, as it will when we choose $r$ according to $\kappa$ and $\kappa \to 0$). After this approximation, we can write $\delta$ as the value at $r$ of the function

$$
f_{b,c,d}(r) = \frac{br - c}{r(r - d)}.
$$
with
\[ b = \frac{\kappa(n + 1)}{2}, \quad c = (M(n - 1) + 1) \frac{\kappa(n + 1)}{2} + \frac{n - 1}{4}, \quad d = M(n - 1) + 1. \]

The function \( f_{b,c,d}(r) \) attains a local extremum at \( r = b^{-1}(c \pm \sqrt{c^2 - abc}) \); using the values for \( b, c, d \) above and simplifying using \( \kappa \to 0 \) we see that we should choose \( r \) to be the nearest integer to
\[ r \approx \frac{n - 1}{n + 1} \cdot \frac{1}{\kappa}. \]

This choice of \( r \) satisfies \( \Theta = \Theta_{n,r} > M \) if \( \kappa \) is sufficiently small (relative to \( n, d \)). We now apply this in the expression above for \( \delta \), now further approximating \( \Theta \) by \( r/(n - 1) \), and we see that in the limit as \( \kappa \to 0 \) we obtain a savings over \( H^n \) of the form \( H^n q^{-\delta} \), in which
\[ \delta \approx \frac{(n + 1)^2}{4(n - 1)} \kappa^2. \]

The significance of this savings is that it is independent of the degree \( d \) of \( g \), due to the application of the multi-dimensional Vinogradov Mean Value Theorem. In particular, it is as strong as the savings of the first author and Xu \([PX19]\) in the purely multiplicative case.

7. Appendix: Further remarks on Vinogradov systems

In this appendix, we briefly remark on three aspects of the proof of Theorem 1.2. First, we consider how improvements to \( \Theta \) (recall Remark 5.2) would lead to a setting in which one would require the sharp results of \([GZ19]\) for all \( n \geq 3 \), in addition to the results of \([PPW13]\). Second, we explain why we only considered \( \mu = M \) in the conclusion of the proof of Theorem 1.2 or equivalently, why the current Burgess method in this setting leads to consideration of very large \( r \). Third, we briefly state a more general result for systems \( G \) other than the standard system \((2.1)\).

7.1. Remarks on the codimension \( \Theta \). Let \( G \) denote the standard system \((2.1)\) of monomials in \( n \) variables of degree at most \( d \). In the proof of Theorem 1.2 we applied the results of \([PPW13]\) to bound \( J_r(G,X) \) when \( n \geq 3 \), and only required the stronger results of \([GZ19]\) when \( n = 2 \). We now remark that if one could improve the stratification of Xu for complete multiplicative character sums (in the sense of the discussion in \([PX19]\ §8.2)\), then one would require the results of \([GZ19]\) for all \( n \geq 2 \).

Precisely, we have seen in the argument in \([4]\) that if the known upper bound is \( J_r(G,X) \leq X^{2\mu r - \mu + \varepsilon} \), then the result of the Burgess method developed in this paper must restrict to values of \( r \) for which \( \Theta > \mu \). Here \( \Theta \) is the codimension of the first exceptional subscheme \( X_1 \) arising in the stratification of Xu \([Xu18]\); see Remark 5.2 for a rough idea. Currently, for \( n \geq 2 \), Xu has obtained \( \Theta \geq [(r - 1)/(n - 1)] \). The appendix in \([PX19]\ §8.2) outlines conjectural possibilities for improvements to the codimension leading to the value for \( \Theta \); for example, one might hope to prove that \( \Theta = r \) is possible. In dimension \( n = 2 \) this is nearly attained already by \( \Theta = [(r - 1)/(n - 1)] \), but is significantly different from the current result for large \( n \). Let us suppose that one could prove \( \Theta \geq r/\alpha(n) \) for some function \( 1 \leq \alpha(n) \leq n - 1 \), leading to the restriction \( r > \mu \alpha(n) \) in the method of proof for Theorem 1.2. In particular, if we use \( \mu = M \) and \( \alpha(n) \) is not too large, then in order to obtain our theorem unconditionally for all \( r > M \alpha(n) \) we would require the results of \([GZ19]\) for \( J_r(G,X) \) for those \( r \) with \( M \alpha(n) < r < R(d + 1) \), while \([PPW13]\) would continue to apply for \( r > R(d + 1) \).

