Algebra & Number Theory

Volume 14
2020
No. 7

Burgess bounds for short character sums evaluated at forms

Lillian B. Pierce and Junyan Xu

msp

ALGEBRA AND NUMBER THEORY 14:7 (2020)
https://doi.org/10.2140/ant.2020.14.1911

Burgess bounds for short character sums evaluated at forms

Lillian B. Pierce and Junyan Xu

We establish a Burgess bound for short multiplicative character sums in arbitrary dimensions, in which the character is evaluated at a homogeneous form that belongs to a very general class of "admissible" forms. This n-dimensional Burgess bound is nontrivial for sums over boxes of sidelength at least q^β, with $\beta > 1/2 - 1/(2(n+1))$. This is the first Burgess bound that applies in all dimensions to generic forms of arbitrary degree. Our approach capitalizes on a recent stratification result for complete multiplicative character sums evaluated at rational functions, due to the second author.

1. Introduction

The celebrated Burgess bound [1957] proves that for χ a nonprincipal multiplicative Dirichlet character modulo a prime q, the character sum

$$S(N, H) = \sum_{\substack{x \in \mathbb{Z} \\ N < x \le N+H}} \chi(x)$$

is bounded for every integer $r \ge 1$ by

$$|S(N, H)| \ll_r H^{1-1/r} q^{(r+1)/4r^2} \log q. \tag{1-1}$$

From this it can be deduced that $S(N, H)$ admits a nontrivial bound $o(H)$ for H as small as $H = q^{1/4+\kappa}$, for any $\kappa > 0$. Bounds for $S(N, H)$ have many applications, and as we survey in Section 1C, Burgess's influential work set records that remain the best known today.

This paper proves the first n-dimensional Burgess bound for short multiplicative character sums evaluated at generic homogeneous polynomial arguments of arbitrarily large degree. Let χ be a nonprincipal multiplicative character modulo a prime q. Let $n \ge 1$ be a fixed dimension, and $F \in \mathbb{Z}[x_1, \dots, x_n]$ a form of degree D. For any $\boldsymbol{N} = (N_1, \dots, N_n) \in \mathbb{R}^n$, $\boldsymbol{H} = (H_1, \dots, H_n) \in \mathbb{R}^n_{\ge 1}$, define

$$S(F; \boldsymbol{N}, \boldsymbol{H}) = \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^n \\ x_i \in (N_i, N_i + H_i]}} \chi(F(\boldsymbol{x})). \tag{1-2}$$

Given $\boldsymbol{H}$, we will define $\|\boldsymbol{H}\| = H_1 \cdots H_n$, so that in particular, $\|\boldsymbol{H}\|$ is a trivial bound for $|S(F; \boldsymbol{N}, \boldsymbol{H})|$. Previous to the work of this paper, when the lengths H_i are short, that is $\ll q^{1/2+\varepsilon}$, nontrivial bounds for

MSC2010: 11L40.
Keywords: character sums.

$S(F; \boldsymbol{N}, \boldsymbol{H})$ of the form $o(\|\boldsymbol{H}\|)$ were known only in special cases, such as when F is a product of n linear forms that are linearly independent over $\mathbb{F}_q$, or when $n = 2$ and F is a binary quadratic form (see Section 1C2 for details). In this paper, we prove nontrivial bounds for $S(F; \boldsymbol{N}, \boldsymbol{H})$ in any dimension n for a very general class of "admissible" forms F, as long as

$$\|\boldsymbol{H}\| H_{\min} \gg q^{n/2+\kappa},$$

for some $\kappa > 0$, where $H_{\min} = \min_i H_i$. In particular, this is satisfied when $\boldsymbol{H} = (H, \dots, H)$ with $H = q^{\beta_n+\kappa}$ for any $\kappa > 0$, where

$$\beta_n = \frac{1}{2} - \frac{1}{2(n+1)}. \tag{1-3}$$

1A. ***Statement of the main theorem.*** We now provide a formal statement of the condition that a form F must satisfy in order to be "admissible" for our main result. We only need to rule out those forms F for which a nontrivial bound for $S(F; \boldsymbol{N}, \boldsymbol{H})$ would naively fail, such as when F is a perfect Δ-th power and χ is order Δ, or when F can be made to depend on fewer than n variables.

Condition ((Δ, q)-admissible). Let q be a fixed prime and $\Delta \geq 1$ a fixed integer. We will say that a polynomial $f \in \mathbb{F}_q[x_1, \dots, x_n]$ is (Δ, q)-admissible if the following holds. Factorize $f = g^{\Delta}h$, where $g, h \in \mathbb{F}_q[x_1, \dots, x_n]$ and h is Δ-th power-free over $\mathbb{F}_q$. Then h has the property that it cannot be made independent of (at least) one variable after a linear transformation, i.e., there exists no linear change of variables $A \in \mathrm{GL}_n(\mathbb{F}_q)$ such that $h(\boldsymbol{x}A) \in \mathbb{F}_q[x_2, \dots, x_n]$.

See Section 3 for further details on this condition, and a precise definition of being Δ-th power-free. If $\Delta \geq 2$ is a fixed integer, any form $F \in \mathbb{Z}[x_1, \dots, x_n]$ such that $F = G^{\Delta}H$ with $G, H \in \mathbb{Z}[x_1, \dots, x_n]$ where H is Δ-th power-free and nondegenerate with respect to changes of variables in $\mathrm{GL}_n(\mathbb{Z})$, has the property that its reduction modulo q is (Δ, q)-admissible for all but finitely many primes q; see Lemma 3.4. For any D, the form $x_1^D + \cdots + x_n^D$ is an example of such a form. Moreover, such forms are generic among all forms in $\mathbb{Z}[x_1, \dots, x_n]$ of degree at most D, since those that violate the conditions depend on fewer parameters.

Our main result is as follows:

Theorem 1.1. *Let χ be a nonprincipal multiplicative Dirichlet character of order Δ modulo a prime q. Let $n \geq 2$ be fixed. For each $r \geq 1$, define*

$$\Theta = \Theta_{n,r} = \left\lfloor \frac{r-1}{n-1} \right\rfloor.$$

Let $\boldsymbol{H} = (H_1, \dots, H_n) \in \mathbb{R}^n_{\geq 1}$ have maximum element $H_{\max}$ and minimum element $H_{\min}$, and assume $H_{\max}H_{\min} < q^{1+1/(2\Theta)}$. For every degree D form $F \in \mathbb{Z}[x_1, \dots, x_n]$ such that its reduction modulo q is (Δ, q)-admissible, then uniformly in $\boldsymbol{N} = (N_1, \dots, N_n)$, for every integer $r \geq 1$,

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll \|\boldsymbol{H}\|^{1-(1/2r)} H_{\min}^{-1/2r} q^{(n\Theta+1)/4r\Theta} (\log q)^{n+1}, \tag{1-4}$$

in which the implied constant depends only on D, Δ, n, r and is otherwise independent of F.

If one carries through our method of proof in the case $n = 1$, we may define $\Theta = \Theta_{1,r} = r$ for all $r \geq 1$ and recover the Burgess bound (1-1) (up to the power of the logarithm). This result may also be extended to apply to a rational function $F = f_1/f_2$ if it is appropriately regarded as $f_1 f_2^{\Delta-1}$. Analogous to the proof of the Pólya–Vinogradov inequality, Fourier-based methods can prove a nontrivial bound for $S(F; \boldsymbol{N}, \boldsymbol{H})$ for suitable forms F when $H_i \gg q^{1/2+\varepsilon}$ (see Section A.3); thus the upper bound restriction on H_i in the hypothesis of the theorem is compatible with our interest in the range $H_i \ll q^{1/2+\varepsilon}$.

For purposes of comparison, we state a direct corollary of Theorem 1.1 in the case that all coordinates of $\boldsymbol{H}$ are of equal size.

Corollary 1.2. *Assume the hypotheses of Theorem 1.1. Suppose $\boldsymbol{H} = (H, H, \dots, H)$ with $H < q^{1/2+1/(4\Theta)}$. For every degree D form $F \in \mathbb{Z}[x_1, \dots, x_n]$ such that its reduction modulo q is (Δ, q)-admissible, then uniformly in $\boldsymbol{N} = (N_1, \dots, N_n)$, for every integer $r \geq 1$,*

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll H^{n-(n+1)/2r} q^{(n\Theta+1)/4r\Theta} (\log q)^{n+1}, \tag{1-5}$$

in which the implied constant depends only on D, Δ, n, r and is otherwise independent of F.

Remark. The upper bound in Burgess's original work is independent of the order of the character. In our results, there is a possible dependence on the order Δ. In our approach, this possible dependence on Δ appears in the constant C'' provided by Theorem 4.4, which we apply to count the number of tuples $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})$ that could result in "large" complete character sum values. In turn, the constant C'' provided by Theorem 4.4 can depend on Δ because it encodes the sum of the degrees of the irreducible components of certain subschemes, that are determined by n, r, D and Δ. See Section 4 and [Xu 2018, Section 2.1] for details.

1B. ***The strength of Theorem 1.1: quantifications.*** In general, for a Burgess-style result such as (1-1), which holds for a range of integers r, to assess its strength for H near the lower-bound threshold that yields a nontrivial bound for $|S(N, H)|$, we must compute which value of r produces maximum savings. For example, in Burgess's original result, if $H = q^{1/4+\kappa}$ the bound (1-1) with parameter value r yields the upper bound $|S(N, H)| \ll Hq^{-\delta}$ where $\delta = (4\kappa r - 1)/(4r^2)$. Computing the maximum of δ with respect to r, we see that by choosing r to be the nearest integer to $1/(2\kappa)$ we may obtain the best value $\delta \approx \kappa^2$.

We perform an analogous optimization of our result, summarized in two corollaries.

Corollary 1.3. *For each $n \geq 2$ and $r \geq 1$, Theorem 1.1 with the parameters n, r provides a nontrivial upper bound $|S(F; \boldsymbol{N}, \boldsymbol{H})| = o_{n,r,\Delta,D}(\|\boldsymbol{H}\|)$ for all $\boldsymbol{H} = (H, \dots, H)$ with $H = q^\beta$ with β in the range*

$$\frac{1}{2} - \frac{\Theta - 1}{2\Theta(n+1)} < \beta \leq \frac{1}{2} + \frac{1}{4\Theta}, \tag{1-6}$$

in which $\Theta = \Theta_{n,r} = \lfloor (r-1)/(n-1) \rfloor$. This range includes a nonempty interval of $\beta < \frac{1}{2}$ as soon as $r \geq 2n - 1$, so that $\Theta = \Theta_{n,r} > 1$. In particular, this range always requires $\beta > \beta_n$ with

$$\beta_n := \frac{1}{2} - \frac{1}{2(n+1)}. \tag{1-7}$$

For $H = q^{\beta_n+\kappa}$ for small κ, we obtain a nontrivial bound $\|\boldsymbol{H}\|q^{-\delta_n}$ with savings approximately of size

$$\delta_n \approx \frac{(n+1)^2}{4(n-1)}\kappa^2$$

as $\kappa \to 0$.

Remark. In the approximate value of the savings δ_n stated above (and similarly in Corollary 1.4 below), the symbol $\approx$ is used to denote that certain approximations were used in the derivation, such as (i) letting the value $\theta_1 = \lfloor (r-1)/(n-1) \rfloor$ be approximated by $r/(n-1)$, which is reasonable, for fixed n, for the very large values of $r = r(\kappa)$ that are chosen when $\kappa \to 0$; and (ii) considering the limiting case $\kappa \to 0$ in order to isolate the main term in computations. More details are given in Section 7A.

The threshold β_n defined in (1-7) has appeared in n-dimensional Burgess bounds that were previously proved in very special cases, such as [Burgess 1967b; 1968; Davenport and Lewis 1963] (see Section 1C2 for an overview, including the stronger results [Chang 2008; 2009; Bourgain and Chang 2010; Heath-Brown 2016]).

In full generality, our main result Theorem 1.1 is in fact stronger than Corollary 1.3, as it can allow one or more of the lengths H_i to be smaller than q^{β_n}, as long as other H_i are commensurably larger. We cannot let H_i vary in a completely uncontrolled fashion, since our savings comes from the smallest parameter $H_{\min}$; thus we assume that $\boldsymbol{H}$ is proportionate, in the sense that $H_{\min} \geq \|\boldsymbol{H}\|^{c_0/n}$ for some constant $0 < c_0 \leq 1$. The relation (1-9) below shows that if we take c_0 smaller so that $H_{\min}$ becomes a decreasing proportion of $\|\boldsymbol{H}\|$, for our Burgess bound to be nontrivial the geometric mean $\|\boldsymbol{H}\|^{1/n}$ is forced into an ever shorter range near $q^{1/2}$.

Corollary 1.4. *For each $n \geq 2$ and $r \geq 1$, Theorem 1.1 with the parameters n, r provides a nontrivial upper bound $|S(F; \boldsymbol{N}, \boldsymbol{H})| = o_{n,r,\Delta,D}(\|\boldsymbol{H}\|)$ for all $\boldsymbol{H} = (H_1, \dots, H_n)$ with $H_{\min}H_{\max} < q^{1+1/2\Theta}$ as long as $H_{\min} \geq \|\boldsymbol{H}\|^{c_0/n}$ for some $0 < c_0 \leq 1$ and $\|\boldsymbol{H}\|^{1/n} = q^{\beta}$ with β in the range*

$$\frac{1}{2} - \frac{c_0\Theta - 1}{2\Theta(n+c_0)} < \beta \leq \frac{1}{2} + \frac{1}{4\Theta}, \tag{1-8}$$

in which $\Theta = \Theta_{n,r} = \lfloor (r-1)/(n-1) \rfloor$. This range includes a nonempty interval of $\beta < \frac{1}{2}$ as soon as $r \geq (1/c_0 + 1)(n-1) + 1$, so that $c_0\Theta > 1$. In particular this range always requires $\|\boldsymbol{H}\|^{1/n} = q^{\beta}$ with

$$\beta > \beta_{n,c_0} = \frac{1}{2} - \frac{c_0}{2(n+c_0)} \geq \frac{1}{2} - \frac{1}{2(n+1)}. \tag{1-9}$$

Alternatively, we can state that Theorem 1.1 obtains a nontrivial upper bound if

$$\|\boldsymbol{H}\| H_{\min} \gg q^{n/2+\kappa}$$

for some small κ. As $\kappa \to 0$ we obtain a nontrivial bound $\|\boldsymbol{H}\|q^{-\delta_n}$ with

$$\delta_n \approx \frac{(n+c_0)^2}{4(n-1)}\kappa^2.$$

1C. ***Overview of previous literature.*** To situate our results, we recall previous literature on Burgess bounds and in particular for the sums $S(F; \boldsymbol{N}, \boldsymbol{H})$.

1C1. *The classical Burgess bound.* For any integer $q \geq 1$, the Pólya–Vinogradov inequality states that $|S(N, H)| \ll q^{1/2} \log q$ (see [Pólya 1918; Vinogradov 1918] or for a modern treatment [Iwaniec and Kowalski 2004, §12.4]). This provides a nontrivial upper bound for $|S(N, H)|$ as long as $H \gg q^{1/2+\varepsilon}$ for some $\varepsilon > 0$. When H is shorter than this range, the sum is considered to be "short," and obtaining an $o(H)$ bound is much more difficult. Conjecturally, under the generalized Riemann hypothesis, a bound as strong as $|S(0, H)| \ll_\varepsilon H^{1/2} q^\varepsilon$ should hold for all $\varepsilon > 0$, thus leading to a nontrivial upper bound in any range $H \gg_\varepsilon q^{3\varepsilon}$ (see, e.g., [Iwaniec and Kowalski 2004, Equation (12.54)], or [Friedlander et al. 2013, Equation (9.6)]; see also the more general Conjecture C_n in [Friedlander et al. 2013, §9]).

Burgess [1957; 1962; 1963; 1986] showed that for χ a primitive character to a prime modulus q, for all integers $r \geq 1$, the inequality (1-1) holds, with an implied constant uniform in N, yielding a nontrivial bound for $H \gg q^{1/4+\kappa}$, $\kappa > 0$. More generally, with $\log q$ replaced by q^ε for arbitrarily small ε, Burgess proved that this bound also holds for cube-free moduli q for all $r \geq 1$, and for any integer q, for $r \leq 3$. The $q^{1/4+\kappa}$ "threshold" of the Burgess bound remains essentially unimproved since its inception, despite significant interest, due to its applications.

As a consequence of (1-1), Burgess [1963] proved a landmark subconvexity bound

$$\left|L\left(\tfrac{1}{2} + it, \chi\right)\right| \ll_{t,\varepsilon} q^{1/4-1/16+\varepsilon}$$

for all $\varepsilon > 0$, with χ a nonprincipal Dirichlet character modulo q as above; there is a corresponding hybrid subconvexity bound $\left|L\left(\frac{1}{2} + it, \chi\right)\right| \ll_\varepsilon (|t|q)^{1/4-1/16+\varepsilon}$ of Heath-Brown [1980]. This remains the best bound known to hold for all Dirichlet L-functions. (Special cases of the modulus q in which a better subconvexity bound is known include: smooth moduli [Graham and Ringrose 1990; Goldmakher 2010; Chang 2014; Irving 2016]; prime-power moduli [Postnikov 1956; Barban et al. 1964; Gallagher 1972; Milićević 2016], or powerful moduli [Iwaniec 1974]. Most recently, Petrow and Young [2018] proved a better Weyl-strength subconvex estimate of size $q^{1/6}$ for all cube-free moduli q.[1]) There is great interest in establishing bounds of at least the strength of a "Burgess exponent" for analogous bounds in higher rank contexts. For example, in the GL(2) setting, the "Burgess bound" for an L-function of a Hecke cusp form g twisted by a primitive Dirichlet character χ modulo q is $\left|L\left(\frac{1}{2} + it, g \otimes \chi\right)\right| \ll_{g,\varepsilon} q^{1/2-1/8+\varepsilon}$ for all $\varepsilon > 0$ (as has been obtained in [Bykovskii 1996; Blomer and Harcos 2008]). Reaching the Burgess exponent in new settings, or even re-proving such Burgess bounds is currently an important proving ground for new methods (e.g., [Munshi 2018] via a GL(2) delta method and subsequently [Aggarwal et al. 2018] via a trivial delta method).

In another direction, the Burgess bound establishes an upper bound for the least quadratic nonresidue $n(p)$ modulo a prime p. Vinogradov conjectured that $n(p) \ll_\varepsilon p^\varepsilon$ for every $\varepsilon > 0$; Burgess's

[1] After the writing of this paper, Petrow and Young [2019] posted a preprint proving a Weyl-strength subconvex estimate for all Dirichlet L-functions, without a restriction on the modulus q. This breakthrough result also implies a small improvement on the classical Burgess upper bound (1-1) for the cases $r = 2, 3$ but not for larger r.

bound (1-1) proves $n(p) \ll_\varepsilon p^{(4\sqrt{e})^{-1}+\varepsilon}$ for any $\varepsilon > 0$, which held the record from [Burgess 1957] until the quantification in [Banks and Guo 2017]. In this vein, there are continued efforts toward the goal of improving the inequality (1-1) directly, such as reducing the power of the logarithm (see [Iwaniec and Kowalski 2004, Equation (12.58) and Remark p. 329; Kerr et al. 2019]), deducing improvements in certain special cases from conjectural improvements on the Pólya–Vinogradov inequality [Fromm and Goldmakher 2019], and making connections to the Elliott–Halberstam conjecture and "Type II sums" in sieve methods [Tao 2015, Conjecture 1.5, Theorem 1.6, Remark 1.7].

1C2. *Previous literature on special cases of* $S(F; \boldsymbol{N}, \boldsymbol{H})$. In the n-dimensional setting of the sums $S(F; \boldsymbol{N}, \boldsymbol{H})$, previous literature mainly focused on two special cases. Burgess [1967b; 1968] considered the case in which

$$F(\boldsymbol{x}) = \prod_{i=1}^{n} L_i(\boldsymbol{x}) \tag{1-10}$$

is a product of n linear forms $L_i \in \mathbb{Z}[x_1, \dots, x_n]$ that are linearly independent over $\mathbb{F}_q$, for q prime. In this multilinear setting, he proved a nontrivial bound $|S(F; \boldsymbol{N}, \boldsymbol{H})| = O(\|\boldsymbol{H}\| q^{-\delta})$ for $\boldsymbol{H} = (H, \dots, H)$ and a certain $\delta = \delta(\kappa) > 0$ as long as $H = q^{\beta_n+\kappa}$ for some $\kappa > 0$, with β_n as defined in (1-7). Bourgain and Chang [2010] incorporated ideas from additive combinatorics to improve this significantly, proving a nontrivial bound $|S(F; \boldsymbol{N}, \boldsymbol{H})| = O(\|\boldsymbol{H}\| q^{-\delta})$ for $\boldsymbol{H} = (H, \dots, H)$ and a certain $\delta = \delta(\kappa) > 0$ as long as $H = q^{1/4+\kappa}$ for some $\kappa > 0$, thus obtaining an n-dimensional result as strong as the original Burgess threshold in each dimension.

