On the Bias of Reed-Muller Codes over Odd Prime Fields

Paul Beame*
University of Washington
beame@cs.washington.edu

Shayan Oveis Gharan†
University of Washington
shayan@cs.washington.edu

Xin Yang*
University of Washington
yx1992@cs.washington.edu

April 6, 2020

Abstract

We study the bias of random bounded-degree polynomials over odd prime fields and show that, with probability exponentially close to 1, n-variate polynomials of degree d over $\mathbb{F}_p$ have bias at most $p^{-\Omega(n/d)}$. This also yields an exponential tail bound on the weight distribution of Reed-Muller codes over odd prime fields. These results generalize bounds of Ben-Eliezer, Hod, and Lovett who proved similar results over $\mathbb{F}_2$. Our bounds are based on an extremal property of the rank of sub-matrices of the generator matrices of Reed-Muller codes over odd prime fields that generalizes a property shown by Keevash and Sudakov for the case of $\mathbb{F}_2$. Our tail bounds on the bias can be used to derive exponential lower bounds on the time for space-bounded learning of bounded-degree polynomials from their evaluations over odd prime fields.

*Research supported in part by NSF grant CCF-1524246
†Research supported in part by NSF grant CCF-1552097 and ONR-YI grant N00014-17-1-2429

1 Introduction

Reed-Muller codes are among the oldest error correcting codes, first introduced by Muller [20] and Reed [22] in the 1950s. These codes were initially defined in terms of bounded-degree multivariate polynomials over $\mathbb{F}_2$ but the same definition can be applied over any finite field. To be more precise, the (d, n) Reed-Muller code over finite field $\mathbb{F}$, denoted $RM_{\mathbb{F}}(d, n)$, takes the message as the coefficients of some n-variate polynomial of degree at most d over $\mathbb{F}$, and the encoding is simply the evaluation of that polynomial over all possible inputs chosen from $\mathbb{F}^n$.

A function $f : \mathbb{F}^n \to \mathbb{F}$ is *balanced* if elements of $\mathbb{F}$ occurs an equal number of times as an output of f. The bias of a function f with co-domain $\mathbb{F}$ is a measure of the fractional deviation of f from being balanced. Since each codeword in a Reed-Muller code is the evaluation of a (polynomial) function over all elements of its domain, the definition of bias directly applies to the codewords of a Reed-Muller code.

Some elements of a Reed-Muller code are very far from balanced (for example the 0 polynomial yields the all-0 codeword, and the codeword for the polynomial $1 + x_1x_2$ has value 1 much more frequently than average) but since, as we might expect, randomly-chosen polynomials behave somewhat like randomly-chosen functions, most codewords are close to being balanced. We quantify that statement and show that for all prime fields, only an exponentially small fraction of Reed-Muller codewords (equivalently, an exponentially small fraction of polynomials of bounded degree) have as much an exponentially small deviation from perfect balance. That is, at most an exponentially small fraction of polynomials have more than an exponentially small bias. Such a result is already known for the case of $\mathbb{F}_2$ [6] so we will only need to prove the statement for odd prime fields.

We now define bias formally and discuss its applications. In the case that $f : \mathbb{F}_2^n \to \mathbb{F}_2$, the *bias* of f,

$$\text{bias}(f) := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)} = \mathbf{Pr}_{x \in_R \mathbb{F}_2^n}[f(x) = 0] - \mathbf{Pr}_{x \in_R \mathbb{F}_2^n}[f(x) = 1].$$

More generally, for p a prime, $\omega = e^{2\pi i/p}$, and $j \in \mathbb{F}_p^*$, we define the *j-th order bias* of $f : \mathbb{F}_p^n \to \mathbb{F}_p$ as

$$\text{bias}_j(f) := \frac{1}{p^n} \sum_{x \in \mathbb{F}_p^n} \omega^{j \cdot f(x)}.$$

Prior uses of bias over these larger co-domains often focus only on the case of a single j (e.g., [8, 13]) since they consider structural implications of bias. However, the use of different values of j is essential for the applications of bias to bounding the imbalance of functions and codewords since, for $p > 3$, one can have functions with 1st-order bias 0 that are very far from balanced. It turns out that it is necessary and sufficient to bound $|\text{bias}_j(f)|$ for all $j \in \mathbb{F}_p^*$ (or, equivalently, all integers j with $1 \leqslant j \leqslant (p-1)/2$ since $|\text{bias}_j(f)| = |\text{bias}_{-j}(f)|$) in order to bound the imbalance: A standard exponential summation argument (e.g., Proposition 2.1 in [3, 4]), shows that for every $b \in \mathbb{F}_p$,

$$\left| \mathbf{Pr}_{x \in_R \mathbb{F}_p^n}[f(x) = b] - \frac{1}{p} \right| \leqslant \max_{j \in \mathbb{F}_p^*} |\text{bias}_j(f)|.$$

For Reed-Muller codes, the bias of a codeword exactly determines its fraction (number) of non-zero entries, which is called the *weight* of the codeword. (In the case of $\mathbb{F}_2$ the bias is determined by the weight but that is not true for $\mathbb{F}_p$ for odd prime p, where the *generalized* weight [14] is required.) The distribution of weights of codewords in Reed-Muller codes over

$\mathbb{F}_2$ plays a critical role in many applications in coding theory and in many other applications in theoretical computer science. As a consequence, the weight distribution of Reed-Muller codes over $\mathbb{F}_2$ has been the subject of considerable study. For degrees $d = 1$ and $d = 2$, the exact weight distribution (and hence the distribution of the bias) for $RM_{\mathbb{F}_2}(2, n)$ has been known for roughly 50 years [23, 19]. For other degrees, precise bounds are only known for weights up to 2.5 times the minimum distance of such codes [15, 16] but this is very far from the balanced regime.