7.2. Remarks on intermediate ranges of \( r \). In the proof of Theorem 1.2 we remarked that we need only consider the upper bound \((6.2)\) when \( \mu = M \), where \( M = M(G) \) is the weight of the associated system of Diophantine equations (the sum of the total degrees), and \( r \) is very large. One might ask whether one could consider other values for \( \mu \), and correspondingly smaller values \( r \). Here we explain why the Burgess method developed in this paper only allows the regime of \( r \) in which \( \mu = M = M(G) \).
In the current discussion we can take $G$ to be any reduced monomial translation-dilation invariant system. Precisely, the question is: what must $\mu$ be in order for both (6.2) and $\Theta = \lfloor (r-1)/(n-1) \rfloor > \mu$ to hold? Given any reduced monomial translation-invariant system $G$, suppose there is a sequence of positive integers $K_j = K_j(G)$ for $1 \leq j \leq n$ such that for all $X \geq 1$, for all $r \geq 1$,

\begin{equation}
J_r(G, X) \ll n \cdot \sum_{j=1}^{n} X^{2rj+(n-j)K_j}.
\end{equation}

(Note that in this notation, $K_n$ plays the role of $M(G).$) In particular, by the breakthrough work of Guo and Zhang [GZ19], this is now known for the standard system $G$ in (2.1) of monomials in $n$ variables with total degree at most $d$; in this case $K_j = \frac{j}{j+1}(\frac{d}{j} + d).

We claim that if a bound of the form (7.1) holds, then in order for both (6.2) and $\Theta = \lfloor (r-1)/(n-1) \rfloor > \mu$ to hold (and hence certainly $r > \mu$), we must have $\mu = K_n$ (so that the $j = n$ term dominates in (7.1)). Indeed, suppose that $r$ is such that the $j$-th term dominates in (7.1), for some $1 \leq j \leq n$. In the notation of (6.2), this would impose $\mu = (2r-1)(n-j) + K_j$. Then in order to have $r > \mu$ we must at least have $r > (2r-1)(n-j)$, which can only hold if $n = j$. (Similarly, the term $X^{rn}$ cannot dominate, since that would impose $\mu = rn$, but the condition $r > \mu$ could not hold.) This proves the claim. (Even if, for example, the codimension $\Theta$ could be improved to $r$, the analogue of (6.3) would still require $r > \mu$, leading to $\mu = M(G)$ via the same argument given above.) Thus it appears that significant innovations to the method would be required, in order to be able to apply counts for Vinogradov systems where any term with $j < n$ dominates in (7.1).

7.3. Remarks on other systems. Let $G$ be any reduced monomial translation-dilation invariant system, in any dimension $n \geq 2$ and with degree $d(G) \geq 1$. Parsell, Prendiville and Wooley proved that (7.1) holds for any $r > R(G)(d(G) + 1)$, in which case the $j = n$ term dominates (with $K_n(G) = M(G)$), and the upper bound is $J_r(G, X) \ll X^{2rn-M(G)+\epsilon}$. More recently, Guo and Zorin-Kranich [GZK20] have proved that a sharp upper bound of the form (7.1), with appropriately defined $K_j(G)$, holds for all $r \geq 1$, for more general systems $G$, which we now describe. Fix a tuple $(k_1, \ldots, k_n)$ of positive integers. Fix an integer $k$. Let $G$ be the system defined according to the set of exponents

\begin{equation}
\Lambda(G) = \{ \beta : \beta_1 \leq k_1, \ldots, \beta_n \leq k_n, 1 \leq |\beta| \leq k \}.
\end{equation}

If in particular $k = k_1 = \cdots = k_n = d$ then this is the standard system (2.1). If $k = k_1 + \cdots + k_n$ then this is known as an Arkhipov-Clucharikov-Karatsuba system. For the systems Guo and Zorin-Kranich handle we can thus obtain a generalization of Theorem 7.1 in the largest range of $r$ allowed by the Burgess method developed in this paper. We record the conclusion of this discussion:

**Theorem 7.1.** Fix $n \geq 2$ and $d, D \geq 1$. Let $q$ be a fixed prime, and let $\chi$ be a non-principal character of order $\Delta$ and conductor $q$. Let $F \in \mathbb{Z}[x_1, \ldots, x_n]$ be a form of degree $D$ such that its reduction modulo $q$ is $(\Delta, q)$-admissible. Let $G$ be a reduced monomial translation-dilation invariant system with rank $R(G)$ and weight $M(G)$ (containing linear monomials in each variable). For each integer $r \geq 1$, define $\Theta = \lfloor (r-1)/(n-1) \rfloor$. Then

\begin{equation}
|T(F, G; N, H)| \ll H^{n-\frac{n+1}{2r}} q^{\frac{n(\Theta-M(G))}{4r(\Theta-M(G))}+1} q^{\frac{\epsilon}{2}},
\end{equation}

for every integer $r$ such that $r > R(G)(d(G) + 1)$ and $\Theta > M(G)$, and for every $H = (H, \ldots, H)$ with $H < q^{1/2+1/4(\Theta-M(G))}$. Furthermore if $G$ is a system of the type (7.2) then we may take any $r$ such that $\Theta > M(G)$. The implied constant could depend on $G, n, D, \Delta, r, \epsilon$ but is otherwise independent of $F$. 
Acknowledgements

Pierce is partially supported by NSF CAREER grant DMS-1652173, a Sloan Research Fellowship, and the AMS Joan and Joseph Birman Fellowship. Pierce thanks the Hausdorff Center for Mathematics and the initiative “A Room of One’s Own” for focused time. Pierce also thanks the nine individuals who helped to tend her children during the main time period in which this work took place.

References

[BD16] J. Bourgain and C. Demeter. Mean value estimates for Weyl sums in two dimensions. *J. Lond. Math. Soc. (2)*, 94(3):814–838, 2016.
[BDG16] J. Bourgain, C. Demeter, and L. Guth. Proof of the main conjecture in Vinogradov’s mean value theorem for degrees higher than three. *Ann. of Math. (2)*, 184(2):633–682, 2016.
[BDG17] J. Bourgain, C. Demeter, and S. Guo. Sharp bounds for the cubic Parsell-Vinogradov system in two dimensions. *Adv. Math.*, 320:827–875, 2017.
[BI86] E. Bombieri and H. Iwaniec. On the order of \(\zeta(1/2+it)\). *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 13:449–472, 1986.
[Bur57] D. A. Burgess. The distribution of quadratic residues and non-residues. *Mathematika*, 4:106–112, 1957.
[Bur63] D. A. Burgess. On character sums and \(L\)-series II. *Proc. London Math. Soc.*, 3:524–536, 1963.
[GZ19] S. Guo and R. Zhang. On integer solutions of Parsell-Vinogradov systems. *Invent. Math.*, 218:1–81, 2019.
[GZK20] S. Guo and P. Zorin-Kranich. Decoupling for moment manifolds associated to Arkhipov-Chubarikov-Karatsuba systems. *Adv. Math.*, 360:(to appear), 2020.
[HBP15] D. R. Heath-Brown and L. B. Pierce. Burgess bounds for short mixed character sums. *J. LMS*, 91:693–708, 2015.
[Ker14] B. Kerr. Some mixed character sums. *arXiv:1410.3587*, 2014.
[Pie16] L. B. Pierce. Burgess bounds for multi-dimensional short mixed character sums. *J. Number Theory*, 163:172–210, 2016.
[PPW13] S. Parsell, S. M. Prendiville, and T. D. Wooley. Near-optimal mean value estimates for multidimensional Weyl sums. *Geom. Funct. Anal.*, 23:1962–2024, 2013.
[PX19] L. B. Pierce and J. Xu. Burgess bounds for short character sums evaluated at forms. *arXiv:1907.03108*, 2019.
[PY19] I. Petrow and M. P. Young. The fourth moment of Dirichlet \(L\)-functions along a coset and the Weyl bound. *arXiv:1908.10346*, 2019.
[Woo16] T. D. Wooley. The cubic case of the main conjecture in Vinogradov’s mean value theorem. *Adv. Math.*, 294:532–561, 2016.
[Woo19] Trevor D. Wooley. Nested efficient congruencing and relatives of Vinogradov’s mean value theorem. *Proc. Lond. Math. Soc. (3)*, 118(4):942–1016, 2019.
[Xu18] J. Xu. Stratification for Multiplicative Character Sums. *International Mathematics Research Notices*, 05 2018. rny096.

Department of Mathematics, Duke University, 120 Science Drive, Durham NC 27708 USA
E-mail address: pierce@math.duke.edu