The second case in which significant results are known is in dimension $n = 2$ when F is a binary quadratic form. In this special case, the work of Burgess above, for bilinear sums in $n = 2$, combined with results of Davenport and Lewis [1963] on analogues of the Burgess bound over $\mathbb{F}_{q^2}$, initially provided a nontrivial upper bound for $S(F, \boldsymbol{N}, \boldsymbol{H})$ for q prime and $H_i > q^{1/3+\kappa}$ (that is $H_i > q^{\beta_2+\kappa}$ with β_2 as in (1-7)), where F is any binary quadratic form that is not a perfect square over $\mathbb{F}_q$; that is for any $F(x_1, x_2) = x_1^2 + ax_1x_2 + bx_2^2$ with $a^2 \not\equiv 4b \pmod q$. Chang [2009, Theorem 11] introduced ideas of additive combinatorics to this setting, and improved this to a nontrivial upper bound for $S(F, \boldsymbol{N}, \boldsymbol{H})$ for $H_1, H_2 > q^{1/4+\kappa}$, i.e., a 2-dimensional result as strong as the original Burgess threshold in each dimension. Most recently, Heath-Brown [2016, Theorem 3] proved that this latter result continues to hold for any odd square-free modulus q such that $(q, \det(F)) = 1$.

Remark. It remains an interesting open question to bound $S(F; \boldsymbol{N}, \boldsymbol{H})$, with all H_i as short as $q^{1/4+\kappa}$ for $\kappa > 0$, in the general case of an arbitrary form F whose reduction modulo q is (Δ, q)-admissible.

1C3. *Further related literature.* We briefly mention certain other results that are related to multivariate sums similar to $S(F; \boldsymbol{N}, \boldsymbol{H})$, although not of exactly the type we consider in this paper. Davenport and Lewis [1963] considered the case of χ a nonprincipal character of $\mathbb{F}_{q^n}^*$ for q prime, and a linear form $F(x_1, \dots, x_n) = \omega_1 x_1 + \cdots + \omega_n x_n$ for $\omega_1, \dots, \omega_n$ a fixed basis of $\mathbb{F}_{q^n}$. They proved that $S(F, \boldsymbol{N}, \boldsymbol{H}) = O(\|\boldsymbol{H}\| q^{-\delta})$ for some $\delta = \delta(\kappa)$ where $H_i > q^{\beta_n+\kappa}$ for some $\kappa > 0$. As remarked in [Burgess 1968], for

$n = 2$ this provides a corresponding upper bound for $S(F; \boldsymbol{N}, \boldsymbol{H})$ in the case that F is an irreducible binary quadratic form $x_1^2 + ax_1x_2 + bx_2^2$ over $\mathbb{F}_q$, in which case $\chi(F(x_1, x_2))$ is a character $(\mathrm{mod}\ q)$ of $x_1 + \omega x_2 \in \mathbb{Q}(\omega)$, where $\omega = \frac{1}{2}a + \frac{1}{2}\sqrt{a^2 - 4b}$. In the n-dimensional setting, if one assumes that $\omega_1, \dots, \omega_n$ is a certain special type of basis (such as a power basis), stronger results have also been obtained; Burgess [1967a] and Karatsuba [1970] proved nontrivial upper bounds in the stronger range $H_i > q^{1/4+\kappa}$. Without such special assumptions, Chang [2008] improved on Davenport and Lewis (for $n \geq 5$) by proving a nontrivial bound as soon as $\|\boldsymbol{H}\| > q^{2n/5+\kappa}$ for some $\kappa > 0$, and furthermore in dimension $n = 2$ Chang [2009, Theorem 5] proved that $H_i \gg q^{1/4+\kappa}$ suffices for any $\kappa > 0$. See also [Chang 2008] for certain results of Burgess-type for multiplicative character sums over sumsets.

Finally, we mention work on mixed character sums in multivariate settings, of the form

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

with a polynomial $g \in \mathbb{R}[x_1, \dots, x_n]$. In [Pierce 2016] the first author proved nontrivial upper bounds for such multivariate sums in the regime $H_i \gg q^{1/4+\kappa}$, in the special case $F(\boldsymbol{x}) = x_1 \cdots x_n$ (generalizing [Heath-Brown and Pierce 2015] in dimension $n = 1$). This was later generalized by Kerr [2014] to the case of $F(\boldsymbol{x})$ multilinear as in (1-10). A second paper in this series will prove Burgess bounds for $S(F, g; \boldsymbol{N}, \boldsymbol{H})$ for any form F whose reduction modulo q is (Δ, q)-admissible.

1D. ***Outline of the paper.*** We present in Section 2 a heuristic overview of the proof of Theorem 1.1, which illustrates how the stratification result of the second author [Xu 2018] plays a key role. In Section 3 we gather together the lemmas we need to motivate and utilize the condition of (Δ, q)-admissibility. In Section 4 we give a convenient restatement and strengthening of the stratification results of [Xu 2018]; we expect this version will be of independent interest in other applications. In Section 5 we begin the n-dimensional Burgess argument, reaching the key new novel steps in Section 6, which carries out the stratification and a Menshov–Rademacher argument involving permuting variables. In Section 7 we complete the Burgess argument and choose parameters optimally; subsequently we verify the corollaries. In the Appendix we provide a comparison illustrating the utility of the stratification, and a conditional argument that assumes a stronger stratification result, which shows that the threshold β_n is stable under such an improvement.

1E. ***Notation.*** We will use the notation $\boldsymbol{x} \in (\boldsymbol{N}, \boldsymbol{N} + \boldsymbol{H}]$ to denote the range of a sum over a box $(\boldsymbol{N}, \boldsymbol{N} + \boldsymbol{H}] = \prod_i (N_i, N_i + H_i]$, and will let $\|\boldsymbol{H}\| = \prod_i H_i$ for any tuple $\boldsymbol{H}$. We will write $\boldsymbol{a}q$ to mean $(a_1q, \dots, a_nq)$ and $\boldsymbol{a}/q$ to mean $(a_1/q, \dots, a_n/q)$. We will also use notations such as $\boldsymbol{x} \leq \boldsymbol{a}$ to denote $x_i \leq a_i$ for $i = 1, \dots, n$ and $\boldsymbol{a}$ $(\mathrm{mod}\ p)$ to mean we regard each a_i $(\mathrm{mod}\ p)$. We define $H_{\max} := \max_i H_i$ and $H_{\min} := \min_i H_i$. We will use the Vinogradov notation $A \ll B$ to denote that there exists a constant C such that $|A| \ll CB$, and $A \ll_\kappa B$ to denote that C may depend on the parameter κ. In the following work, all implied constants may depend on n, r, D, $\Delta := \mathrm{ord}\ \chi$ and possibly an arbitrarily small $\varepsilon > 0$ without further specification, but will never depend on $\boldsymbol{N}$, $\boldsymbol{H}$, q.

2. Method of proof: an overview

In this section we recall the main points of the Burgess method in dimension 1, outline the difficulties that arise in $n \geq 2$ dimensions, and then sketch how we overcome these difficulties to obtain our main theorem.

At its heart, the Burgess method in the classical 1-dimensional setting builds from a character sum $S(N; H)$ of length $H < q$, a character sum over a complete set of residues modulo q. Doing so by a Fourier expansion only works efficiently if the sum is not too short, that is, if the character sum is of length at least $H \gg q^{1/2+\varepsilon}$ (see Section A.3). When $H \ll q^{1/2+\varepsilon}$, Burgess's method instead dissects and translates the "short" sum into many "short-short" sums of length H/p for some prime p of size roughly P (with P to be chosen optimally in terms of H and q). Averaging this process over sufficiently many choices of p, the short-short sums become distributed across a long interval of length q. If this process is done with enough redundancy, the starting points of these short-short sums nearly cover a full set of residues modulo q. Simultaneously, Burgess considers not just an average of these short-short sums, but a $2r$-th moment, leading to the study of a sum over starting points m, of the form

$$\sum_{m} \max_{k \leq 2H/P} |S(m, k)|^{2r}. \tag{2-1}$$

At this point, positivity allows one to sum over all $1 \leq m \leq q$ so that the sum over m is a complete set of residues. The Menshov–Rademacher technique allows one to deduce a bound for this maximal moment (2-1) from a bound for the nonmaximal moment

$$\sum_{m \pmod q} |S(m, k)|^{2r} \tag{2-2}$$

in which we think of m as varying over the starting points of the short-short sums and $k \leq 2H/P$ as being the new short-short length. We may write (2-2) equivalently as

$$\sum_{x_1,\dots,x_{2r} \in (0,k]} \sum_{m \pmod q} \chi(F_{\{\boldsymbol{x}\}}(m)), \tag{2-3}$$

in which $F_{\{\boldsymbol{x}\}}(m) = (m + x_1)(m + x_2)^{\Delta-1} \cdots (m + x_{2r})^{\Delta-1}$. If $F_{\{\boldsymbol{x}\}}(m)$ is not a perfect Δ-th power modulo q, then the Weil bound $O(q^{1/2})$ applies to the sum over m, and we say the tuple $\boldsymbol{x}$ is "good" (which is the generic case); otherwise $\boldsymbol{x}$ is "bad" (which is a sparse case) and we apply the trivial bound $O(q)$ to the sum over m. Balancing the contributions of these two cases leads to the optimal choice of P and the Burgess bound (1-1).

Generalizing this argument to the n-dimensional case, we will prove in (5-9) that for any $r \geq 1$,

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll (\log P) P^{n-1/2r} \|\boldsymbol{H}\|^{-1/2r} \Bigg(\sum_{\substack{\boldsymbol{m} \\ |m_i| \leq 2q}} \max_{\boldsymbol{k} \leq 2\boldsymbol{H}/P} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \Bigg)^{1/(2r)}. \tag{2-4}$$

This will ultimately reduce the problem of bounding $|S(F; \boldsymbol{N}, \boldsymbol{H})|$ to bounding

$$\sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \le \sum_{\substack{\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(2r)} \\ \boldsymbol{x}^{(i)} \in (\boldsymbol{0}, \boldsymbol{k}]}} \Big| \sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big|, \tag{2-5}$$

where we define for each collection $\{\boldsymbol{x}\} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\}$ with $\boldsymbol{x}^{(i)} \in \mathbb{Z}^n$ the polynomial

$$F_{\{\boldsymbol{x}\}}(\boldsymbol{X}) = F_{\{\boldsymbol{x}\}}(X_1, \dots, X_n) = \prod_{i=1}^{2r} (F(\boldsymbol{X} + \boldsymbol{x}^{(i)}))^{\delta(i)}, \tag{2-6}$$

where $\delta(i) = +1$ if i is odd and $\Delta - 1$ if i is even. One would hope that if the $\boldsymbol{x}^{(i)}$ are appropriately independent (that is, the "good" case), a generalization of the Weil bound would yield square-root cancellation, that is an $O(q^{n/2})$ bound for

$$\Big| \sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big|. \tag{2-7}$$

But achieving such a bound has been a critical barrier to generalizing the Burgess method to this n-dimensional setting. One difficulty is that the leading form of $F_{\{\boldsymbol{x}\}}(\boldsymbol{X})$ (the homogeneous part of highest degree) defines a highly singular projective variety, whereas previous literature on Weil bounds required this either to be a nonsingular projective variety [Katz 2002], or could only allow certain singular varieties that are not general enough for our application; see, e.g., [Rojas-León 2005; 2006].

Moreover, in dimensions $n \ge 2$, as well as the two extremal cases in which the sum (2-7) is $O(q^{n/2})$ or $O(q^n)$, there may be intermediate cases $O(q^{(n+j-1)/2})$ for $j = 1, \dots, n+1$. Indeed, suppose we partition the $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r}$ into the following types: those belonging to a good set denoted by $\mathrm{Good}(\boldsymbol{k})$, are such that

$$\Big| \sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big| \le C q^{n/2}$$

for a certain constant C, and those belonging to the j-th bad set, denoted by $\mathrm{Bad}_j(\boldsymbol{k})$, are such that

$$\Big| \sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big| > C q^{(n+j-1)/2} \qquad \text{but} \qquad \Big| \sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big| \le C q^{(n+j)/2}.$$

Then according to this dissection,

$$\sum_{\boldsymbol{m} \,(\mathrm{mod}\, q)} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \le C |\mathrm{Good}(\boldsymbol{k})| q^{n/2} + C \sum_{j=1}^{n} |\mathrm{Bad}_j(\boldsymbol{k})| q^{(n+j)/2}. \tag{2-8}$$

Certainly $|\mathrm{Good}(\boldsymbol{k})| \le \|\boldsymbol{k}\|^{2r}$. The real question is how to bound $|\mathrm{Bad}_j(\boldsymbol{k})|$ for each $j = 1, \dots, n$. The recent work of the second author [Xu 2018] proves a set of bounds that are perfectly suited for our purposes. This takes the form of a "stratification," in the spirit of [Fouvry 2000, Proposition 1.0; Laumon 2000, Proposition 3.2, Theorem 3.3; Fouvry and Katz 2001, Theorems 1.1 and 1.2].

To give the flavor of this stratification, we state here a special case in dimension $n = 2$ (see Section 4 and in particular Theorem 4.4 for the full setting). Let $\boldsymbol{k} = (k, k, \dots, k)$ be a fixed tuple in $\mathbb{Z}^n$, with $k \geq 1$. We will consider tuples $\boldsymbol{x} \in \mathbb{Z}^n$ that lie in the box $\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{k}]$, and more generally, a collection $\{\boldsymbol{x}\} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\} \in (\mathbf{0}, \boldsymbol{k}]^{2r}$ of $2r$ such n-tuples.

Theorem 2.1. *Let $n = 2$ and let $r, \Delta, D \geq 1$ be fixed. There exists a constant $C = C(n, r, D)$ and a constant $C'' = C''(n, r, \Delta, D)$ such that the following holds. For any prime q, for any nonprincipal multiplicative Dirichlet character χ of order Δ modulo q, and for any $F \in \mathbb{F}_q[x_1, x_2]$ that is (Δ, q)-admissible, for every tuple $\boldsymbol{k} = (k, k)$,*

$$\#\left\{\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\} \in (\mathbf{0}, \boldsymbol{k}]^{2r} : \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m}))\Big| > Cq^{(n+j-1)/2}\right\} \leq C'' \begin{cases} k^{4r} & \text{if } j = 0, \\ k^{3r+1} & \text{if } j = 1, \\ k^{2r} & \text{if } j = 2. \end{cases}$$

We can interpret this as follows: for $n = 2$, the trivial bound for the number of collections $\{\boldsymbol{x}\}$ in the box $(\mathbf{0}, \boldsymbol{k}]^{2r} \subset \mathbb{Z}^{2nr}$ is k^{4r}. When $j = 1$, we see that at most $O(k^{3r+1})$ collections can violate square-root cancellation, i.e., as soon as $r \geq 2$, generically square-root cancellation holds. This stratification, in its general formulation (Theorem 4.4) is the key input which allows us to prove the Burgess bound for all dimensions $n \geq 2$. (See Section A.1 for a further demonstration of why the full stratification is useful.)

We stated Theorem 2.1 for simplicity in the case where the length $\boldsymbol{k}$ has identical values in each coordinate. In our argument, we must instead allow $\boldsymbol{k} = (k_1, \dots, k_n)$ with k_i varying independently; in particular this arises in the step when we deduce a bound for the maximal moment in (2-4) from the nonmaximal moment in (2-5). This raises another difficulty in the multi-dimensional setting, which we now outline.

Xu's stratification (Theorem A, Theorem 4.1) shows that for each $j = 1, \dots, n$, the collections $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\}$ counted by $\mathrm{Bad}_j(\boldsymbol{k})$ lie on a certain subscheme over $\mathbb{F}_q$ with a certain codimension. In general, fix a dimension R, let $X \subset \mathbb{A}^R_{\mathbb{F}_q}$ be a subscheme of codimension ϖ and let U be the sum of the degrees of its irreducible components. Given subsets $M_i \subset \mathbb{F}_q$, define the "box" $M = \prod_{i=1}^R M_i \subset \mathbb{A}^R_{\mathbb{F}_q}$. We require a bound for

$$I_X(M) := \#\bigg(X(\mathbb{F}_q) \cap \prod_{i=1}^R M_i\bigg), \tag{2-9}$$

which depends only on the codimension of X and the degree of the irreducible components of X. A trivial bound, best possible if $\operatorname{codim} X = 0$, is $I_X(M) \leq \|M\| := \prod_{i=1}^R |M_i|$. We need to improve on this when $\operatorname{codim} X \geq 1$.

To gain an intuition, consider the case of $R = 2$ and X of codimension 1. A naive hope might be that $I_X(M) \leq \|M\|^{1/2} = \|M\| \cdot |M_1|^{-1/2}|M_2|^{-1/2}$. But this need not be true. Supposing for example that $|M_1| \approx 1$ is very small while $|M_2|$ is very large, it could happen that X of codimension 1 lies along the subset M_2 of M, thus leading to the bound $I_X(M) \gg |M_2| = \|M\| \cdot |M_1|^{-1} \approx \|M\|$. Thus in general, in estimating $I_X(M)$ we can only expect to save factors corresponding to the smallest components of the "box" M.

Concretely, for X of codimension ϖ, we use the fact that if

$$1 \le |M_1| \le |M_2| \le \cdots \le |M_R| < \infty, \tag{2-10}$$

then by [Xu 2018, Lemma 1.7],

$$I_X(M) \le U\|M\| \cdot |M_1|^{-1}|M_2|^{-1}\cdots|M_\varpi|^{-1}, \tag{2-11}$$

where U is the sum of the degrees of the irreducible components of X. In our setting, we will apply such bounds in Section 6 when using a Menshov–Rademacher technique to deduce a bound for a maximal moment from a nonmaximal moment. Here, in order to guarantee the ordering (2-10) we must permute variables in a delicate argument, and apply rearrangement inequalities in order to conclude. This is another novel aspect of our method.

3. Preliminaries

3A. ***Power-free conditions.*** We say that $F \in \mathbb{Z}[x_1, \dots, x_n]$ is d-th power-free if each nonconstant irreducible factor of F over $\mathbb{Z}$ (or equivalently over $\mathbb{Q}$, by Gauss's lemma) appears with multiplicity strictly less than d. In general, given a field k, we say that $F \in k[x_1, \dots, x_n]$ is d-th power-free over k when $F = cF_1^{a_1}F_2^{a_2}\cdots F_\ell^{a_\ell}$ with $c \in k^\times$, all $a_i < d$ and all $F_i \in k[x_1, \dots, x_n]$ are irreducible and pairwise nonassociate. (We say $G, G' \in k[x_1, \dots, x_n]$ are nonassociate if there is no $c \in k^\times$ such that $G = cG'$.) (Note that a constant may be regarded as a d-th power-free polynomial. In lemmas or theorems which assume a polynomial is d-th power-free, other hypotheses will rule out the case of the polynomial being a constant.)

To be precise, we recall that the property of being d-th power-free may be specified equivalently over a field k or the separable closure of k:

Lemma 3.1. *Let k be a field and k^s its separable algebraic closure, so $k^s = \bar{k}$ if k is perfect, and in particular if k is finite. Then for any $F \in k[x_1, \dots, x_n]$, F is a perfect d-th power over k (up to a nonzero multiplicative constant) if and only if F is a perfect d-th power over k^s. Similarly, F is d-th power-free over k if and only if F is d-th power-free over k^s.*

Proof. We begin with the second claim. Certainly if F is d-th power-free over k^s then it is over k. For the other direction, write $F = cF_1^{a_1}F_2^{a_2}\cdots F_\ell^{a_\ell}$ with $c \in k^\times$, all $a_i < d$ and all $F_i \in k[x_1, \dots, x_n]$ irreducible and pairwise nonassociate. For each such F_i, we recall from [Xu 2018, Lemma 3.15 (2)] that the fact that F_i is irreducible over k implies that F_i is square-free as a polynomial in $k^s[x_1, \dots, x_n]$. Thus upon factoring F_i over k^s we have $F_i = G_{i,1}G_{i,2}\cdots G_{i,b_i}$ in which each $G_{i,j}$ is irreducible in k^s and as j varies the $G_{i,j}$ are pairwise nonassociate. Thus the factorization of F over k^s is

$$c(G_{1,1}^{a_1}G_{1,2}^{a_1}\cdots G_{1,b_1}^{a_1})(G_{2,1}^{a_2}G_{2,2}^{a_2}\cdots G_{2,b_2}^{a_2})\cdots(G_{\ell,1}^{a_\ell}G_{\ell,2}^{a_\ell}\cdots G_{\ell,b_\ell}^{a_\ell}).$$

Next we recall from [Xu 2018, Lemma 3.15 (3)] that if F_i, $F_{i'} \in k[x_1, \dots, x_n]$ are nonassociate irreducible polynomials, then F_i and $F_{i'}$ have no common factors in $k^s[x_1, \dots, x_n]$. From this we conclude that

$G_{i,j}$ and $G_{i',j'}$ are nonassociate when $(i, j) \neq (i', j')$. Thus as i and j vary the $G_{i,j}$ are all pairwise nonassociate over k^s, so that F remains d-th power-free over k^s.

Finally, if F is a perfect d-th power over k then it also is over k^s. In the other direction, if F is a perfect d-th power over k^s, then in the factorization above, all a_i must be multiples of d, so that F factors over k as $F = cG^d$ with $G = F_1^{a_1/d} \cdots F_\ell^{a_\ell/d}$. □

3B. ***Translation invariance conditions.*** It is natural to impose on F that it be appropriately nondegenerate, in the sense that it cannot be made independent of one or more variables. Indeed, if there exists a linear change of coordinates $\boldsymbol{x} \mapsto \boldsymbol{x}A$ with $A \in \mathrm{GL}_n(\mathbb{Z})$ such that $F(\boldsymbol{x}A) \in \mathbb{Z}[x_2, \dots, x_n]$ then we would not expect $|S(F; \boldsymbol{N}, \boldsymbol{H})|$ to obey bounds of the full n-dimensional strength that we will obtain. We will require several equivalent formulations for the condition that F is nondegenerate in this sense.

We recall six equivalent statements about a polynomial $F \in \mathbb{Z}[x_1, \dots, x_n]$ having the property that it can be made independent of one of the indeterminates by a linear change of coordinates over $\mathbb{Z}$.

Lemma 3.2 [Xu 2018, Lemma 3.20]. *Let $F \in \mathbb{Z}[x_1, \dots, x_n]$ and let $\boldsymbol{x} = (x_1, \dots, x_n)$ be the row vector of indeterminates. Then the following are equivalent*:

(1) *F is invariant under some nontrivial translation in $\overline{\mathbb{Q}}^n$, i.e., there exists $0 \neq \boldsymbol{m} \in \overline{\mathbb{Q}}^n$ such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$.*

(2) *F is invariant under some nontrivial translation in $\mathbb{Z}^n$, i.e., there exists $0 \neq \boldsymbol{m} \in \mathbb{Z}^n$ such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$.*

(3) *F can be made independent of one of the indeterminates by a linear change of coordinates, i.e., there exists $A \in \mathrm{GL}_n(\mathbb{Z})$ such that $F(\boldsymbol{x}A) \in \mathbb{Z}[x_2, \dots, x_n]$.*

(4) *When viewed as a morphism $\mathbb{A}^n_{\mathbb{Z}} \to \mathbb{A}^1_{\mathbb{Z}}$, F factors through a linear map $\mathbb{A}^n_{\mathbb{Z}} \to \mathbb{A}^{n-1}_{\mathbb{Z}}$, i.e., there exists an integral $n \times (n-1)$ matrix B and $f \in \mathbb{Z}[x_2, \dots, x_n]$ such that $F(\boldsymbol{x}) \equiv f(\boldsymbol{x}B)$.*

(5) *For almost all prime numbers q (all but finitely many), the reduction of F modulo q is invariant under some nontrivial translation in $\overline{\mathbb{F}}_q^n$.*

(6) *For infinitely many prime numbers q, the reduction of F modulo q is invariant under some nontrivial translation in $\overline{\mathbb{F}}_q^n$.*

We now also require an analogue of this over a field k.

Lemma 3.3. *Let k be a perfect field. Let $F \in k[x_1, \dots, x_n]$ and suppose that* $\deg F < \operatorname{char} k$ *if k is of positive characteristic. Let $\boldsymbol{x}$ be the row vector $(x_1, \dots, x_n)$ of indeterminates. Then the following are equivalent*:

(1) *F is invariant under some nontrivial translation in $\bar{k}^n$, i.e., there exists $\boldsymbol{0} \neq \boldsymbol{m} \in \bar{k}^n$ such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$.*

(2) *F is invariant under some nontrivial translation in k^n, i.e., there exists $\boldsymbol{0} \neq \boldsymbol{m} \in k^n$ such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$.*

(3) *F can be made independent of one of the indeterminates by a linear change of coordinates, i.e., there exists $A \in \mathrm{GL}_n(k)$ such that $F(\boldsymbol{x}A) \in k[x_2, \dots, x_n]$.*

(4) *When viewed as a morphism $\mathbb{A}_k^n \to \mathbb{A}_k^1$, F factors through a linear map $\mathbb{A}_k^n \to \mathbb{A}_k^{n-1}$, i.e., there exists an $n \times (n-1)$ matrix B with entries in k, and $f \in k[x_2, \dots, x_n]$ such that $F(\boldsymbol{x}) \equiv f(\boldsymbol{x}B)$.*

Proof of Lemma 3.3. (1) $\Longrightarrow$ (2): Let $\mathbf{0} \neq \boldsymbol{m} = (m_1, \dots, m_n) \in \bar{k}^n$ be such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$, and assume without loss of generality that $m_1 \neq 0$. By iteration, $F(\boldsymbol{x} + t\boldsymbol{m}) - F(\boldsymbol{x}) \equiv 0$ as a function of $\boldsymbol{x}$, for all $t \in \mathbb{Z}$, hence for all $t \in k_0$, the prime field inside k, if $\operatorname{char} k > 0$. We consider separately the case of characteristic zero: if $\operatorname{char} k = 0$ we directly conclude that $F(\boldsymbol{x} + t\boldsymbol{m}) - F(\boldsymbol{x}) = 0$ as a polynomial in $\bar{k}[x_1, \dots, x_n, t]$ since a nonzero polynomial cannot have infinitely many roots (namely all $t \in \mathbb{Z}$). In the positive characteristic case, we learn that $t^{\operatorname{char} k} - t$ divides $F(\boldsymbol{x} + t\boldsymbol{m}) - F(\boldsymbol{x})$ as polynomials in $\bar{k}[x_1, \dots, x_n, t]$. Under the assumption $\deg F < \operatorname{char} k$, it therefore must be the case that

$$F(\boldsymbol{x} + t\boldsymbol{m}) - F(\boldsymbol{x}) = 0 \tag{3-1}$$

as a polynomial in $\bar{k}[x_1, \dots, x_n, t]$. Now let E be the field generated by $m_1, \dots, m_n$ over k, and choose any $t \in E$ such that $\mathrm{Tr}_{E/k}(t) \in k \setminus \{0\}$; such a t is guaranteed as long as E/k is separable, which holds because we assumed that k is perfect. (For $\boldsymbol{u} \in E^n$ we will let $\mathrm{Tr}_{E/k}(\boldsymbol{u}) = (\mathrm{Tr}_{E/k}(u_1), \dots, \mathrm{Tr}_{E/k}(u_n))$.) Then since $m_1 \neq 0$, we see that also $\mathrm{Tr}_{E/k}(t\boldsymbol{m}/m_1) \in k^n \setminus \{\mathbf{0}\}$. We will now observe that $F(\boldsymbol{x} + \mathrm{Tr}_{E/k}(t\boldsymbol{m}/m_1)) \equiv F(\boldsymbol{x})$, concluding the proof of (2). Since F has coefficients in k, then for any $\sigma \in \mathrm{Gal}(E/k)$, we have $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \sigma(t\boldsymbol{m}/m_1))$. Consequently, we have $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \mathrm{Tr}_{E/k}(t\boldsymbol{m}/m_1))$, as desired.

(2) $\Longrightarrow$ (3): Let $\mathbf{0} \neq \boldsymbol{m} = (m_1, \dots, m_n) \in k^n$ be such that $F(\boldsymbol{x}) \equiv F(\boldsymbol{x} + \boldsymbol{m})$; then proceeding as in the previous argument, this implies that $F(\boldsymbol{x} + t\boldsymbol{m}) \equiv F(\boldsymbol{x})$ as polynomials in $k[x_1, \dots, x_n, t]$. We will show that there exists $A \in \mathrm{GL}_n(k)$ such that $\boldsymbol{m} = (1, 0, \dots, 0)A$. Once we have this matrix, we observe that $F(\boldsymbol{x}A) \equiv F(\boldsymbol{x}A + \boldsymbol{m}) \equiv F((\boldsymbol{x} + (1, 0, \dots, 0))A)$ so that upon defining $G(\boldsymbol{x}) = F(\boldsymbol{x}A)$, we have that $G(\boldsymbol{x})$ is invariant under translation $\boldsymbol{x} \mapsto \boldsymbol{x} + (1, 0, \dots, 0)$. Consequently, when regarded as polynomial in x_1,

$$G(x_1, x_2, \dots, x_n) - G(0, x_2, \dots, x_n) \tag{3-2}$$

has all integers as its roots, and hence all elements in k_0 as its roots if $\operatorname{char} k > 0$. In the characteristic zero case, this implies that (3-2) is the zero polynomial in x_1, and hence $F(\boldsymbol{x}A) = G(\boldsymbol{x}) = G(0, x_2, \dots, x_n) \in k[x_2, \dots, x_n]$. If $\operatorname{char} k > 0$, we learn that $x_1^{\operatorname{char} k} - x_1$ divides (3-2), but since $\deg F < \operatorname{char} k$, it must be the case that (3-2) is identically the zero polynomial over k. Hence as before we have $F(\boldsymbol{x}A) = G(\boldsymbol{x}) = G(0, x_2, \dots, x_n) \in k[x_2, \dots, x_n]$, concluding the proof.

Finally, we construct the matrix A. Note that $\mathrm{GL}_n(k)$ acts transitively on nonzero vectors in k^n, since any such vector is an element in a basis for k^n, and there exists a unique element in $\mathrm{GL}_n(k)$ mapping one ordered basis to another. Thus in particular there exists $A \in \mathrm{GL}_n(k)$ such that $\boldsymbol{m} = (1, 0, \dots, 0)A$, as desired.

(3) $\Longrightarrow$ (4): Suppose that $F(\boldsymbol{x}A) \equiv f(x_2, \dots, x_n)$ for some $f \in k[x_2, \dots, x_n]$, so $F(\boldsymbol{x}) \equiv F((\boldsymbol{x}A^{-1})A) \equiv f((\boldsymbol{x}A^{-1})_2, \dots, (\boldsymbol{x}A^{-1})_n)$, where $A^{-1} \in \mathrm{GL}_n(k)$. Then it suffices to define B to be the matrix constructed of the last $n - 1$ columns of A^{-1}.

(4) $\Longrightarrow$ (1): Now we suppose that there exists such an $n \times (n-1)$ matrix B and $f \in k[x_2, \dots, x_n]$ such that $F(\boldsymbol{x}) \equiv f(\boldsymbol{x}B)$. Since multiplication by B is a linear map from k^n to k^{n-1}, the nullspace of this map is nontrivial and hence there exists $\mathbf{0} \neq \boldsymbol{m} \in k^n$ such that $\boldsymbol{m}B = \mathbf{0}$; consequently

$$F(\boldsymbol{x}) \equiv f(\boldsymbol{x}B) \equiv f(\boldsymbol{x}B + \boldsymbol{m}B) \equiv f((\boldsymbol{x}+\boldsymbol{m})B) \equiv F(\boldsymbol{x}+\boldsymbol{m}).$$

This implies (2), which certainly implies (1), concluding the proof of the lemma. □

3C. ***All but finitely many primes.*** In the introduction we stated that forms that are "admissible" over $\mathbb{Z}$ are "admissible" over $\mathbb{F}_q$ for all but finitely many primes. The formal statement is here:

Lemma 3.4. *Let $\Delta \geq 1$ be fixed. Let $F \in \mathbb{Z}[x_1, \dots, x_n]$ and suppose F factors as $F = G^{\Delta}H$ with $G, H \in \mathbb{Z}[x_1, \dots, x_n]$ and H being Δ-th power-free over $\mathbb{Z}$. Furthermore, assume that H is nondegenerate over $\mathbb{Z}$, in the sense that there is no $A \in \mathrm{GL}_n(\mathbb{Z})$ such that $H(\boldsymbol{x}A) \in \mathbb{Z}[x_2, \dots, x_n]$. Then for all but finitely many primes q, the reduction of F modulo q is (Δ, q)-admissible.*

Proof. For a fixed prime q, we reduce F, G, H modulo q to $f, g, h \in \mathbb{F}_q[x_1, \dots, x_n]$. By Lemma 3.22 of [Xu 2018], since H is Δ-th power-free over $\mathbb{Z}$ then for all but finitely many primes q, h is Δ-th power-free over $\mathbb{F}_q$. Letting Q_1 denote this finite set of exceptional primes, then for all $q \notin Q_1$, we have that $f = g^{\Delta}h$ with h being Δ-th power-free over $\mathbb{F}_q$. As a consequence of Lemma 3.2 (6), since H is nondegenerate over $\mathbb{Z}$, the reduction h of H modulo q can be invariant under a nontrivial translation in $\overline{\mathbb{F}}_q^n$ only for finitely many primes q; we will call this exceptional set Q_2. Finally, let Q_3 denote the primes $q \leq \deg F$. We now proceed to consider the primes $q \notin Q_1 \cup Q_2 \cup Q_3$; for such primes, h is not invariant under any nontrivial translation in $\mathbb{F}_q^n$, and (3) in Lemma 3.3 shows that h cannot be made independent of any indeterminate by a linear change of variables in $\mathrm{GL}_n(\mathbb{F}_q)$. This proves the lemma. (Here we excluded primes in Q_3 because we cite Lemma 3.3, but we note that the specific implications of this lemma that we employ here do not need the assumption $q > \deg F$.) □

3D. ***Permutations of variables.*** Within the Burgess method, we will use a variant of the Menshov–Rademacher method for deducing bounds for maximal moments from nonmaximal moments. To carry out this argument in our setting, we will need to reorder the variables $x_1, x_2, \dots, x_n$ in $F(\boldsymbol{x})$ so that a corresponding tuple of parameters $(k_1, \dots, k_n)$ satisfies the ordering $k_1 \leq k_2 \leq \cdots \leq k_n$. Thus we are led to consider forms resulting from F when the variables are permuted. For any permutation π of $\{1, \dots, n\}$, define the form $F_\pi(\boldsymbol{X})$ from the form $F(\boldsymbol{X})$ by setting $F_\pi(X_1, \dots, X_n) = F(X_{\pi(1)}, \dots, X_{\pi(n)})$.

Lemma 3.5. *Let $\Delta \geq 1$ and a prime q be fixed. If a form $F \in [x_1, \dots, x_n]$ has the property that its reduction modulo q is (Δ, q)-admissible, then for every permutation π on a set of n elements, the reduction modulo q of F_π is (Δ, q)-admissible.*

Proof. Letting f denote the reduction of F modulo q, we write $f = g^{\Delta}h$ with h being Δ-th power-free over $\mathbb{F}_q$; then correspondingly for the permuted versions, if f_π denotes the reduction of F_π modulo q then $f_\pi = (g_\pi)^{\Delta}h_\pi$ with h_π being Δ-th power-free over $\mathbb{F}_q$. Moreover, h can be made independent of at least one variable after a $\mathrm{GL}_n(\mathbb{F}_q)$ change of variable if and only if h_π can. □

4. The stratification of complete character sums

In this section we recall the stratification of complete character sums proved in [Xu 2018] and show how to deduce a slightly stronger formulation that we believe will be of independent interest, as well as being useful in this paper. In our presentation, we will replace the dimension r in the original work by the dimension $2r$ in each instance, but the content of this section would apply in an analogous way for any dimension r (odd or even).

4A. ***The stratification obtained by Xu.*** We first recall the statement of [Xu 2018, Theorem 1.1 and Corollary 1.8] in our setting of dimension $2r$. For each fixed $n, r \geq 1$, we define a set of parameters θ_j for $1 \leq j \leq n$, as follows:

$$\theta_j = \begin{cases} 0 & \text{if } j = 0, \\ j\lfloor (r-1)/(n-1) \rfloor & \text{if } j = 1, \dots, n-2, \\ r-1 & \text{if } j = n-1, \\ nr & \text{if } j = n. \end{cases} \tag{4-1}$$

Note that this differs superficially from the definition of θ_j in [Xu 2018, p. 2]: we are working with dimension $2r$ in place of r and the floor function results in slightly different formulas. (Precisely, from Xu's work we may take

$$\theta_1 = \lfloor (2r-1)/(2n-2) \rfloor = \lfloor (r-1)/(n-1) \rfloor \tag{4-2}$$

and then we set $\theta_j = j\theta_1$ for $1 \leq j \leq n-1$. In particular for $j = n-1$,

$$\theta_{n-1} = \lfloor (2r-1)/2 \rfloor = r-1.$$

In fact, Xu's original theorem allows a slightly larger value of $\theta_j \geq j\theta_1$; it is later apparent in (7-13) that this slightly larger choice, leading to a slightly stronger version of Theorem A, would not significantly improve our current application.)

Theorem A [Xu 2018, Theorem 1.1]. *Let integers $n, r, \Delta, D \geq 1$ be fixed. There exist integers $C = C(n, r, D) \geq 1$ and $C' = C'(n, r, \Delta, D) \geq 1$ and a finite set $\mathcal{S} = \mathcal{S}(n, r, \Delta, D)$ of primes such that the following holds.*

Let κ denote a finite field with algebraic closure denoted by $\bar{\kappa}$. For each $1 \leq i \leq 2r$, let $\chi_i : \kappa^\times \to \mathbb{C}^\times$ be a multiplicative character (extended to κ by setting $\chi_i(0) = 0$) and assume that $d_i := \operatorname{ord}(\chi_i) | \Delta > 0$. Let $F_i \in \kappa(x_1, \dots, x_n)$ be a d_i-th power-free rational function of degree at most D and assume that

$$T_{F_i} := \{\boldsymbol{m} \in \bar{\kappa}^n : F_i(\boldsymbol{x}) \equiv F_i(\boldsymbol{x} + \boldsymbol{m})\} \tag{4-3}$$

is finite for each $1 \leq i \leq 2r$. Then upon defining

$$S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) := \sum_{\boldsymbol{m} \in \kappa^n} \prod_{i=1}^{2r} \chi_i(F_i(\boldsymbol{m} + \boldsymbol{x}^{(i)})), \tag{4-4}$$

we have that whenever $\operatorname{char}\kappa \notin \mathcal{S}$, *there exist subschemes*

$$\mathbb{A}^{2nr}_\kappa = X_0 \supset X_1 \supset X_2 \supset \cdots \supset X_n$$

such that for each $1 \le j \le n$,

(1) *the sum of the degrees of irreducible components of* X_j *is at most* C';

(2) $\dim X_j \le 2nr - \theta_j$ *with* θ_j *defined as in* (4-1);

(3) *for all* $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}^{2nr}(\kappa) \setminus X_j(\kappa)$,

$$|S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| \le C(\#\kappa)^{(n+j-1)/2}.$$

Remark. The hypothesis that for each $1 \le i \le 2r$, χ_i has order $d_i > 0$ could allow that χ_i is the principal character. However, if χ_i is principal then $d_i = 1$ and hence to be d_i-th power-free $F_i \in \kappa(x_1, \dots, x_n)$ must be a constant function. This implies that T_{F_i} is infinite, thus excluding this possibility from the theorem. This remark applies to Corollary B, below, and the resulting Theorem 4.1 and Corollary 4.2.

Remark. Note that while the subschemes X_j may depend on the F_i, the parameters C, C', θ_j depend only on the maximum degree D of the F_i. The constant $C = C(n, r, D)$ noted above is bounded above by $(4r(D+1)+1)^n$ as computed in [Katz 2001, Theorem 11] (see the remarks following Theorem 1.1 in [Xu 2018], again recalling that we have $2r$ in place of r). The constant C' (and later C'' in Corollary 4.3) only depends on n, r, D, Δ but is not explicitly determined. For further details on the dependence of C' and C'' see [Xu 2018, Section 2.1, and the remark following Theorem 1.1] and Corollary 4.3.

Next, we must convert Theorem A into a count for the number of points $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}^{2nr}(\kappa)$ (later denoted as collections $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\}$) such that a given upper bound for $|S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})|$ holds. It is convenient to define for each $0 \le j \le n$ and any sequence $1 \le k_1 \le k_2 \le \cdots \le k_n < \infty$ the function

$$B_{n,r}(j; \boldsymbol{k}) = B_{n,r}(j; k_1, \dots, k_n) = \begin{cases} 1 & \text{if } j = 0, \\ k_1^{\theta_j} = k_1^{j\lfloor (r-1)/(n-1)\rfloor} & \text{if } j = 1, \dots, n-2, \\ k_1^{\theta_{n-1}} = k_1^{r-1} & \text{if } j = n-1, \\ (k_1 \cdots k_{n/2})^{2r} & \text{if } j = n,\ n \text{ even}, \\ (k_1 \cdots k_{(n-1)/2})^{2r} k_{(n+1)/2}^r & \text{if } j = n,\ n \text{ odd}. \end{cases} \tag{4-5}$$

(To aid comparison to Xu's original notation [2018] involving parameters denoted by n_0, η, we make the simple observation that for $j = n$ and $\theta_j = nr$, if n is even we write $nr = (n/2)2r$ (with $n_0 = n/2$ and $\eta = 0$) and when n is odd we write $nr = ((n-1)/2) \cdot 2r + r$ (with $n_0 = (n-1)/2$ and $\eta = r$), so that Xu's original notation results in the expression stated above for $B_{n,r}(j, \boldsymbol{k})$ in the case $j = n$, and with ambient dimension $2r$.)

Corollary B [Xu 2018, Corollary 1.8]. *Let integers* $n, r, \Delta, D \ge 1$ *be fixed. There exist integers* $C = C(n, r, D) \ge 1$ *and* $C' = C'(n, r, \Delta, D) \ge 1$ *and a finite set* $\mathcal{S} = \mathcal{S}(n, r, \Delta, D)$ *of primes such that the following holds.*

Let κ denote a finite field with algebraic closure denoted by $\bar{\kappa}$. For each $1 \le i \le 2r$, let $\chi_i : \kappa^\times \to \mathbb{C}^\times$ be a multiplicative character (extended to κ by setting $\chi_i(0) = 0$) and assume that $d_i := \operatorname{ord}(\chi_i) | \Delta > 0$. Let $F_i \in \kappa(x_1, \dots, x_n)$ be a d_i-th power-free rational function of degree at most D and assume that T_{F_i} as defined in (4-3) *is finite for each $1 \le i \le 2r$.*

Let $\{M_i\}_{i=1}^n$ be subsets of κ such that $1 \le |M_1| \le |M_2| \le \cdots \le |M_n| < \infty$, and define $M := \prod_{i=1}^n M_i \subset \kappa^n$ and $\|M\| := \prod_{i=1}^n |M_i|$. Then whenever $\operatorname{char} \kappa \notin \mathcal{S}$, for each $1 \le j \le n$,

$$\#\big\{(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in M^{2r} : |S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| > C(\#\kappa)^{(n+j-1)/2}\big\} \le C' \|M\|^{2r} B_{n,r}(j; |M_1|, \dots, |M_n|)^{-1}. \quad (4\text{-}6)$$

4B. ***The stratification in our setting.*** Now we state the new versions of Theorem A and Corollary B that we use in this paper.

Theorem 4.1. *Let $\kappa = \mathbb{F}_q$, q prime. The result of Theorem A holds if we replace the hypothesis that for each $1 \le i \le 2r$, the form F_i is d_i-th power-free and T_{F_i} is finite, by the weaker hypothesis that for each $1 \le i \le 2r$, the reduction modulo q of the form F_i is (d_i, q)-admissible, and for each $1 \le i \le 2r$, $\deg F_i < q$.*

Corollary 4.2. *Let $\kappa = \mathbb{F}_q$, q prime. The result of Corollary B holds if we replace the hypothesis that for each $1 \le i \le 2r$, the form F_i is d_i-th power-free and T_{F_i} is finite, by the weaker hypothesis that for each $1 \le i \le 2r$, the reduction modulo q of the form F_i is (d_i, q)-admissible.*

In the above corollary, we are able to omit the condition $\deg F_i < q$ seen in Theorem 4.1 by possibly enlarging C'; see (4-9). Through similar considerations, we can remove consideration of the set $\mathcal{S}$, as we record here:

Corollary 4.3. *In addition, given n, r, Δ, D in either Corollary B or Corollary 4.2, we may take the set $\mathcal{S} = \mathcal{S}(n, r, \Delta, D)$ to be the empty set, at the expense of replacing $C' = C'(n, r, \Delta, D)$ by a possibly larger constant $C'' = C''(n, r, \Delta, D)$.*

Within the Burgess argument, we will consider the sum

$$\sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) = \sum_{\boldsymbol{m} \pmod q} \prod_{i=1}^{2r} \chi_i(F(\boldsymbol{m} + \boldsymbol{x}^{(i)}))$$

in which we have fixed a multiplicative Dirichlet character χ of order Δ modulo a prime q and then set $\chi_i = \chi$ if i is odd and $\chi_i = \bar{\chi}$ if i is even. This is then clearly of the form $S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})$ as defined in (4-4), with the choice that all the F_i are equal to our fixed $F \in \mathbb{F}_q[x_1, \dots, x_n]$ of degree D. For later reference, we record the following immediate consequence of Theorem 4.1 and Corollary 4.3 (choosing the subset M_i of κ to be $(0, k_i]$ for each $1 \le i \le 2r$).

Theorem 4.4. *Let integers $n, r, \Delta, D \ge 1$ be fixed. Then there exist constants $C = C(n, r, D)$ and $C'' = C''(n, r, \Delta, D)$ such that the following holds.*

Fix a prime q and let χ be a multiplicative Dirichlet character of order Δ modulo q. Suppose $F \in \mathbb{Z}[x_1, \dots, x_n]$ has the property that its reduction modulo q is (Δ, q)-admissible and define $F_{\{\boldsymbol{x}\}}(\boldsymbol{X})$ accordingly as in (2-6). *Then for every $1 \le j \le n$, for every tuple $\boldsymbol{k} = (k_1, \dots, k_n) \in \mathbb{Z}^n$ with $1 \le k_1 \le k_2 \le \cdots \le k_n \le q$,*

$$\#\left\{(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in (\boldsymbol{0}, \boldsymbol{k}]^{2r} : \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m}))\Big| > Cq^{(n+j-1)/2}\right\} \le C''\|\boldsymbol{k}\|^{2r} B_{n,r}(j;\boldsymbol{k})^{-1}, \quad (4\text{-}7)$$

in which $B_{n,r}(j;\boldsymbol{k})$ is defined as in (4-5).

Remark. Note that the trivial upper bound in (4-7) is $\|\boldsymbol{k}\|^{2r}$. The fundamental consequence of Theorem 4.4 is that it shows that *generically* among $\{\boldsymbol{x}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r}$, square-root cancellation holds, as soon as r is sufficiently large relative to n. Precisely, as soon as $r \ge n$, so that the exponent $\lfloor (r-1)/(n-1) \rfloor$ appearing in $B_{n,r}(1;\boldsymbol{k})$ is strictly positive, the number of $\{\boldsymbol{x}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r}$ such that square-root cancellation is violated is $O(\|\boldsymbol{k}\|^{2r} k_1^{-1})$, which suffices for our claim, as long as k_1 is at least a positive power of $\|\boldsymbol{k}\|$.

Also, to aid in understanding the role of the function $B_{n,r}(j;\boldsymbol{k})$ in this result, we note that the bound (4-7) is in the format of (2-11) with the choice $R = 2nr$ and with the R-dimensional box being

$$(0, k_1] \times \cdots \times (0, k_1] \times \cdots \times (0, k_n] \times \cdots \times (0, k_n],$$

in which each factor $(0, k_i]$ appears $2r$ times. Thus when $\theta_j \le 2r$ (this holds for $j \le n-1$), we only save factors of k_1. In the final case $j = n$ when $\theta_n = nr$, we save some factors of k_i for $1 \le i \le \lceil n/2 \rceil$ as well. This leads to the definition of $B_{n,r}(j;\boldsymbol{k})$. Finally, we remark for later reference that by construction, under the hypotheses of the theorem, we always have

$$\|\boldsymbol{k}\|^{2r} B_{n,r}(j;\boldsymbol{k})^{-1} \ge 1. \quad (4\text{-}8)$$

Remark. Conjecturally, one might hope to improve the result of Theorem 4.4 by proving that one can take larger values for the codimension θ_j used to define the savings factor $B_{n,r}(j;\boldsymbol{k})^{-1}$. (Precise implications may be found in Section A.2, where we show that even the conjecturally best possible values for the codimension do not significantly change our main result.) For comparison, in the most extreme case, it is not hard to see that we must have $\theta_n \le nr$, and hence certainly must also have $\theta_j \le nr$ for all $j \le n$. For recall from Theorem A that X_n is a subscheme of $\mathbb{A}_\kappa^{2nr}$ such that for all $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}_\kappa^{2nr} \setminus X_n(\kappa)$, $|S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| \le C(\#\kappa)^{n-1/2}$. In fact, Xu's paper shows the stronger result that for all finite extensions k/κ and $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}_k^{2nr} \setminus X_n(k)$, it holds that $|S_k(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| \le C(\#k)^{n-1/2}$; here S_k is defined analogously to S but summing over $\boldsymbol{m} \in k^n$ and with χ_i replaced by $\chi_i(N_{k/\kappa}(\cdot))$. That is, all $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})$ such that $|S_k(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| > C(\#k)^{n-1/2}$ must lie in $X_n(k)$. Then we claim that $\dim X_n \ge nr$ (and consequently $\theta_n \le nr$). To see this, we consider any tuple $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})$ for which $\boldsymbol{x}^{(j)} = \boldsymbol{x}^{(j+r)}$ for $j = 1, \dots, r$. There are $(\#k)^{nr}$ such tuples, and each of them has the property that $|S_k(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| = (\#k)^n$. This is $> C(\#k)^{n-1/2}$ if $\#k$ is sufficiently large (which we can choose it to be).

4C. ***Deduction of the corollaries.*** Corollary 4.2 follows from Theorem 4.1 in an identical fashion to how Corollary B follows from Theorem A in [Xu 2018] (see Corollary 1.8 in that paper) and we do not repeat the proof here. We only note that in Corollary 4.2 we no longer need to assume that $\deg F_i < q$. For indeed, suppose that for some $1 \le i \le 2r$ we have $\deg F_i \ge \operatorname{char}\kappa$. Then we note that trivially

$$|S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})| \le (\#\kappa)^{2nr} = q^{2nr} \le D^{2nr} \tag{4-9}$$

for all $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}_\kappa^{2nr}$, so that upon enlarging C' if necessary so that $C' \ge D^{2nr}$ the results of Corollary 4.2 hold (here we also use the fact (4-8)).

To obtain Corollary 4.3 in which formally $\mathcal{S} = \varnothing$, we note that for any $q \in \mathcal{S}(n, r, \Delta, D)$, we may write the trivial upper bound $C'\|M\|^{2r} \le C'q^{2nr}$ on the right-hand side of (4-6). Thus in order to state a version of Corollary B or Corollary 4.2 with $\mathcal{S} = \varnothing$, we simply replace C' in the statement of the corollary by $C''(n, r, \Delta, D) = \max\{C', q^{2nr} : q \in \mathcal{S}(n, r, \Delta, D)\}$.

4D. ***Proof of Theorem 4.1.*** Theorem 4.1 follows from a small modification inside the proof of Theorem A in [Xu 2018]. To be clear, we will state exactly the change that is made (recalling that in our setting we use $2r$ where [Xu 2018] uses r; our modifications would of course work for any dimension r).

The main idea is that even if there is an $i \in \{1, \dots, 2r\}$ such that F_i is not d_i-th power-free, we can write $F_i = G_i^{d_i}\widetilde{F}_i$ in which $\widetilde{F}_i$ is d_i-th power-free, and has $T_{\widetilde{F}_i}$ finite, under the assumption that the reduction of F_i modulo q is (d_i, q)-admissible, and then at a key moment in the proof we work with $\widetilde{F}_i$ instead of F_i. To be precise, recall that for any $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \kappa^{2nr}$,

$$S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) := \sum_{\boldsymbol{m} \in \kappa^n} \prod_{i=1}^{2r} \chi_i(F_i(\boldsymbol{m} + \boldsymbol{x}^{(i)})). \tag{4-10}$$

We also define for any tuple $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in \kappa^{2ns}$ and each $1 \le i \le 2r$ the function

$$T_i(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) = \sum_{\boldsymbol{x} \in \kappa^n} \chi_i(F_{i,\{\boldsymbol{m}\}}(\boldsymbol{x})),$$

in which

$$F_{i,\{\boldsymbol{m}\}}(\boldsymbol{x}) := \prod_{j=1}^{s} F_i(\boldsymbol{m}^{(j)} + \boldsymbol{x}) \prod_{j=s+1}^{2s} F_i(\boldsymbol{m}^{(j)} + \boldsymbol{x})^{d_i - 1}.$$

The proof of Theorem A and hence of Corollary B relies on four ingredients.

(I) The first ingredient [Xu 2018, Proposition 1.2] is an identity between $2s$-moments of the sums S with $2r$-multilinear averages of the sums T_i, namely

$$\sum_{(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}) \in \kappa^{2nr}} |S(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)})|^{2s} = \sum_{(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in \kappa^{2ns}} \prod_{i=1}^{2r} T_i(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}). \tag{4-11}$$

We will refer to the left-hand side as the moment $M_\kappa(r, s)$; it has a natural generalization to a moment $M_k(r, s)$ defined in an appropriately analogous manner over any finite extension k/κ, with $\chi_i(\cdot)$ replaced by $\chi_i(N_{k/\kappa}(\cdot))$.

(II) The second ingredient [Xu 2018, Proposition 1.5] relates the moments $M_k(r, s)$ for finite extensions k/κ to the dimension of the subschemes X_j.

(III) The third ingredient [Xu 2018, Proposition 1.6 (a)] is an upper bound of $O((\#k)^{ns})$ (that is, of square-root strength) for the number of tuples $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in k^{2ns}$ such that $F_{i,\{\boldsymbol{m}\}}(\boldsymbol{x})$ is a perfect d_i-th power in $\bar{k}(x_1, \dots, x_n)$.

(IV) The fourth ingredient [Xu 2018, Proposition 1.6 (b)] is an application of the Weil bound to save one factor of $(\#k)^{1/2}$ off the trivial bound $(\#k)^n$ for an n-dimensional character sum that is a generalization of T_i in an extension k/κ. Precisely, it is the statement that uniformly in finite extensions k/κ and tuples $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in k^{2ns}$, if $F_{i,\{\boldsymbol{m}\}} \in k(x_1, \dots, x_n)$ is not a perfect d_i-th power in $\bar{k}(x_1, \dots, x_n)$, then

$$\sum_{\boldsymbol{x} \in k^n} \chi_i(\mathrm{N}_{k/\kappa}(F_{i,\{\boldsymbol{m}\}}(\boldsymbol{x})) = O((\#k)^{n-1/2}). \tag{4-12}$$

These four ingredients are applied in a bootstrapping process. The general philosophy is that a weak bound with very few exceptions can be bootstrapped into a stronger bound with possibly more exceptions. More precisely, ingredients (III) and (IV) are the initial input, showing that a small savings holds for the sums T_i, aside from possibly $O((\#k)^{ns})$ many (that is, square-root many) exceptional choices of $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in k^{2ns}$. For the possible exceptional choices, a trivial upper bound of $O((\#k)^n)$ is applied in place of (4-12).

This input step provides a savings, on average, for the sums T_i on the multilinear right-hand side of the identity (4-11) in ingredient (I) and hence for the moment of S on the left-hand side of (I). Ingredient (II) then expresses this savings on the moment of S as a stratification in terms of a lower bound on $\mathrm{codim} X_j$ for each j. This result holds uniformly for sums S of the shape (4-10). Since each T_i is also a sum of this shape (with s, χ_i, F_i defined appropriately), the resulting bound for sums S can be applied to each T_i, yielding an improvement over the initial savings for T_i. This argument then bootstraps to prove the final result.

With this outline in hand, we may now briefly verify Theorem 4.1. The only point at which this argument utilized the assumption that each F_i is d_i-th power-free and has T_{F_i} finite was at the initial input to the bootstrapping, when ingredient (III) was used once; see [Xu 2018, §2.3]. Thus all we must do is show that this step, namely [Xu 2018, Proposition 1.6 (a)] can be proved under the alternate assumption that for each i, the reduction of F_i modulo q is (d_i, q)-admissible, and that $\deg F_i < q$. We will replace [Xu 2018, Proposition 1.6 (a)] with the following proposition.

Proposition 4.5. *Let $\kappa = \mathbb{F}_q$ with q prime and fix a rational function $F \in \kappa(x_1, \dots, x_n)$ with $\deg F < q$ such that the reduction of F modulo q is (d, q)-admissible. Fix $s \in \mathbb{N}$. Then for each finite extension k/κ, the number of $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2s)}) \in k^{2ns}$ such that $F_{\{\boldsymbol{m}\}}(\boldsymbol{x})$ is a perfect d-th power over $\bar{\kappa}$ is at most $O((\#k)^{ns})$.*

Once this has been proved, this replaces [Xu 2018, Proposition 1.6 (a)] as ingredient (III) in Xu's proof, and the results of Theorem A and Corollary B follow under our alternative hypotheses, thus verifying Theorem 4.1.

We prove Proposition 4.5 as follows. Suppose that F has the property that its reduction modulo q is (d, q)-admissible. Write $F = G^d \widetilde{F}$ in which $G, \widetilde{F} \in \kappa(x_1, \dots, x_n)$ and $\widetilde{F}$ is d-th power-free, with $\deg \widetilde{F} < \operatorname{char} \kappa$. Under the assumption that the reduction modulo q of F is (d, q)-admissible, by definition there is no linear change of variables $A \in \mathrm{GL}_n(\mathbb{F}_q)$ such that $\widetilde{F}(\boldsymbol{x}A) \in \mathbb{F}_q[x_2, \dots, x_n]$. Hence by Lemma 3.3 (which we may apply since $\deg \widetilde{F} < q$), the only value of $\boldsymbol{m} \in \bar{\kappa}^n$ such that $\widetilde{F}(\boldsymbol{x}) \equiv \widetilde{F}(\boldsymbol{x}+\boldsymbol{m})$ is $\boldsymbol{m} = \boldsymbol{0}$, so that

$$T_{\widetilde{F}} := \{\boldsymbol{m} \in \bar{\kappa}^n \mid \widetilde{F}(\boldsymbol{x}) \equiv \widetilde{F}(\boldsymbol{x}+\boldsymbol{m})\}$$

is certainly finite. Thus we may apply the following lemma to $\widetilde{F}$ (which we quote without repeating the proof):

Lemma 4.6 [Xu 2018, Lemma 3.16]. *Fix r, $D \geq 1$. There exists $C_0 = C_0(r, D)$ such that the following holds. Let κ be a finite field and κ_0 its prime field. Let $H \in \kappa(x_1, \dots, x_n)$ be a d-th power-free rational function of degree at most D, and assume that*

$$T_H = \{\boldsymbol{m} \in \bar{\kappa}^n \mid H(\boldsymbol{x}) \equiv H(\boldsymbol{x}+\boldsymbol{m})\}$$

is finite. For any finite extension k/κ and $\{a_i\}_{i=1}^{2r} \subset \mathbb{Z}$ such that $\gcd(d, a_i) = 1$ for each $1 \leq i \leq 2r$, let P_H be the collection of tuples $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2r)}) \in k^{2nr}$ such that the rational function $\prod_{i=1}^{2r} H(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i}$ is a perfect d-th power over $\bar{\kappa}$. Then

$$\#P_H \leq C_0(\#k)^{nr}(\#T_H)^r. \tag{4-13}$$

Now note that for any such set of exponents $\{a_i\}_{i=1}^{2r}$, we have that $\prod_{i=1}^{2r} \widetilde{F}(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i}$ is a perfect d-th power if and only if

$$\prod_{i=1}^{2r} F(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i} = \prod_{i=1}^{2r} \widetilde{F}(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i} \Bigg(\prod_{i=1}^{2r} G(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i} \Bigg)^d$$

is. Thus if we define P_F, respectively $P_{\widetilde{F}}$, to be the collection of tuples $(\boldsymbol{m}^{(1)}, \dots, \boldsymbol{m}^{(2r)}) \in k^{2nr}$ such that the rational function $\prod_{i=1}^{2r} F(\boldsymbol{x}+\boldsymbol{m}^{(i)})^{a_i}$ is a perfect d-th power over $\bar{\kappa}$ (or analogously for $\widetilde{F}$), we have by Lemma 4.6 that

$$\#P_F = \#P_{\widetilde{F}} \leq C_0(\#k)^{nr}.$$

This concludes the verification of Proposition 4.5 and hence of Theorem 4.1.

5. Initiation of the Burgess argument

We now derive the first steps of the Burgess method, generalizing the approach of [Gallagher and Montgomery 2010; Heath-Brown 2013] from the one-variable case and [Pierce 2016] in the multi-variable case. The central complications that are distinctive to our new stratified setting are mainly addressed in Section 6.

We make the observation that given a character sum $S(F; \boldsymbol{N}, \boldsymbol{H})$ with data F, $\boldsymbol{N}$, $\boldsymbol{H}$, we may reorder the variables $x_1, \dots, x_n$ so that the lengths $H_1, \dots, H_n$ satisfy

$$H_1 \leq H_2 \leq \cdots \leq H_n. \tag{5-1}$$

In particular, if the reduction of F modulo q is (Δ, q)-admissible, then it stays (Δ, q)-admissible after any reordering of the variables (Lemma 3.5). We will assume (5-1) from now on, and will prove the statement of Theorem 1.1 with $H_1 = H_{\min}$ and $H_n = H_{\max}$.

Fix a prime $p \nmid q$ such that $p \le H_{\min}$, and split each coordinate $x_i \in (N_i, N_i + H_i]$ into residue classes modulo p by writing $\boldsymbol{x} = q\boldsymbol{a} + p\boldsymbol{m}$ with $\boldsymbol{a}, \boldsymbol{m} \in \mathbb{Z}^n$, where $0 \le a_i < p$ and $m_i \in (N_i', N_i' + H_i']$, for which we define

$$N_i' = (N_i - a_i q)/p, \quad H_i' = H_i/p$$

for each $i = 1, \dots, n$. Then

$$S(F; \boldsymbol{N}, \boldsymbol{H}) = \sum_{\boldsymbol{a} \pmod p} \sum_{\boldsymbol{m} \in (\boldsymbol{N}', \boldsymbol{N}' + \boldsymbol{H}']} \chi(F(q\boldsymbol{a} + p\boldsymbol{m})).$$

By the fact that χ has period q, the homogeneity of F and the multiplicativity of χ,

$$S(F; \boldsymbol{N}, \boldsymbol{H}) = \chi(p^D) \sum_{\boldsymbol{a} \pmod p} \sum_{\boldsymbol{m} \in (\boldsymbol{N}', \boldsymbol{N}' + \boldsymbol{H}']} \chi(F(\boldsymbol{m})),$$

so that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \le \sum_{\boldsymbol{a} \pmod p} |S(F; \boldsymbol{N}', \boldsymbol{H}')|.$$

We now average over a set $\mathcal{P}$ of primes, $\mathcal{P} = \{P < p \le 2P : p \nmid q\}$, so that $|\mathcal{P}| \gg P(\log P)^{-1}$. We will later choose P so that

$$P \le H_i, \qquad 1 \le i \le n. \tag{5-2}$$

Then

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \le \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \sum_{\boldsymbol{a} \pmod p} |S(F; \boldsymbol{N}', \boldsymbol{H}')|. \tag{5-3}$$

Here we recall that $\boldsymbol{N}'$ and $\boldsymbol{H}'$ depend on $\boldsymbol{a}$ and p. We now average over the starting points $\boldsymbol{N}'$ and the lengths $\boldsymbol{H}'$ in order to make them independent of $\boldsymbol{a}$, p.

Lemma 5.1. *Fix $\boldsymbol{M} \in \mathbb{R}^n$ and $\boldsymbol{L} \in \boldsymbol{R}^n_{\ge 1}$. For any $\boldsymbol{K} \le \boldsymbol{L}$,*

$$|S(F; \boldsymbol{M}, \boldsymbol{K})| \le 2^{2n} \|\boldsymbol{L}\|^{-1} \sum_{\boldsymbol{m} \in (\boldsymbol{M} - \boldsymbol{L}, \boldsymbol{M}]} \max_{\boldsymbol{k} \le 2\boldsymbol{L}} |S(F; \boldsymbol{m}, \boldsymbol{k})|.$$

Proof. The first observation is that for any fixed $\boldsymbol{M}, \boldsymbol{K}$, and any fixed $\boldsymbol{m}$ with $m_i \le M_i$ for all $i = 1, \dots, n$,

$$S(F; \boldsymbol{M}, \boldsymbol{K}) = \sum_{\substack{\boldsymbol{\delta} = (\delta_1, \dots, \delta_n) \\ \delta_i \in \{0,1\}}} (-1)^{\sigma(\boldsymbol{\delta})} S(F; \boldsymbol{m}, \boldsymbol{M} - \boldsymbol{m} + (\boldsymbol{1} - \boldsymbol{\delta}) \cdot \boldsymbol{K}), \tag{5-4}$$

where $\sigma(\boldsymbol{\delta}) = \sum_i \delta_i$. We will prove this momentarily, but first we see how it implies Lemma 5.1. For any $K_i \le L_i$, and any m_i with $M_i - L_i < m_i \le M_i$, we have

$$0 \le M_i - m_i + (1 - \delta_i) K_i \le 2L_i,$$

for either choice of $\delta_i \in \{0, 1\}$. Hence for any $\boldsymbol{m}$ with $\boldsymbol{M} - \boldsymbol{L} \le \boldsymbol{m} \le \boldsymbol{M}$, (5-4) shows that

$$|S(F; \boldsymbol{M}, \boldsymbol{K})| \le 2^n \max_{\boldsymbol{k} \le 2\boldsymbol{L}} |S(F; \boldsymbol{m}, \boldsymbol{k})|. \tag{5-5}$$

There are at least $L_i/2$ integers with $M_i - L_i < m_i \le M_i$, so that there are at least $2^{-n}\|\boldsymbol{L}\|$ values $\boldsymbol{m}$ in the range $\boldsymbol{M} - \boldsymbol{L} < \boldsymbol{m} \le \boldsymbol{M}$, and the lemma now follows by averaging (5-5) over these values.

Finally, we prove (5-4). This is simple to see in the case of dimension $n = 1$: for any $m_1 \le M_1$, we observe that as intervals,

$$(M_1, M_1 + K_1] = (m_1, M_1 + K_1] \setminus (m_1, M_1].$$

Thus in terms of the sums over these intervals,

$$S(F; M_1, K_1) = S(F; m_1, M_1 - m_1 + K_1) - S(F, m_1, M_1 - m).$$

For general dimensions $n \ge 1$, (5-4) follows from the principle of inclusion-exclusion. Explicitly, the sum $S(F; \boldsymbol{M}, \boldsymbol{K})$ has a range of summation over the box $W := \prod_{i=1}^n (M_i, M_i + K_i]$. Temporarily, for ease of notation, let $S(W)$ denote this sum, with range of summation over the box W; similarly, given any box W' defined below, we will let $S(W')$ denote the corresponding sum, with range of summation over the box W'.

Suppose $\boldsymbol{m} \le \boldsymbol{M}$ is fixed. Let W^* be the larger box $W^* := \prod_{i=1}^n (m_i, M_i + K_i]$. For each $j = 1, \dots, n$, let $W^{(j)}$ be the box defined as a product of intervals for $1 \le i \le n$, with "short" interval $(m_j, M_j]$ in the j-th coordinate and "long" interval $(m_i, M_i + K_i]$ in every other coordinate, for $i \ne j$. For each j let $\overline{W^{(j)}}$ be the complement of $W^{(j)}$ inside W^*. Precisely, $\overline{W^{(j)}}$ is a product of intervals for $1 \le i \le n$, with interval $(M_j, M_j + K_j]$ in the j-th coordinate and interval $(m_i, M_i + K_i]$ for $i \ne j$. Observe that our box W of interest is $W = \bigcap_{i=1}^n \overline{W^{(j)}}$. Then by the principle of inclusion-exclusion,

$$S(W) = S(W^*) - \sum_{i=1}^n S(W^{(j)}) + \sum_{1 \le j_1 < j_2 \le n} S(W^{(j_1)} \cap W^{(j_2)}) + \cdots + (-1)^n S(W^{(1)} \cap W^{(2)} \cap \cdots \cap W^{(n)}).$$

We can recognize that this is the identity (5-4) as follows. First, on the left-hand side, $S(W)$ is by definition our desired sum $S(F; \boldsymbol{M}, \boldsymbol{K})$. Next, we observe that the entries on the right-hand side are in bijection with the entries on the right-hand side of (5-4). Precisely, fix a tuple $(\delta_1, \dots, \delta_n)$ with $\delta_i \in \{0, 1\}$. Suppose that this tuple has k entries of 1, occurring at the indices $j_1, \dots, j_k$. If $k = 0$ then in the notation of the identity (5-4), the sum $S(F; \boldsymbol{m}, \boldsymbol{M} - \boldsymbol{m} + (\boldsymbol{1} - \boldsymbol{\delta}) \cdot \boldsymbol{K})$ is precisely the sum $S(W^*)$. If $1 \le k \le n$, then $S(F; \boldsymbol{m}, \boldsymbol{M} - \boldsymbol{m} + (\boldsymbol{1} - \boldsymbol{\delta}) \cdot \boldsymbol{K})$ is precisely $S(W^{(j_1)} \cap \cdots \cap W^{(j_k)})$ appearing in the inclusion-exclusion identity above. In each case, the signs $(-1)^{\sigma(\boldsymbol{\delta})} = (-1)^k$ agree as well. □

We apply Lemma 5.1 to (5-3) with $L_i = H_i/P \ge 1$. We obtain

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \le |\mathcal{P}|^{-1} \sum_{p \in \mathcal{P}} \sum_{\boldsymbol{a} \pmod p} 2^{2n} \|\boldsymbol{H}\|^{-1} P^n \sum_{\boldsymbol{m} \in (\boldsymbol{N}' - \boldsymbol{H}/P, \boldsymbol{N}']} \max_{\boldsymbol{k} \le 2\boldsymbol{H}/P} |S(F; \boldsymbol{m}, \boldsymbol{k})|. \tag{5-6}$$

After rearranging,

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \le |\mathcal{P}|^{-1} 2^{2n} \|\boldsymbol{H}\|^{-1} P^n \sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m}) \max_{\boldsymbol{k} \le 2\boldsymbol{H}/P} |S(F; \boldsymbol{m}, \boldsymbol{k})|, \tag{5-7}$$

where we have defined

$$\mathcal{A}(\boldsymbol{m}) = \#\left\{ \boldsymbol{a},\, p : p \in \mathcal{P}, \boldsymbol{a} = (a_1, \dots, a_n), 0 \le a_i < p : \frac{N_i - a_i q}{p} - \frac{H_i}{P} < m_i \le \frac{N_i - a_i q}{p}, i = 1, \dots, n \right\}.$$

We record the following facts about $\mathcal{A}(\boldsymbol{m})$, whose proof we defer to Section 7B.

Lemma 5.2. *The quantity $\mathcal{A}(\boldsymbol{m})$ vanishes unless $\boldsymbol{m}$ satisfies $|m_i| \le 2q$ for each i. Moreover, if*

$$H_i P < q, \qquad 1 \le i \le n, \tag{5-8}$$

then

$$\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m}) \ll P\|\boldsymbol{H}\|, \qquad \sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m})^2 \ll P\|\boldsymbol{H}\|.$$

Applying Hölder's inequality twice to (5-7), we obtain

$$\begin{aligned} &|S(F; \boldsymbol{N}, \boldsymbol{H})| \\ &\quad \ll |\mathcal{P}|^{-1} \|\boldsymbol{H}\|^{-1} P^n \Big(\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m}) \Big)^{1-1/r} \Big(\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m})^2 \Big)^{1/(2r)} \Big(\sum_{\substack{\boldsymbol{m} \\ |m_i| \le 2q}} \max_{\boldsymbol{k} \le 2\boldsymbol{H}/P} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \Big)^{1/(2r)}. \end{aligned}$$

Thus applying the results of Lemma 5.2 shows that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll (\log P) P^{n-1/2r} \|\boldsymbol{H}\|^{-1/2r} \Big(\sum_{\substack{\boldsymbol{m} \\ |m_i| \le 2q}} \max_{\boldsymbol{k} \le 2\boldsymbol{H}/P} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \Big)^{1/(2r)}. \tag{5-9}$$

It is sufficient to look at the internal sum over $\boldsymbol{m}$ modulo q; in fact obtaining this complete character sum is the main accomplishment of the manipulations up to this point.

We ignore for the moment the maximum over $\boldsymbol{k} \le 2\boldsymbol{H}/P$ and focus first on estimating the nonmaximal moment. We rewrite $S(F; \boldsymbol{m}, \boldsymbol{k})$ as

$$S(F; \boldsymbol{m}, \boldsymbol{k}) = \sum_{\boldsymbol{x} \in (\boldsymbol{m}, \boldsymbol{m}+\boldsymbol{k}]} \chi(F(\boldsymbol{x})) = \sum_{\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{k}]} \chi(F(\boldsymbol{m}+\boldsymbol{x})),$$

so that upon expansion,

$$\sum_{\boldsymbol{m} \pmod q} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} = \sum_{\boldsymbol{m} \pmod q} \Big| \sum_{\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{k}]} \chi(F(\boldsymbol{m}+\boldsymbol{x})) \Big|^{2r} \le \sum_{\substack{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)} \\ \boldsymbol{x}^{(i)} \in (\boldsymbol{0}, \boldsymbol{k}]}} \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big|, \tag{5-10}$$

where $F_{\{\boldsymbol{x}\}}$ is defined in terms of the original form F and the collection $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(2r)}\}$ by (2-6).

6. Stratification and a Menshov–Rademacher argument

6A. ***Application of the stratification for character sums.*** We now come to a critical novel step, which is to estimate how often we obtain a certain quality of upper bound for the complete character sum

$$\sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})).$$

For this, we call upon the stratification of complete character sums stated in Theorem 4.4. Let us suppose that $k_1 \le \ldots \le k_n \le q$. For each $1 \le j \le n$, define

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

in which $C = C(n, r, D)$ is the constant provided by Theorem 4.4. Then by Theorem 4.4, we have $(\boldsymbol{0}, \boldsymbol{k}]^{2r} =: Y_0 \supset Y_1 \supset Y_2 \supset \cdots \supset Y_n \supset Y_{n+1} := \varnothing$, and $\#Y_j \le C''\|\boldsymbol{k}\|^{2r} B_{n,r}(j; \boldsymbol{k})^{-1}$. Upon employing the decomposition $(\boldsymbol{0}, \boldsymbol{k}]^{2r} = \coprod_{j=0}^{n} Y_j \setminus Y_{j+1}$ in (5-10) we have

$$\begin{aligned} \sum_{\boldsymbol{m} \pmod q} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} &\le \sum_{j=0}^{n} \sum_{\{\boldsymbol{x}\} \in Y_j \setminus Y_{j+1}} \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m})) \Big| \\ &\le \sum_{j=0}^{n} (\#Y_j) C q^{(n+(j+1)-1)/2} \\ &\le C \cdot C'' \|\boldsymbol{k}\|^{2r} \sum_{j=0}^{n} q^{(n+j)/2} B_{n,r}(j; \boldsymbol{k})^{-1}. \end{aligned} \tag{6-1}$$

To summarize, we have proved:

Lemma 6.1. *Fix $n, r, D, \Delta \ge 1$, a prime q, and a multiplicative Dirichlet character χ of order Δ modulo q. Then for every form $F \in \mathbb{Z}[X_1, \ldots, X_n]$ of degree D such that its reduction modulo q is (Δ, q)-admissible, the following holds. For every $\boldsymbol{k} = (k_1, \ldots, k_n)$ with $1 \le k_1 \le k_2 \le \cdots \le k_n \le q$, then*

$$\sum_{\boldsymbol{m} \pmod q} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \ll_{n,r,D,\Delta} \|\boldsymbol{k}\|^{2r} \sum_{j=0}^{n} q^{(n+j)/2} B_{n,r}(j; \boldsymbol{k})^{-1}, \tag{6-2}$$

with $B_{n,r}(j; \boldsymbol{k})$ as defined in (4-5)*, and with the implicit constant dependent on n, r, D, Δ but independent of q, χ, F, $\boldsymbol{k}$.*

However recall that the actual quantity we must bound in (5-9) is a moment of $|S(F; \boldsymbol{m}, \boldsymbol{k})|$ that includes a maximum over $\boldsymbol{k} \le 2\boldsymbol{H}/P$. To do so, we will employ a Menshov–Rademacher argument.

6B. ***A Menshov–Rademacher argument with permutations.*** The Menshov–Rademacher argument [Menshov 1923; Rademacher 1922] may be employed in a wide variety of circumstances; in general it allows one to replace a supremum of a function $|f(u_j)|$ over an index set of size U by a sum of differences $|f(u_j) - f(u_{j-1})|$ over an index set of size $O(\log U)$. In our present setting, these differences

are differences of partial sums over a box, which are themselves partial sums of the same kind, so that the Menshov–Rademacher device is a useful tool.

However a typical Menshov–Rademacher argument would not immediately apply in our case, since we cannot save a power of $\|\boldsymbol{k}\|$ but typically only a power of $k_{\min}$, the shortest side of the box. We see this phenomenon is already present in Lemma 6.1, since we have assumed an ordering $k_1 \le k_2 \le \cdots \le k_n$ for the side-lengths of the box. Even if we assume in the beginning that we have such an ordering, certain internal steps in the Menshov–Rademacher argument do not preserve such an ordering, and thus arranging the argument so that we may apply Lemma 6.1 will require delicate considerations of permutations of the variables.

Our main result in this section is the following:

Proposition 6.2. *If $1 \le K_1 \le K_2 \le \cdots \le K_n \le q$, then*

$$\sum_{\substack{\boldsymbol{m} \\ |m_i| \le 2q}} \max_{\boldsymbol{k} \le \boldsymbol{K}} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} \ll_{n,r,D,\Delta} \|\boldsymbol{K}\|^{2r} (\log K_n)^{2nr} \sum_{j=0}^{n} q^{(n+j)/2} \widetilde{B}_{n,r}(j; \boldsymbol{K})^{-1} \tag{6-3}$$

with the implicit constant dependent on n, r, D, Δ but independent of q, χ, $\boldsymbol{K}$, and with

$$\widetilde{B}_{n,r}(j; \boldsymbol{K}) = \begin{cases} 1 & \text{if } j = 0, \\ K_1^{\theta_j} & \text{if } j = 1, \dots, n-1, \\ (K_1 \cdots K_{n/2})^{(2r-1)} & \text{if } j = n, \ n \text{ even}, \\ (K_1 \cdots K_{(n-1)/2})^{(2r-1)} (K_{(n+1)/2})^{r} & \text{if } j = n, \ n \text{ odd}. \end{cases} \tag{6-4}$$

Note that $\widetilde{B}_{n,r}(n; \boldsymbol{K})^{-1}$ loses one power in decay compared to $B_{n,r}(n; \boldsymbol{K})^{-1}$. In contrast, in the one-dimensional case, no decay is lost when passing from the nonmaximal estimate to the maximal estimate; this minor loss will not affect our final outcome, and we explain why it arises below.

6C. ***A dyadic decomposition.*** It suffices to prove Proposition 6.2 for $\boldsymbol{K} = (K_1, \dots, K_n)$ where each K_i is a power of 2, say $K_i = 2^{t_i}$, in which case we naturally have $t_1 \le t_2 \le \cdots \le t_n$, under the hypotheses of the proposition. We fix $\boldsymbol{m}$ and suppose that $S(F; \boldsymbol{m}, \boldsymbol{k})$ assumes its maximum for $\boldsymbol{k} \le \boldsymbol{K}$ at $\boldsymbol{k} = (k_1, \dots, k_n)$, and for each i we decompose

$$k_i = \sum_{\delta_i \in \mathcal{D}_i} 2^{t_i - \delta_i},$$

where $\mathcal{D}_i$ is a set of distinct nonnegative integers $\delta_i \le t_i$. Let $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ denote a tuple with $\delta_i \in \mathcal{D}_i$ for each i. Let $\boldsymbol{t} = (t_1, \dots, t_n)$ and define the notation $2^{\boldsymbol{t}-\boldsymbol{\delta}} = (2^{t_1-\delta_1}, \dots, 2^{t_n-\delta_n})$. Given a tuple $\boldsymbol{v}$, we define $2^{\boldsymbol{t}-\boldsymbol{\delta}} \circ \boldsymbol{v} = (2^{t_1-\delta_1} v_1, \dots, 2^{t_n-\delta_n} v_n)$. Furthermore let $\boldsymbol{V}_{\boldsymbol{m},\boldsymbol{\delta}}$ be an n-tuple defined such that the i-th entry is $(\boldsymbol{V}_{\boldsymbol{m},\boldsymbol{\delta}})_i = 2^{t_i-\delta_i} v_{\boldsymbol{m},\boldsymbol{\delta},i}$, where

$$v_{\boldsymbol{m},\boldsymbol{\delta},i} = \sum_{e_i \in \mathcal{D}_i, e_i < \delta_i} 2^{\delta_i - e_i} < 2^{\delta_i}.$$

Then we may express

$$S(F; \boldsymbol{m}, \boldsymbol{k}) = \sum_{\boldsymbol{\delta}, \delta_i \in \mathcal{D}_i} S(F; \boldsymbol{m} + \boldsymbol{V}_{\boldsymbol{m},\boldsymbol{\delta}}, 2^{\boldsymbol{t}-\boldsymbol{\delta}}).$$

After taking absolute values, we can obtain an upper bound by increasing the sum to run over all tuples $\boldsymbol{\delta}$ with $\delta_i \le t_i$. We get

$$|S(F;\boldsymbol{m},\boldsymbol{k})| \le \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} |S(F;\boldsymbol{m}+\boldsymbol{V}_{\boldsymbol{m},\boldsymbol{\delta}},2^{\boldsymbol{t}-\boldsymbol{\delta}})|.$$

Then by Hölder's inequality,

$$|S(F;\boldsymbol{m},\boldsymbol{k})|^{2r} \le \Bigg(\prod_{i=1}^{n}(t_i+1)\Bigg)^{2r-1} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \sum_{\substack{\boldsymbol{v}\\ 0\le v_i<2^{\delta_i}}} |S(F;\boldsymbol{m}+2^{\boldsymbol{t}-\boldsymbol{\delta}}\circ\boldsymbol{v},2^{\boldsymbol{t}-\boldsymbol{\delta}})|^{2r},$$

where we have possibly enlarged the right-hand side by summing over all possible values for $v_{\boldsymbol{m},\boldsymbol{\delta},i}$ up to 2^{δ_i}. Recalling that for each $\boldsymbol{m}$ we chose $\boldsymbol{k}$ above to be the length at which the maximum is attained, we then have

$$\begin{aligned}\sum_{\substack{\boldsymbol{m}\\ |m_i|\le 2q}} \max_{\boldsymbol{k}\le\boldsymbol{K}} |S(F;\boldsymbol{m},\boldsymbol{k})|^{2r} &\le T^{2r-1} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \sum_{\substack{\boldsymbol{v}\\ 0\le v_i<2^{\delta_i}}} \sum_{\substack{\boldsymbol{m}\\ |m_i|\le 2q}} |S(F;\boldsymbol{m}+2^{\boldsymbol{t}-\boldsymbol{\delta}}\circ\boldsymbol{v},2^{\boldsymbol{t}-\boldsymbol{\delta}})|^{2r}\\ &\ll T^{2r-1} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \|2^{\boldsymbol{\delta}}\| \sum_{\boldsymbol{m}\ (\mathrm{mod}\ q)} |S(F;\boldsymbol{m},2^{\boldsymbol{t}-\boldsymbol{\delta}})|^{2r},\end{aligned}$$

where $T := \prod_{i=1}^{n}(t_i+1)$. Now we perform the key step that accommodates the fact that we only achieve a saving in the smallest direction of the box.

6D. ***Application of the nonmaximal upper bound.*** We would like to apply Lemma 6.1 to the innermost sums over $\boldsymbol{m}$. Fix $\boldsymbol{\delta}$ (with $0\le\delta_i\le t_i$ for each i). Notice that we may not have $t_1-\delta_1\le t_2-\delta_2\le\cdots\le t_n-\delta_n$ even though $t_1\le t_2\le\cdots\le t_n$, so we may need to reorder $\boldsymbol{t}-\boldsymbol{\delta}$ before applying Lemma 6.1. Let σ be a permutation of $\{1,2,\dots,n\}$ (depending on $\boldsymbol{\delta}$) such that

$$t_{\sigma(1)}-\delta_{\sigma(1)} \le t_{\sigma(2)}-\delta_{\sigma(2)} \le \cdots \le t_{\sigma(n)}-\delta_{\sigma(n)}. \tag{6-5}$$

Given an n-tuple $\boldsymbol{x}$, let $\boldsymbol{x}_\sigma = (x_{\sigma(1)},x_{\sigma(2)},\dots,x_{\sigma(n)})$, and $2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma} = (2^{t_{\sigma(1)}-\delta_{\sigma(1)}},\dots,2^{t_{\sigma(n)}-\delta_{\sigma(n)}})$. Recall the discussion on permutations of variables in Section 3D; for any permutation π of $\{1,\dots,n\}$ define the form $F_\pi(\boldsymbol{X})$ by setting $F_\pi(X_1,\dots,X_n)=F(X_{\pi(1)},\dots,X_{\pi(n)})$.

Letting σ^{-1} be the permutation inverse of σ, then

$$S(F;\boldsymbol{m},2^{\boldsymbol{t}-\boldsymbol{\delta}}) = \sum_{\boldsymbol{x}\in(\boldsymbol{m},\boldsymbol{m}+2^{\boldsymbol{t}-\boldsymbol{\delta}})} \chi(F(\boldsymbol{x})) = \sum_{\boldsymbol{x}_\sigma\in(\boldsymbol{m}_\sigma,\boldsymbol{m}_\sigma+2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})} \chi(F_{\sigma^{-1}}(\boldsymbol{x}_\sigma)) = S(F_{\sigma^{-1}};\boldsymbol{m}_\sigma,2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma}).$$

Note that $\boldsymbol{m}_\sigma$ ranges over all n-tuples with coordinates modulo q as $\boldsymbol{m}$ does, and that the last argument in $S(F_{\sigma^{-1}};\boldsymbol{m}_\sigma,2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})$ satisfies the requirement (6-5) so that we may apply Lemma 6.1 with n, r, D, Δ, q, χ as before but now to the form $F_{\sigma^{-1}}$; here we use the uniformity of Lemma 6.1 with respect to the form. (We recall from Lemma 3.5 that the reduction modulo q of $F_{\sigma^{-1}}$ is (Δ,q)-admissible if and only if F is.)

We may conclude (using the fact that $\|2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma}\| = \|2^{\boldsymbol{t}-\boldsymbol{\delta}}\| = \|2^{\boldsymbol{t}}\| \cdot \|2^{\boldsymbol{\delta}}\|^{-1}$) that

$$\begin{aligned}\sum_{\substack{\boldsymbol{m}\\ |m_i|\le 2q}} \max_{\boldsymbol{k}\le \boldsymbol{K}} |S(F;\boldsymbol{m},\boldsymbol{k})|^{2r} &\ll_{n,r,D,\Delta} T^{2r-1} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \|2^{\boldsymbol{\delta}}\| \|2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma}\|^{2r} \sum_{j=0}^{n} q^{(n+j)/2} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1}\\ &= T^{2r-1} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \|2^{\boldsymbol{t}}\|^{2r} \|2^{\boldsymbol{\delta}}\|^{-(2r-1)} \sum_{j=0}^{n} q^{(n+j)/2} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1}\\ &= T^{2r-1} \|2^{\boldsymbol{t}}\|^{2r} \sum_{j=0}^{n} q^{(n+j)/2} \sum_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \|2^{\boldsymbol{\delta}}\|^{-(2r-1)} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1}\\ &\le T^{2r} \|\boldsymbol{K}\|^{2r} \sum_{j=0}^{n} q^{(n+j)/2} \max_{\substack{\boldsymbol{\delta}\\ 0\le\delta_i\le t_i}} \big\{\|2^{\boldsymbol{\delta}}\|^{-(2r-1)} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1}\big\}. \end{aligned} \tag{6-6}$$

Here we recall that $\|2^{\boldsymbol{t}}\| = \|\boldsymbol{K}\|$. Note that in the case of dimension $n=1$ (and $\theta_1 = r$), the sum over j is comprised of the two terms

$$q^{n/2}2^{-\delta(2r-1)} + q^n 2^{-\delta(2r-1)}(2^{t-\delta})^{-r} \le q^{n/2} + q^n 2^{-rt} = q^{n/2} + q^n K^{-r},$$

familiar from the classical 1-dimensional Burgess argument.

Now in general for $n \ge 2$ we must reinterpret $B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1}$ in terms of the coordinates of $\boldsymbol{K} = 2^{\boldsymbol{t}}$, in which we recall that $t_1 \le t_2 \le \cdots \le t_n$; this argument is more complicated, and in particular for $j = n$ we will get a positive power of 2^δ we cannot ignore as in the case of $n = 1$ (e.g., compare to the top line of [Heath-Brown 2013, p. 204]). We summarize the necessary result:

Lemma 6.3. *For $\boldsymbol{K} = 2^{\boldsymbol{t}}$ with $t_1 \le \cdots \le t_n$, for each $\boldsymbol{\delta} \le \boldsymbol{t}$, let σ be a permutation of the indices $\{1,\dots,n\}$ such that $t_{\sigma(1)} - \delta_{\sigma(1)} \le \cdots \le t_{\sigma(n)} - \delta_{\sigma(n)}$. Then*

$$\|2^{\boldsymbol{\delta}}\|^{-(2r-1)} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1} \le \begin{cases} 1 & \text{if } j=0,\\ K_1^{-\theta_j} & \text{if } j=1,\dots,n-1,\\ (K_1\cdots K_{n/2})^{-(2r-1)} & \text{if } j=n,\ n \text{ even},\\ (K_1\cdots K_{(n-1)/2})^{-(2r-1)} K_{(n+1)/2}^{-r} & \text{if } j=n,\ n \text{ odd}. \end{cases}$$

Once we apply this lemma to (6-6), upon noting that $T^{2r} \le 2^{2nr}(\log_2 K_n)^{2nr} \ll (\log K_n)^{2nr}$, we have proved Proposition 6.2.

6E. ***Proof of Lemma 6.3: rearrangement.*** We recall the definition of $B_{n,r}(j;\boldsymbol{k})$ in (4-5) so that for a tuple $\boldsymbol{k} = (k_1,\dots,k_n)$ with $k_1 \le k_2 \le \cdots \le k_n$, $B_{n,r}(j;\boldsymbol{k}) = 1$ for $j=0$, $B_{n,r}(j;\boldsymbol{k}) = k_1^{\theta_j}$ for $j = 1,\dots,n-1$, and for $j=n$ we have $B_{n,r}(j;\boldsymbol{k}) = (k_1\cdots k_{n/2})^{2r}$ if n is even and $B_{n,r}(j;\boldsymbol{k}) = (k_1\cdots k_{(n-1)/2})^{2r} k_{(n+1)/2}^r$ if n is odd.

We may quickly dispatch the case $j=0$, in which case

$$\|2^{\boldsymbol{\delta}}\|^{-(2r-1)} B_{n,r}(j;2^{\boldsymbol{t}_\sigma-\boldsymbol{\delta}_\sigma})^{-1} = \|2^{\boldsymbol{\delta}}\|^{-(2r-1)} \le 1.$$

For the remaining cases of $j \geq 1$, it is helpful to invert, and take the logarithm, and prove for fixed $\boldsymbol{\delta}$ and fixed j a lower bound for the quantity

$$\log_2\bigl(\|2^{\boldsymbol{\delta}}\|^{(2r-1)} B_{n,r}(j; 2^{\mathbf{t}_\sigma - \boldsymbol{\delta}_\sigma})\bigr).$$

First we consider the case of $1 \leq j \leq n-1$; for each of these j (using the fact that $\sum_i \delta_{\sigma(i)} = \sum_i \delta_i$),

$$\begin{aligned}\log_2\bigl(\|2^{\boldsymbol{\delta}}\|^{(2r-1)} B_{n,r}(j; 2^{\mathbf{t}_\sigma - \boldsymbol{\delta}_\sigma})\bigr) &= (2r-1)\biggl(\sum_{i=1}^{n} \delta_{\sigma(i)}\biggr) + \theta_j(t_{\sigma(1)} - \delta_{\sigma(1)}) \\ &= \theta_j t_{\sigma(1)} + (2r-1-\theta_j)\delta_{\sigma(1)} + (2r-1)\sum_{i=2}^{n} \delta_{\sigma(i)} \geq \theta_j t_{\sigma(1)}.\end{aligned}$$

Here we used that for $1 \leq j \leq n-1$ we have $\theta_j \leq r-1$, and moreover $\delta_{\sigma(i)} \geq 0$ for all $1 \leq i \leq n$. Thus

$$\|2^{\boldsymbol{\delta}}\|^{-(2r-1)} B_{n,r}(j; 2^{\mathbf{t}_\sigma - \boldsymbol{\delta}_\sigma})^{-1} \leq 2^{-\theta_j t_{\sigma(1)}} \leq 2^{-\theta_j t_1} = K_1^{-\theta_j},$$

upon recalling that $t_1 \leq t_2 \leq \cdots \leq t_n$.

Now we turn to the more complicated case of $j = n$. First we assume that n is even. Now we have

$$\log_2\bigl(\|2^{\boldsymbol{\delta}}\|^{(2r-1)} B_{n,r}(j; 2^{\mathbf{t}_\sigma - \boldsymbol{\delta}_\sigma})\bigr) = (2r-1)\sum_{i=1}^{n} \delta_{\sigma(i)} + 2r\sum_{i=1}^{n/2}(t_{\sigma(i)} - \delta_{\sigma(i)}).$$

It is convenient to set temporarily for each $i = 1, \ldots, n$ the parameter $\Theta_{n,i} = 2r$ if $1 \leq i \leq n/2$ and $\Theta_{n,i} = 0$ if $n/2 < i \leq n$. Then upon recalling that each $\delta_i \leq t_i$, we have

$$\begin{aligned}&\log_2\bigl(\|2^{\boldsymbol{\delta}}\|^{(2r-1)} B_{n,r}(j; 2^{\mathbf{t}_\sigma - \boldsymbol{\delta}_\sigma})\bigr) \\ &\quad= (2r-1)\sum_{i=1}^{n} \delta_{\sigma(i)} + \sum_{i=1}^{n} \Theta_{n,i}(t_{\sigma(i)} - \delta_{\sigma(i)}) = \sum_{i=1}^{n} \Theta_{n,i} t_{\sigma(i)} + \sum_{i=1}^{n}(2r-1-\Theta_{n,i})\delta_{\sigma(i)} \\ &\quad\geq \sum_{i=1}^{n} \Theta_{n,i} t_{\sigma(i)} + \sum_{\substack{i=1 \\ 2r-1-\Theta_{n,i}<0}}^{n} (2r-1-\Theta_{n,i}) t_{\sigma(i)} + \sum_{\substack{i=1 \\ 2r-1-\Theta_{n,i}\geq 0}}^{n} (2r-1-\Theta_{n,i}) \cdot 0 \\ &\quad= \sum_{\substack{i=1 \\ 2r-1-\Theta_{n,i}<0}}^{n} (2r-1) t_{\sigma(i)} + \sum_{\substack{i=1 \\ 2r-1-\Theta_{n,i}\geq 0}}^{n} \Theta_{n,i} t_{\sigma(i)}. \qquad (6\text{-}7)\end{aligned}$$

(Here in the inequality, equality can occur for those $\boldsymbol{\delta}$ such that $\delta_i = t_i$ for all $1 \leq i \leq n/2$ and $\delta_i = 0$ for $n/2 < i \leq n$. The inequality is where we see that if $\Theta_{n,i} = 2r$, we must replace $\delta_{\sigma(i)}$ by $t_{\sigma(i)}$ rather than by 0; this is why in the final statement of the inequality we lose slightly in the maximal moment, compared to the nonmaximal moment. This effect is not possible in dimension $n = 1$.) Now by the definition of $\Theta_{n,i}$, the first sum is over $1 \leq i \leq n/2$; the second sum is over $n/2 < i \leq n$, in which range $\Theta_{n,i} = 0$ so that the second sum is vacuous. Now we use the following simple observation.

Lemma 6.4 (rearrangement inequality). *Let $t_1 \le t_2 \le \cdots \le t_n$ be a fixed nondecreasing sequence of real numbers and $a_1 \ge \cdots \ge a_n$ a fixed nonincreasing sequence of real numbers. Then for any permutation σ on $\{1, \dots, n\}$, and for any $1 \le M \le n$,*

$$\sum_{i=1}^{M} a_i t_i \le \sum_{i=1}^{M} a_i t_{\sigma(i)}. \tag{6-8}$$

This is a variant of a standard rearrangement inequality; for completeness we give a brief proof in Section 7B. Applying this observation in (6-7) with $M = n/2$, we have shown that

$$\log_2\big(\|2^{\delta}\|^{(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})\big) \ge (2r-1) \sum_{i=1}^{n/2} t_i,$$

so that in the case of $j = n$ even,

$$\|2^{\delta}\|^{-(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})^{-1} \le (2^{t_1} \cdots 2^{t_{n/2}})^{-(2r-1)} = (K_1 \cdots K_{n/2})^{-(2r-1)}.$$

The argument is similar for $j = n$ with n odd, and we only specify the necessary changes, starting with

$$\log_2\big(\|2^{\delta}\|^{(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})\big) = (2r-1) \sum_{i=1}^{n} \delta_{\sigma(i)} + 2r \sum_{i=1}^{(n-1)/2} (t_{\sigma(i)} - \delta_{\sigma(i)}) + r(t_{\sigma((n+1)/2)} - \delta_{\sigma((n+1)/2)}).$$

It is convenient to set temporarily for each $i = 1, \dots, n$ the parameter $\Theta_{n,i} = 2r$ if $1 \le i \le (n-1)/2$, $\Theta_{n,(n+1)/2} = r$ and $\Theta_{n,i} = 0$ if $(n+1)/2 < i \le n$. With this notation, the argument then proceeds as before, until we reach the statement of (6-7), which now holds with this new definition of $\Theta_{n,i}$. Now the first sum on the right-hand side of (6-7) is over $1 \le i \le (n-1)/2$, while the second sum on the right-hand side is over $i \ge (n+1)/2$, and has its only nonzero contribution coming from $i = (n+1)/2$. We may conclude that

$$\log_2\big(\|2^{\delta}\|^{(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})\big) \ge \sum_{i=1}^{(n-1)/2} (2r-1) t_{\sigma(i)} + r t_{\sigma((n+1)/2)}.$$

We now apply (6-8) from Lemma 6.4 to conclude that

$$\log_2\big(\|2^{\delta}\|^{(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})\big) \ge \sum_{i=1}^{(n-1)/2} (2r-1) t_i + r t_{(n+1)/2},$$

or equivalently,

$$\|2^{\delta}\|^{-(2r-1)} B_{n,r}(j; 2^{t_\sigma - \delta_\sigma})^{-1} \le (2^{t_1} \cdots 2^{t_{(n-1)/2}})^{-(2r-1)} (2^{t_{(n+1)/2}})^{-r} = (K_1 \cdots K_{(n-1)/2})^{-(2r-1)} K_{(n+1)/2}^{-r}.$$

This completes the proof of Lemma 6.3.

7. Conclusion of the Burgess argument

We now apply Proposition 6.2 to (5-9) with $\boldsymbol{K} = 2\boldsymbol{H}/P$, recalling that we are working under the assumption (5-1) that $H_1 \le H_2 \le \cdots \le H_n$. (Also recall that $K_n \le H_n < q$.) We conclude that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})|^{2r} \ll_{n,r,\Delta,D} (\log q)^{2r(n+1)} P^{2nr-1} \|\boldsymbol{H}\|^{-1} \left(\frac{\|\boldsymbol{H}\|}{P^n}\right)^{2r} \sum_{j=0}^{n} q^{(n+j)/2} \widetilde{B}_{n,r}(j; \boldsymbol{H}/P)^{-1}$$

$$= (\log q)^{2r(n+1)} \|\boldsymbol{H}\|^{2r-1} P^{-1} q^{n/2} \sum_{j=0}^{n} q^{j/2} \widetilde{B}_{n,r}(j; \boldsymbol{H}/P)^{-1}, \tag{7-1}$$

in which we recall the definition of the function $\widetilde{B}_{n,r}(\,\cdot\,,\,\cdot\,)$ from (6-4).

At this stage of the Burgess argument in the one-dimensional setting $n = 1$, one knows that $\theta_0 = 0$ and $\theta_1 = r$, so that the sum over $j \in \{0, 1\}$ contributes $(1 + q^{1/2}(H_1/P)^{-r})$. To balance this, we would then choose P to be an integer with

$$\tfrac{1}{2} H_1 q^{-1/(2r)} \le P \le H_1 q^{-1/(2r)}. \tag{7-2}$$

Thus when $n = 1$, we recover

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll \|\boldsymbol{H}\|^{1-1/r} q^{(r+1)/4r^2} (\log q)^2,$$

which agrees with Burgess's statement (1-1) up to a logarithm. Now for $n \ge 2$, we observe:

Lemma 7.1. *For $1 \le K_1 \le K_2 \le \cdots \le K_n$,*

$$\sum_{j=0}^{n} q^{j/2} \widetilde{B}_{n,r}(j; \boldsymbol{K})^{-1} \ll_{n,r} 1$$

precisely when

$$q^{1/2} K_1^{-\theta_1} \le 1. \tag{7-3}$$

Under this assumption, the sum over j is dominated by the terms with $j = 0, 1$.

In particular, this lemma (whose proof we defer to Section 7B) shows that the sum is $\gg 1$ if the relation (7-3) does not hold; hence it is advantageous to assume (7-3). Under this assumption, we can dominate the sum by the terms with $j = 0, 1$ and hence we conclude from (7-1) that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})|^{2r} \ll_{n,r,\Delta,D} (\log q)^{2r(n+1)} \|\boldsymbol{H}\|^{2r-1} P^{-1} q^{n/2} (1 + q^{1/2}(H_1/P)^{-\theta_1}).$$

To balance the last two terms within parentheses, we choose P to be an integer with

$$\frac{H_1}{2q^{1/(2\theta_1)}} \le P \le \frac{H_1}{q^{1/(2\theta_1)}}, \tag{7-4}$$

where we recall that $\theta_1 = \lfloor (r-1)/(n-1) \rfloor$ if $n \ge 2$. We recall that earlier in (5-2) and (5-8) we had the requirements that $P \le H_i$ for all i and $H_i P < q$ for all i. The first is clearly true; the second we may verify as long as we assume $H_n H_1 < q^{1+1/(2\theta_1)}$, as we do in our theorem statement.

With this choice for P, we have

$$|S(F; \boldsymbol{N}, \boldsymbol{H})|^{2r} \ll_{n,r,\Delta,D} \|\boldsymbol{H}\|^{2r-1} H_1^{-1} q^{n/2+1/(2\theta_1)} (\log q)^{2r(n+1)},$$

and hence we conclude that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll_{n,r,\Delta,D} \|\boldsymbol{H}\|^{1-1/(2r)} H_1^{-1/(2r)} q^{(n\theta_1+1)/4r\theta_1} (\log q)^{n+1}.$$

This proves Theorem 1.1, upon recalling that we have reduced to the setting in which $H_{\min} = H_1$, $H_{\max} = H_n$, and we have set

$$\Theta = \theta_1 = \left\lfloor \frac{r-1}{n-1} \right\rfloor$$

if $n \geq 2$.

7A. ***Proof of Corollaries 1.3 and 1.4.*** Below, we prove Corollaries 1.3 and 1.4 simultaneously; for the case of Corollary 1.3, simply set $c_0 = 1$ in each instance below. We recall from Theorem 1.1 that

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll \|\boldsymbol{H}\|^{1-1/(2r)} H_{\min}^{-1/(2r)} q^{(n\Theta+1)/4r\Theta} (\log q)^{n+1}, \tag{7-5}$$

where for every $r \geq 1$ we have set

$$\Theta = \Theta_{n,r} = \left\lfloor \frac{r-1}{n-1} \right\rfloor.$$

First let us determine for a given $n \geq 2$ the threshold governing for which lengths $\boldsymbol{H}$ the bound (7-5) is nontrivial, that is $o(\|\boldsymbol{H}\|)$, under the assumption that $\|\boldsymbol{H}\|^{1/n} = q^{\beta}$ and $H_{\min} \gg \|\boldsymbol{H}\|^{c_0/n} = q^{c_0\beta}$. Then the bound (7-5) is nontrivial as long as

$$\left(1 - \frac{1}{2r}\right) n\beta - \frac{c_0}{2r}\beta + \frac{n\Theta+1}{4r\Theta} < n\beta,$$

that is,

$$\beta > \beta_{n,r} := \frac{n\Theta+1}{2\Theta(n+c_0)} = \frac{1}{2} - \frac{c_0\Theta - 1}{2\Theta(n+c_0)}. \tag{7-6}$$

Given $n \geq 2$ and $0 < c_0 \leq 1$, as long as we take r sufficiently large that $\Theta = \Theta_{n,r} > 1/c_0$, we have $\beta_{n,r} < \frac{1}{2}$. (In particular if $c_0 = 1$, note that $\Theta_{n,r} > 1$ when $(r-1)/(n-1) \geq 2$, or equivalently, $r \geq 2n-1$.) On the other hand, note that for fixed $n \geq 2$, for all $r \geq 1$

$$\frac{1}{2} - \frac{c_0\Theta - 1}{2\Theta(n+c_0)} > \frac{1}{2} - \frac{c_0}{2(n+c_0)} \geq \frac{1}{2} - \frac{1}{2(n+1)},$$

for all $0 < c_0 \leq 1$, and this a limitation on the range of β for which the bound is nontrivial.

We now compute that the bound (7-5) is of the form $\|\boldsymbol{H}\| q^{-\delta}$ where

$$\delta = \frac{n+c_0}{2r}\beta - \frac{n\Theta+1}{4r\Theta}. \tag{7-7}$$

We make the approximation that $\Theta = (r-1)/(n-1)$, which is an identity when $n = 2$, and will not be far from the truth, when we later take r very large. Then we compute that after this approximation, δ can be represented as a function of r by

$$f_{a,b,c}(r) = \frac{a}{r} - \frac{br-c}{r(r-1)},$$

where

$$a = \frac{(n+c_0)\beta}{2}, \qquad b = \frac{n}{4}, \qquad c = \frac{1}{4}.$$

As a function of r, this attains a maximum at

$$r = (a-b)^{-1}\{(a-c) \pm \sqrt{(a-c)^2 - (a-c)(a-b)}\}. \tag{7-8}$$

To have $r > 0$ we must have $(a-b) > 0$, that is

$$\beta > \beta_{n,c_0} := \frac{n}{2(n+c_0)} = \frac{1}{2} - \frac{c_0}{2(n+c_0)}, \tag{7-9}$$

agreeing with our previous observation. Thus from now on we assume $\beta = \beta_{n,c_0} + \kappa$ for some small κ, and we will study how $\delta = \delta_n(\kappa)$ behaves as $\kappa \to 0$. From (7-7) with $\beta = \beta_{n,c_0} + \kappa$ (and without approximating Θ), we see that for any r,

$$\delta = \delta_n(\kappa) := \frac{n+c_0}{2r}\kappa - \frac{1}{4r\Theta}.$$

To maximize this, according to (7-8), as $\kappa \to 0$ we will take r to be the integer closest to

$$\frac{n-1}{\kappa(n+c_0)}.$$

For an approximation, we set $r = (n-1)/(\kappa(n+c_0))$ in the expression for $\delta_n(\kappa)$, and using the approximation $\Theta \approx r/(n-1)$ (which is reasonable as $\kappa \to 0$ since then $r \to \infty$), we see that

$$\delta_n(\kappa) \approx \frac{(n+c_0)^2}{4(n-1)}\kappa^2 \tag{7-10}$$

as $\kappa \to 0$. Here we use $\approx$ because we have made two types of approximations: first, we used approximations of the value of Θ and r for simplicity, which introduced errors that become vanishing small, relative to the true size of $\delta_n(\kappa)$, as $\kappa \to 0$. Second, to simplify certain expressions, we used the fact that we are in the limiting case $\kappa \to 0$.

Alternatively, we can encapsulate the restriction (7-9) by recording it as the restriction

$$\|\boldsymbol{H}\| H_{\min} \gg \|\boldsymbol{H}\|^{1+c_0/n} \gg (q^{n\beta})^{1+c_0/n},$$

where $\beta = \beta_{n,c_0} + \kappa$ for any small $\kappa > 0$. Therefore we will obtain a nontrivial bound as long as $\|\boldsymbol{H}\| H_{\min} \gg q^{n/2+\kappa}$ for any small $\kappa > 0$.

7B. ***Proof of technical lemmas.***

Proof of Lemma 5.2. This argument originates in [Heath-Brown 2013, §4] and is similar but not identical to lemmas in [Heath-Brown and Pierce 2015; Pierce 2016]; for completeness we provide an argument. The first property in Lemma 5.2 is a direct result of the definition of $\mathcal{A}(\boldsymbol{m})$. Since each $\mathcal{A}(\boldsymbol{m})$ is a nonnegative integer,

$$\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m}) \le \sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m})^2,$$

and it suffices to prove the third property. We write

$$\begin{aligned}\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m})^2 &= \sum_{\boldsymbol{m}} \#\left\{p,\, p',\, \boldsymbol{a},\, \boldsymbol{a}' : m_i \le \frac{N_i - a_i q}{p} < m_i + H_i/P, m_i \le \frac{N_i - a_i' q}{p'} < m_i + H_i/P\right\} \\ &\ll \left(\prod_{i=1}^n \frac{H_i}{P}\right) \#\left\{p,\, p',\, \boldsymbol{a},\, \boldsymbol{a}' : 0 \le \left|\frac{N_i - a_i q}{p} - \frac{N_i - a_i' q}{p'}\right| \le H_i/P\right\} \\ &= \|\boldsymbol{H}\| P^{-n} \sum_{p,\, p' \in \mathcal{P}} \mathcal{M}(p,\, p'), \end{aligned} \tag{7-11}$$

where

$$\mathcal{M}(p,\, p') = \#\left\{\boldsymbol{a}, \boldsymbol{a}', 0 \le a_i < p, 0 \le a_i' < p' : 0 \le \left|\frac{N_i - a_i q}{p} - \frac{N_i - a_i' q}{p'}\right| \le H_i/P\right\}.$$

First consider $p = p'$. Then

$$\begin{aligned}\mathcal{M}(p,\, p) &\le \#\big\{\boldsymbol{a}, \boldsymbol{a}' : |(N_i - a_i q) - (N_i - a_i' q)| \le p(H_i/P) \le 2H_i,\ i = 1, \dots, n\big\} \\ &\le \#\big\{\boldsymbol{a}, \boldsymbol{a}' : |a_i - a_i'| \le 2H_i/q < 2,\ i = 1, \dots, n\big\}.\end{aligned}$$

Here we have used $H_i < q$. This shows that once $\boldsymbol{a}$ is chosen, there are at most 3^n choices for $\boldsymbol{a}'$, so that $\mathcal{M}(p,\, p) \ll P^n$ and hence $\sum_{p=p' \in \mathcal{P}} \mathcal{M}(p,\, p') \ll P^{n+1}$, which suffices for our desired bound for (7-11).

Next, consider the case $p \ne p'$. For each $i = 1, \dots, n$ we choose (by Bertrand's postulate) a prime l_i such that

$$\frac{q}{H_i} < l_i \le \frac{2q}{H_i}.$$

(Here we use the assumption that $H_i < q$ for each i.) For each i, let $M_i = [N_i l_i/q]$ or $M_i = [N_i l_i/q] + 1$, so that $l_i \nmid M_i$. Then $|N_i l_i/q - M_i| \le 1$ implies that $|N_i - q M_i/l_i| \le q/l_i$, so that

$$\begin{aligned}\mathcal{M}(p,\, p') &\ll \#\left\{\boldsymbol{a}, \boldsymbol{a}' : \left|\frac{q M_i/l_i - a_i q}{p} - \frac{q M_i/l_i - a_i' q}{p'}\right| \le \frac{H_i}{P} + \frac{q}{l_i p} + \frac{q}{l_i p'},\ i = 1, \dots, n\right\} \\ &\ll \#\big\{\boldsymbol{a}, \boldsymbol{a}' : |(p' - p)M_i - (a_i p' - a_i' p) l_i| \le 12P,\ i = 1, \dots, n\big\}.\end{aligned}$$

Given p, p' and an integer δ, for each fixed index i there is at most one way to choose a_i, a_i' with $0 \le a_i < p$, $0 \le a_i' < p'$ such that $a_i p' - a_i' p = \delta$. Thus

$$\sum_{p \ne p' \in \mathcal{P}} \mathcal{M}(p, p') \ll \#\big\{p \ne p' \in \mathcal{P}, \boldsymbol{u} = (u_1, \dots, u_n), |u_i| \le 12P : M_i(p'-p) \equiv u_i \,(\mathrm{mod}\, l_i),\ i = 1, \dots, n\big\}.$$

Now we use the fact that $l_i \nmid M_i$. Thus for a fixed i, the condition $M_i(p' - p) \equiv u_i \pmod{l_i}$ determines $p' - p$ uniquely modulo l_i, and hence uniquely in $\mathbb{Z}$, as long as $P < l_i$, which is guaranteed by the assumption $P < q/H_i$, that is $PH_i < q$. In particular, there is at most one value for the difference $p' - p$ that will satisfy all n conditions. So we may choose p freely and then p' is determined. As a result, after counting up the possible choices for $\boldsymbol{u}$, we conclude that

$$\sum_{p \ne p' \in \mathcal{P}} \mathcal{M}(p, p') \ll P^{n+1}.$$

Applying this in (7-11), we conclude that

$$\sum_{\boldsymbol{m}} \mathcal{A}(\boldsymbol{m})^2 \ll \|\boldsymbol{H}\| P. \qquad \square$$

Proof of Lemma 6.4. We may restrict our attention to permutations that map $\{1, \dots, M\}$ to itself, or equivalently, we may suppose going forward that $M = n$. For indeed, any indices i that occur in the sums such that $i \le M$ but $\sigma(i) > M$, clearly contribute no more to the left-hand side than to the right-hand side, since $t_i \le t_{\sigma(i)}$.

Now let σ be the permutation that minimizes

$$a_1 t_{\sigma(1)} + \cdots + a_n t_{\sigma(n)}; \tag{7-12}$$

if there is more than one such permutation, we choose σ to be the one with the greatest number of fixed points. We will show that σ is the identity. For suppose otherwise, and let i be the smallest index such that $\sigma(i) \ne i$. Then $\sigma(i) > i$ and hence $t_{\sigma(i)} \ge t_i$. Furthermore, denoting by k the index such that $\sigma(k) = i$, we must also have that $k > i$ and hence $a_k \le a_i$. We then see that $(t_{\sigma(i)} - t_i)(a_i - a_k) \ge 0$, or equivalently, $t_{\sigma(i)} a_i + t_i a_k \ge t_i a_i + t_{\sigma(i)} a_k$. Define a new permutation σ' by $\sigma'(u) = u$ for $u = 1, \dots, i$, $\sigma'(k) = \sigma(i)$, and $\sigma'(u) = \sigma(u)$ for all the remaining $u \in \{i+1, \dots, n\} \setminus \{k\}$. Then we see that $t_{\sigma(i)} a_i + t_{\sigma(k)} a_k \ge t_{\sigma'(i)} a_i + t_{\sigma'(k)} a_k$ so that (7-12) does not increase in value if we replace σ by σ', and σ' must also be a minimizer. Yet σ' has one more fixed point than σ, which is a contradiction. We conclude that σ is the identity. $\square$

Proof of Lemma 7.1. By the definition of the θ_j, the sum over $j = 0, 1, \dots, n-1$ takes the form

$$1 + \sum_{j=1}^{n-1} q^{j/2} K_1^{-\theta_j} = 1 + \sum_{j=1}^{n-2} q^{j/2} K_1^{-j\theta_1} + q^{(n-1)/2} K_1^{-(r-1)} \le 1 + \sum_{j=1}^{n-1} (q^{1/2} K_1^{-\theta_1})^j \tag{7-13}$$

in which $\theta_1 = \lfloor (r-1)/(n-1) \rfloor$. Here we have used the fact that for $j = n-1$, $\theta_{n-1} = r - 1 \ge (n-1)\lfloor (r-1)/(n-1) \rfloor = (n-1)\theta_1$. Now we see from the right-most expression that under the

assumption (7-3), all terms $j \geq 2$ are dominated by $j = 0, 1$. On the other hand, we see from the middle expression that if (7-3) does not hold, then that expression is $\gg 1$, as claimed.

It remains to examine the terms with $j = n$, which we divide into the even and odd cases. For $n \geq 2$ even, the $j = n$ term is

$$q^{n/2}(K_1 \cdots K_{n/2})^{-(2r-1)} \leq q^{n/2} K_1^{-(2r-1)n/2} \leq (q^{1/2} K_1^{-\theta_1})^n,$$

where we have used the ordering $K_1 \leq K_2 \leq \cdots \leq K_n$ and the fact that $(2r-1)(n/2) \geq n(r-1)/(n-1) \geq n\theta_1$ holds when $n \geq 2$. Thus the above expression is $\ll 1$ under the assumption (7-3).

For $n \geq 3$ odd, the $j = n$ term is

$$q^{n/2}(K_1 \cdots K_{(n-1)/2})^{-(2r-1)} K_{(n+1)/2}^{-r} \leq q^{n/2} K_1^{-\{(2r-1)(n-1)/2+r\}} \leq (q^{1/2} K_1^{-\theta_1})^n,$$

upon verifying that $n \geq 2$ suffices to show that $(2r-1)(n-1)/2 + r \geq n(r-1)/(n-1) \geq n\theta_1$. Under the assumption (7-3) we see that the $j = n$ term is also $\ll 1$. This concludes the proof of Lemma 7.1. □

Appendix

A.1. ***The role of the stratification.*** It is useful to remark on the crucial role that the stratification plays in obtaining our main theorem. Suppose that instead of Theorem A we only gained information about X_1, without any further stratification into $X_2, \ldots, X_n$. Then for those $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(2r)}) \in \mathbb{A}^{2nr}_{\mathbb{F}_q} \setminus X_1(\mathbb{F}_q)$, we would have

$$|S(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(2r)})| \leq C q^{n/2},$$

but for those $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(2r)}) \in X_1(\mathbb{F}_q)$ we could have an upper bound as bad as the trivial $(\#\mathbb{F}_q)^n$. We would then have only the instance $j = 1$ of Theorem 4.4, namely

$$\#\left\{\{\boldsymbol{x}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r} : \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m}))\Big| > C q^{n/2}\right\} \leq C'' \|\boldsymbol{k}\|^{2r} k_1^{-\theta_1}.$$

In place of (6-1) and Lemma 6.1 we would now have

$$\begin{aligned}\sum_{\boldsymbol{m} \pmod q} |S(F; \boldsymbol{m}, \boldsymbol{k})|^{2r} &\leq \sum_{\substack{\{\boldsymbol{x}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r} \\ \{\boldsymbol{x}\} \in X_0 \setminus X_1}} \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m}))\Big| + \sum_{\substack{\{\boldsymbol{x}\} \in (\boldsymbol{0}, \boldsymbol{k}]^{2r} \\ \{\boldsymbol{x}\} \in X_1}} \Big| \sum_{\boldsymbol{m} \pmod q} \chi(F_{\{\boldsymbol{x}\}}(\boldsymbol{m}))\Big| \\ &\ll \|\boldsymbol{k}\|^{2r} C q^{n/2} + C'' \|\boldsymbol{k}\|^{2r} q^n k_1^{-\theta_1}.\end{aligned}$$

The second term has the worst growth q^n appearing in (6-2) combined with the least savings $k_1^{-\theta_1} = B_{n,r}(1; \boldsymbol{k})^{-1}$. Proceeding with Menshov–Rademacher and the remaining argument, we would obtain

$$|S(F; \boldsymbol{N}, \boldsymbol{H})|^{2r} \ll (\log q)^{2r(n+1)} \|\boldsymbol{H}\|^{2r-1} P^{-1} q^{n/2} (1 + q^{n/2} \widetilde{B}_{n,r}(1; \boldsymbol{H}/P)^{-1}).$$

The last factor is $1+q^{n/2}(H_1/P)^{-\theta_1}$, which we balance by choosing $P = H_1 q^{-n/(2\theta_1)}$. This is a smaller choice than (7-4), hence provides smaller savings; ultimately this yields the bound

$$|S(F; \boldsymbol{N}, \boldsymbol{H})| \ll \|\boldsymbol{H}\|^{1-1/2r} H_1^{-1/2r} q^{(n\theta_1+n)/4r\theta_1} (\log q)^{n+1}. \tag{A-1}$$

This is worse than our main theorem by a factor of $q^{(n-1)/4r\theta_1}$.

A.2. ***Conditional results: assuming a stronger stratification result.*** In our Theorem 1.1, the larger $\Theta = \Theta_{n,r}$ is as a function of r, the better the bound is asymptotically in n. We briefly explore how one could hope to increase the value of $\Theta_{n,r}$. The key is to improve Theorem A, and hence Theorem 4.4, by obtaining larger values for the codimensions θ_j.

At present, Theorem 4.4 holds with $\theta_0 = 0$ and $\theta_n = nr$; for the intermediate values $1 \le j \le n-1$, we currently obtain values

$$\theta_j = j\left\lfloor \frac{r-1}{n-1} \right\rfloor \approx j\left(\frac{r-1}{n-1}\right). \tag{A-2}$$

However, suppose that in the stratification result of Theorem A (and hence in Theorem 4.1 and its corollaries under the modified hypotheses), we were able to take the larger values

$$\theta_j^\sharp = jr, \qquad 1 \le j \le n. \tag{A-3}$$

This is a natural hypothesis since it is the linear interpolation between $\theta_0 = 0$ and $\theta_n = nr$. In fact, note from the definition (A-2) that we very nearly achieve (A-3) in the case of $n = 2$.

Supposing that we can take $\theta_j^\sharp$ as large as in (A-3) in Theorem A, we could deduce that the statement of Theorem 4.4 would hold with the function $B_{n,r}(j; \boldsymbol{k})$ replaced by the modified function $B_{n,r}^\sharp(j; \boldsymbol{k})$ defined for $0 \le j \le n$ and $\boldsymbol{k} = (k_1, \dots, k_n)$ with $k_1 \le k_2 \le \cdots \le k_n$ by

$$B_{n,r}^\sharp(j; \boldsymbol{k}) = \begin{cases} 1 & \text{if } j = 0, \\ (k_1 \cdots k_{j/2})^{2r} & \text{if } j \ge 1 \text{ is even}, \\ (k_1 \cdots k_{(j-1)/2})^{2r} k_{(j+1)/2}^r & \text{if } j \ge 1 \text{ is odd}. \end{cases}$$

Proceeding through the Burgess argument in this paper with the function $B_{n,r}(j; \boldsymbol{k})$ replaced in each instance by $B_{n,r}^\sharp(j; \boldsymbol{k})$, we would arrive at the analogue of (7-1), which now takes the form

$$|S(F; \boldsymbol{N}, \boldsymbol{H})|^{2r} \ll_{n,r,\Delta,D} (\log q)^{2r(n+1)} \|\boldsymbol{H}\|^{2r-1} P^{-1} q^{n/2} \sum_{j=0}^{n} q^{j/2} \widetilde{B}_{n,r}^\sharp(j; \boldsymbol{H}/P)^{-1}, \tag{A-4}$$

in which for any $\boldsymbol{k}$ with $k_1 \le k_2 \le \cdots \le k_n$ we define

$$\widetilde{B}_{n,r}^\sharp(j; \boldsymbol{k}) = \begin{cases} 1 & \text{if } j = 0, \\ (k_1 \cdots k_{j/2})^{2r-1} & \text{if } j \ge 1 \text{ is even}, \\ (k_1 \cdots k_{(j-1)/2})^{2r-1} k_{(j+1)/2}^r & \text{if } j \ge 1 \text{ is odd}. \end{cases}$$

Recall that we assume $H_1 \le H_2 \le \cdots \le H_n$. We choose (cf. (7-2) and (7-4)) P to be an integer such that

$$\frac{H_1}{2q^{1/(2r-1)}} \le P \le \frac{H_1}{q^{1/(2r-1)}}$$

which balances the $j=0$ and $j=2$ contributions. Under this choice, a simple computation shows that for each $j=0,\dots,n$ we can verify that

$$q^{j/2}\widetilde{B}^{\sharp}_{n,r}(j;\boldsymbol{H}/P)^{-1} \le 1,$$

upon using the fact that $H_1 \le H_2 \le \ldots \le H_n$ and the definition of the function $\widetilde{B}^{\sharp}_{n,r}(\,\cdot\,,\,\cdot\,)$ given above. Applying this in (A-4) would give

$$|S(F;\boldsymbol{N},\boldsymbol{H})| \ll_{n,r,\Delta,D} \|\boldsymbol{H}\|^{1-(1/2r)} H_1^{-1/2r} q^{(n\Theta^{\sharp}+1)/4r\Theta^{\sharp}} (\log q)^{n+1}$$

with the definition

$$\Theta^{\sharp} = \frac{2r-1}{2} = r - \frac{1}{2}.$$

We can compute that this is nontrivial for $\boldsymbol{H}$ satisfying the analogues of (1-6) or (1-8) with $\Theta^{\sharp}$ in place of Θ; in the limit as $n\to\infty$ we quantify the strength of the bound near the threshold $\beta_n = \frac{1}{2} - 1/(2(n+1))$. Letting $\beta = \beta_n + \kappa$, then our bound is of the form $\|\boldsymbol{H}\|q^{-\delta}$, where

$$\delta = \frac{n+1}{2r}\kappa - \frac{1}{4r\Theta^{\sharp}} \approx \frac{n+1}{2r}\kappa - \frac{1}{4r^2}.$$

The maximum

$$\delta \approx \frac{(n+1)^2}{4}\kappa^2$$

is achieved when r is the nearest integer to

$$r \approx \frac{1}{\kappa(n+1)}.$$

Thus this conjectural improvement to the stratification would yield a stronger savings near the threshold β_n, but would not alter the fundamental threshold β_n. Similar computations to those above show that even if we could conjecturally improve the θ_j values to the strongest possible values $\theta_j^{\flat} = nr$ for $1 \le j \le n$, this bound will not be improved substantially, and the threshold β_n will not change.

A.3. ***Fourier methods for incomplete sums that are not short.*** Roughly speaking, the threshold $H_i \le q^{1/2}$ appears as a natural constraint of the ranges for which our main results hold. This is not a deficit, for recall that on the other side of this threshold, different methods, which also rely on Weil bounds, become feasible, for appropriate F. To bound $S(F;\boldsymbol{N},\boldsymbol{H})$ in cases where $H_i \gg q^{1/2}$, an advantageous strategy is to "complete the sum," writing

$$S(F;\boldsymbol{N},\boldsymbol{H}) = \sum_{\substack{\boldsymbol{a}=(a_1,\dots,a_n)\\ a_i \pmod q}} \chi(F(\boldsymbol{a})) \sum_{\substack{\boldsymbol{x}\in\mathbb{Z}^n\\ x_i\in(N_i,N_i+H_i]}} \mathbf{1}_{\boldsymbol{x}\equiv\boldsymbol{a} \pmod q}.$$

One then expands the sum over $\boldsymbol{x}$ using

$$\mathbf{1}_{\boldsymbol{x}\equiv\boldsymbol{a} \pmod q} = \frac{1}{q^n} \sum_{\boldsymbol{k} \pmod q} e_q(\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{a}))$$

so that

$$S(F; \boldsymbol{N}, \boldsymbol{H}) = \frac{1}{q^n} \sum_{\boldsymbol{k} \pmod q} \sum_{\boldsymbol{a} \pmod q} \chi(F(\boldsymbol{a})) e_q(-\boldsymbol{k}\cdot\boldsymbol{a}) \Xi_1(a_1/q)\cdots\Xi_n(a_n/q),$$

in which $\Xi(\alpha) = \min\{H_i, \|\alpha\|^{-1}\}$, where $\|\alpha\|$ denotes the distance from α to the nearest integer. One then aims to show that under appropriate assumptions on the smoothness of $F \in \mathbb{F}_q[x_1, \dots, x_n]$, for generic $\boldsymbol{k}$ a Weil bound applies so that the internal sum over $\boldsymbol{a}$ is $O(q^{n/2})$. (Note that this does require the deep input of a Weil-strength bound for a multi-dimensional mixed character sum; see, e.g., [Katz 2006].) The resulting sum over $\boldsymbol{k}$ is then expected to be roughly on the order of $O(q^{-n/2}\|\boldsymbol{H}\| + q^{n/2}(\log q)^n)$, which is $o(\|\boldsymbol{H}\|)$ in the case that $H_i \gg q^{1/2}\log q$ for each $i = 1, \dots, n$, that is, the setting that is complementary to that of this paper. In a hybrid case, in which some H_i are smaller than $q^{1/2}$ and some are larger, one could adopt a hybrid strategy; this regime is closely related to [Pierce 2006; Heath-Brown and Pierce 2012].

Acknowledgements

The authors thank D. R. Heath-Brown and M. Larsen for helpful advice on this project; the authors also thank the referee for helpful suggestions. Pierce is partially supported by NSF CAREER grant DMS-1652173, a Sloan Research Fellowship and the AMS Joan and Joseph Birman Fellowship. Xu has been partially supported by NSF DMS-1702152.

References

[Aggarwal et al. 2018] K. Aggarwal, R. Holowinsky, Y. Lin, and Q. Sun, "The Burgess bound via a trivial delta method", 2018. To appear in *Ramanujan J.* arXiv

[Banks and Guo 2017] W. D. Banks and V. Z. Guo, "Quadratic nonresidues below the Burgess bound", *Int. J. Number Theory* **13**:3 (2017), 751–759. MR Zbl

[Barban et al. 1964] M. B. Barban, Y. V. Linnik, and N. G. Tshudakov, "On prime numbers in an arithmetic progression with a prime-power difference", *Acta Arith.* **9** (1964), 375–390. MR Zbl

[Blomer and Harcos 2008] V. Blomer and G. Harcos, "Hybrid bounds for twisted L-functions", *J. Reine Angew. Math.* **621** (2008), 53–79. MR Zbl

[Bourgain and Chang 2010] J. Bourgain and M.-C. Chang, "On a multilinear character sum of Burgess", *C. R. Math. Acad. Sci. Paris* **348**:3-4 (2010), 115–120. MR Zbl

[Burgess 1957] D. A. Burgess, "The distribution of quadratic residues and non-residues", *Mathematika* **4** (1957), 106–112. MR Zbl

[Burgess 1962] D. A. Burgess, "On character sums and L-series", *Proc. Lond. Math. Soc.* (3) **12** (1962), 193–206. MR Zbl

[Burgess 1963] D. A. Burgess, "On character sums and L-series, II", *Proc. Lond. Math. Soc.* (3) **13** (1963), 524–536. MR Zbl

[Burgess 1967a] D. A. Burgess, "Character sums and primitive roots in finite fields", *Proc. Lond. Math. Soc.* (3) **17** (1967), 11–25. MR Zbl

[Burgess 1967b] D. A. Burgess, "On the quadratic character of a polynomial", *J. Lond. Math. Soc.* **42** (1967), 73–80. MR Zbl

[Burgess 1968] D. A. Burgess, "A note on character sums of binary quadratic forms", *J. Lond. Math. Soc.* **43** (1968), 271–274. MR Zbl

[Burgess 1986] D. A. Burgess, "The character sum estimate with $r = 3$", *J. Lond. Math. Soc.* (2) **33**:2 (1986), 219–226. MR Zbl

[Bykovskii 1996] V. A. Bykovskii, "A trace formula for the scalar product of Hecke series and its applications", *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* **226**:13 (1996), 14–36. In Russian; translated in *J. Math. Sci. (N.Y.)* **89**:1 (1988), 915–932. MR Zbl

[Chang 2008] M.-C. Chang, "On a question of Davenport and Lewis and new character sum bounds in finite fields", *Duke Math. J.* **145**:3 (2008), 409–442. MR Zbl

[Chang 2009] M.-C. Chang, "Burgess inequality in $\mathbb{F}_{p^2}$", *Geom. Funct. Anal.* **19**:4 (2009), 1001–1016. MR Zbl

[Chang 2014] M.-C. Chang, "Short character sums for composite moduli", *J. Anal. Math.* **123** (2014), 1–33. MR Zbl

[Davenport and Lewis 1963] H. Davenport and D. J. Lewis, "Character sums and primitive roots in finite fields", *Rend. Circ. Mat. Palermo* (2) **12** (1963), 129–136. MR Zbl

[Fouvry 2000] É. Fouvry, "Consequences of a result of N. Katz and G. Laumon concerning trigonometric sums", *Israel J. Math.* **120**:A (2000), 81–96. MR Zbl

[Fouvry and Katz 2001] É. Fouvry and N. Katz, "A general stratification theorem for exponential sums, and applications", *J. Reine Angew. Math.* **540** (2001), 115–166. MR Zbl

[Friedlander et al. 2013] J. B. Friedlander, H. Iwaniec, B. Mazur, and K. Rubin, "The spin of prime ideals", *Invent. Math.* **193**:3 (2013), 697–749. Correction in **202**:2 (2015), 923–925. MR Zbl

[Fromm and Goldmakher 2019] E. Fromm and L. Goldmakher, "Improving the Burgess bound via Pólya–Vinogradov", *Proc. Amer. Math. Soc.* **147**:2 (2019), 461–466. MR Zbl

[Gallagher 1972] P. X. Gallagher, "Primes in progressions to prime-power modulus", *Invent. Math.* **16** (1972), 191–201. MR Zbl

[Gallagher and Montgomery 2010] P. X. Gallagher and H. L. Montgomery, "A note on Burgess's estimate", *Mat. Zametki* **88**:3 (2010), 355–364. In Russian; translated in *Math. Notes* **88**:3-4 (2010), 321–329. MR Zbl

[Goldmakher 2010] L. Goldmakher, "Character sums to smooth moduli are small", *Canad. J. Math.* **62**:5 (2010), 1099–1115. MR Zbl

[Graham and Ringrose 1990] S. W. Graham and C. J. Ringrose, "Lower bounds for least quadratic non-residues", pp. 269–309 in *Analytic number theory* (Allerton Park, IL, 1989), edited by B. C. Berndt et al., Progr. Math. **85**, Birkhäuser, Boston, 1990. MR Zbl

[Heath-Brown 1980] D. R. Heath-Brown, "Hybrid bounds for Dirichlet L-functions, II", *Q. J. Math.* (2) **31**:122 (1980), 157–167. MR Zbl

[Heath-Brown 2013] D. R. Heath-Brown, "Burgess's bounds for character sums", pp. 199–213 in *Number theory and related fields*, edited by J. M. Borwein et al., Springer Proc. Math. Stat. **43**, Springer, 2013. MR Zbl

[Heath-Brown 2016] D. R. Heath-Brown, "Small solutions of quadratic congruences, and character sums with binary quadratic forms", *Mathematika* **62**:2 (2016), 551–571. MR Zbl

[Heath-Brown and Pierce 2012] D. R. Heath-Brown and L. B. Pierce, "Counting rational points on smooth cyclic covers", *J. Number Theory* **132**:8 (2012), 1741–1757. MR Zbl

[Heath-Brown and Pierce 2015] D. R. Heath-Brown and L. B. Pierce, "Burgess bounds for short mixed character sums", *J. Lond. Math. Soc.* (2) **91**:3 (2015), 693–708. MR Zbl

[Irving 2016] A. J. Irving, "Estimates for character sums and Dirichlet L-functions to smooth moduli", *Int. Math. Res. Not.* **2016**:15 (2016), 4602–4633. MR Zbl

[Iwaniec 1974] H. Iwaniec, "On zeros of Dirichlet's L series", *Invent. Math.* **23** (1974), 97–104. MR Zbl

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc. Colloq. Publ. **53**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Karatsuba 1970] A. A. Karatsuba, "Estimates of character sums", *Izv. Akad. Nauk SSSR Ser. Mat.* **34**:1 (1970), 20–30. In Russian; translated in *Math. USSR-Uzv.* **4**:1 (1970), 19–29. MR

[Katz 2001] N. M. Katz, "Sums of Betti numbers in arbitrary characteristic", *Finite Fields Appl.* **7**:1 (2001), 29–44. MR Zbl

[Katz 2002] N. M. Katz, "Estimates for nonsingular multiplicative character sums", *Int. Math. Res. Not.* **2002**:7 (2002), 333–349. MR Zbl

[Katz 2006] N. M. Katz, "On a question of Lillian Pierce", *Forum Math.* **18**:4 (2006), 699–710. MR Zbl

[Kerr 2014] B. Kerr, "Some mixed character sums", preprint, 2014. arXiv

[Kerr et al. 2019] B. Kerr, I. E. Shparlinski, and K. H. Yau, "A refinement of the Burgess bound for character sums", *Michigan Math. J.* (online publication November 2019).

[Laumon 2000] G. Laumon, "Exponential sums and l-adic cohomology: a survey", *Israel J. Math.* **120**:A (2000), 225–257. MR Zbl

[Menshov 1923] D. Menshov, "Sur les séries de fonctions orthogonales", *Fund. Math.* **4**:1 (1923), 82–105. Zbl

[Milićević 2016] D. Milićević, "Sub-Weyl subconvexity for Dirichlet L-functions to prime power moduli", *Compos. Math.* **152**:4 (2016), 825–875. MR

[Munshi 2018] R. Munshi, "A note on Burgess bound", pp. 273–289 in *Geometry, algebra, number theory, and their information technology applications* (Toronto, 2016/Kozhikode, India, 2016), edited by A. Akbary and S. Gun, Springer Proc. Math. Stat. **251**, Springer, 2018. MR

[Petrow and Young 2018] I. Petrow and M. P. Young, "The Weyl bound for Dirichlet L-functions of cube-free conductor", preprint, 2018. arXiv

[Petrow and Young 2019] I. Petrow and M. P. Young, "The fourth moment of Dirichlet L-functions along a coset and the Weyl bound", preprint, 2019. arXiv

[Pierce 2006] L. B. Pierce, "A bound for the 3-part of class numbers of quadratic fields by means of the square sieve", *Forum Math.* **18**:4 (2006), 677–698. MR Zbl

[Pierce 2016] L. B. Pierce, "Burgess bounds for multi-dimensional short mixed character sums", *J. Number Theory* **163** (2016), 172–210. MR Zbl

[Pólya 1918] G. Pólya, "Über die Verteilung der quadratischen Reste und Nichtreste", *Gött. Nachr.* **1918** (1918), 21–29. Zbl

[Postnikov 1956] A. G. Postnikov, "On Dirichlet L-series with the character modulus equal to the power of a prime number", *J. Indian Math. Soc.* (*N.S.*) **20** (1956), 217–226. MR Zbl

[Rademacher 1922] H. Rademacher, "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen", *Math. Ann.* **87**:1-2 (1922), 112–138. MR Zbl

[Rojas-León 2005] A. Rojas-León, "Estimates for singular multiplicative character sums", *Int. Math. Res. Not.* **2005**:20 (2005), 1221–1234. MR Zbl

[Rojas-León 2006] A. Rojas-León, "Purity of exponential sums on $\mathbb{A}^n$", *Compos. Math.* **142**:2 (2006), 295–306. MR Zbl

[Tao 2015] T. Tao, "The Elliott–Halberstam conjecture implies the Vinogradov least quadratic nonresidue conjecture", *Algebra Number Theory* **9**:4 (2015), 1005–1034. MR Zbl

[Vinogradov 1918] I. M. Vinogradov, "Sur la distribution des résidus et non résidus de puissances", *Perm. Univ. Fiz.-Mat. Zh.* **1** (1918), 18–24.

[Xu 2018] J. Xu, "Stratification for multiplicative character sums", *Int. Math. Res. Not.* (online publication May 2018).

Communicated by Roger Heath-Brown
Received 2019-07-17 Revised 2019-12-14 Accepted 2020-02-06

pierce@math.duke.edu *Duke University, Durham, NC, United States*

xu56@indiana.edu *Indiana University, Bloomington, IN, United States*

Algebra & Number Theory

msp.org/ant

EDITORS

MANAGING EDITOR

Bjorn Poonen
Massachusetts Institute of Technology
Cambridge, USA

EDITORIAL BOARD CHAIR

David Eisenbud
University of California
Berkeley, USA

BOARD OF EDITORS

Bhargav Bhatt	University of Michigan, USA
Richard E. Borcherds	University of California, Berkeley, USA
Antoine Chambert-Loir	Université Paris-Diderot, France
J-L. Colliot-Thélène	CNRS, Université Paris-Sud, France
Brian D. Conrad	Stanford University, USA
Samit Dasgupta	Duke University, USA
Hélène Esnault	Freie Universität Berlin, Germany
Gavril Farkas	Humboldt Universität zu Berlin, Germany
Hubert Flenner	Ruhr-Universität, Germany
Sergey Fomin	University of Michigan, USA
Edward Frenkel	University of California, Berkeley, USA
Wee Teck Gan	National University of Singapore
Andrew Granville	Université de Montréal, Canada
Ben J. Green	University of Oxford, UK
Joseph Gubeladze	San Francisco State University, USA
Christopher Hacon	University of Utah, USA
Roger Heath-Brown	Oxford University, UK
János Kollár	Princeton University, USA
Philippe Michel	École Polytechnique Fédérale de Lausanne
Susan Montgomery	University of Southern California, USA
Shigefumi Mori	RIMS, Kyoto University, Japan
Martin Olsson	University of California, Berkeley, USA
Raman Parimala	Emory University, USA
Jonathan Pila	University of Oxford, UK
Irena Peeva	Cornell University, USA
Anand Pillay	University of Notre Dame, USA
Michael Rapoport	Universität Bonn, Germany
Victor Reiner	University of Minnesota, USA
Peter Sarnak	Princeton University, USA
Michael Singer	North Carolina State University, USA
Christopher Skinner	Princeton University, USA
Vasudevan Srinivas	Tata Inst. of Fund. Research, India
J. Toby Stafford	University of Michigan, USA
Shunsuke Takagi	University of Tokyo, Japan
Pham Huu Tiep	University of Arizona, USA
Ravi Vakil	Stanford University, USA
Michel van den Bergh	Hasselt University, Belgium
Akshay Venkatesh	Institute for Advanced Study, USA
Marie-France Vignéras	Université Paris VII, France
Melanie Matchett Wood	University of California, Berkeley, USA
Shou-Wu Zhang	Princeton University, USA

PRODUCTION

production@msp.org
Silvio Levy, Scientific Editor

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2020 is US $415/year for the electronic version, and $620/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

© 2020 Mathematical Sciences Publishers

Algebra & Number Theory

Volume 14 No. 7 2020

p-adic Asai *L*-functions of Bianchi modular forms 1669
DAVID LOEFFLER and CHRIS WILLIAMS

Pro-unipotent harmonic actions and dynamical properties of *p*-adic cyclotomic multiple zeta values 1711
DAVID JAROSSAY

Nouvelles cohomologies de Weil en caractéristique positive 1747
JOSEPH AYOUB

Elliptic curves over totally real cubic fields are modular 1791
MAARTEN DERICKX, FILIP NAJMAN and SAMIR SIKSEK

Motivic Gauss–Bonnet formulas 1801
MARC LEVINE and ARPON RAKSIT

Moments of quadratic twists of elliptic curve *L*-functions over function fields 1853
HUNG M. BUI, ALEXANDRA FLOREA, JONATHAN P. KEATING and EDVA RODITTY-GERSHON

Nonvanishing of hyperelliptic zeta functions over finite fields 1895
JORDAN S. ELLENBERG, WANLIN LI and MARK SHUSTERMAN

Burgess bounds for short character sums evaluated at forms 1911
LILLIAN B. PIERCE and JUNYAN XU

Galois action on the principal block and cyclic Sylow subgroups 1953
NOELIA RIZO, A. A. SCHAEFFER FRY and CAROLINA VALLEJO

Abelian extensions in dynamical Galois theory 1981
JESSE ANDREWS and CLAYTON PETSCHE

1937-0652(2020)14:7;1-V