For general constant degrees, Kaufman, Lovett and Porat [17] give a bound on the weight distribution for Reed-Muller codes over $\mathbb{F}_2$, and Abbe, Shpilka, and Wigderson [1] generalize the result to linear degrees. These results yield tail bounds for the number of codewords with bias approaching 0 and, using the cases for arbitrarily small constant bias, imply good bounds for list-decoding algorithms [11, 17].

Ben-Eliezer, Hod, and Lovett [6] proved sharper bounds showing that the fraction of codewords with more than exponentially small bias (of the form $2^{-c_1 n/d}$ for constant $c_1 > 0$) is at most $2^{-c_2 m} = |RM_{\mathbb{F}_2}(d, n)|^{-c_2}$ for constant $c_2 > 0$ where $m = \log_2 |RM_{\mathbb{F}_2}(d, n)|$ is the dimension of the code. (For $d < n/2$ they also showed that this fraction of codewords is tight by exhibiting a set of codewords in $RM_{\mathbb{F}_2}(d, n)$ of size $|RM_{\mathbb{F}_2}(d, n)|^{c_3}$ for $c_3 > 0$ that has such a bias.) This bound was used by [3, 4, 10] to show that learning bounded degree polynomials over $\mathbb{F}_2$ from their evaluations with success probability $2^{-o(n)}$ requires space $\Omega(nm/d)$ or time $2^{\Omega(n/d)}$.

Our Results We generalize the results of Ben-Eliezer, Hod, and Lovett [6] to show that only an exponentially small fraction of polynomials over prime fields can have non-negligible bias. Formally speaking, let $\mathcal{P}_p(d, n)$ denote the set of polynomials of degree at most d in n variables over $\mathbb{F}_p$, and let $\mathcal{M}_p(d, n)$ denote the set of monic monomials of degree at most d in n variables. (The Reed-Muller code $RM_{\mathbb{F}_p}(d, n)$ has dimension $|\mathcal{M}_p(d, n)|$ and satisfies $|RM_{\mathbb{F}_p}(d, n)| = |\mathcal{P}_p(d, n)|$.)

Our main result is the following theorem:

Theorem 1.1. *For any $0 < \delta < 1/2$ there are constants $c_1, c_2 > 0$ depending on δ such that for any odd prime p, for all integers $d \leqslant \delta n$ and all $j \in \mathbb{F}_p^*$, we have*

$$\mathbf{Pr}_{f \in_R \mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)| > p^{-c_1 n/d}] \leqslant p^{-c_2 |\mathcal{M}_p(d,n)|}.$$

In a related paper [3, 4], we show that this theorem can be plugged into a general theorem on learning finite functions to obtain that any algorithm that learns polynomials over $\mathbb{F}_p$ of degree at most d with probability at least $p^{-O(n)}$ from their evaluations on random inputs either requires time $p^{\Omega(n/d)}$ or space $\Omega(n \cdot |\mathcal{M}_p(d, n)|/d \cdot \log p)$. Similar results can be shown to follow by using the theorem above with the methods of [10]. For details, see [3].

The following corollary of Theorem 1.1 is also immediate:

Corollary 1.2. *For any $0 < \delta < 1/2$ there are constants $c_1, c_2 > 0$ such that for any odd prime p and integers d, n with $d \leqslant \delta n$, the number of codewords of $RM_{\mathbb{F}_p}(d, n)$ of weight at most $1 - 1/p - p^{-c_1 n/d}$ is at most $|RM_{\mathbb{F}_p}(d, n)|^{1-c_2}$.*

There is a limit to the amount that Theorem 1.1 can be improved, as shown by the following proposition:

Proposition 1.3. *For any $0 < \delta < 1/2$ there are constants $c' < 1$ and $c'' > 0$ depending on δ such that for all integers $d \leqslant \delta n$ and all $j \in \mathbb{F}_p^*$, we have*

$$\mathbf{Pr}_{f \in_R \mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)| > p^{-c'' n/d}] \geqslant p^{-c' |\mathcal{M}_p(d,n)|}.$$

As part of our proof of Theorem 1.1, we must prove the following tight bound on the rank of the evaluations of monomials of degree at most d on sets of points. Alternatively this can be seen as the extremal dimension of the span of truncated Reed-Muller codes at sizes that are powers of the field size.

Lemma 1.4. *Let S be a subset of $\mathbb{F}_p^n$ such that $|S| = p^r$. Then the dimension of the subspace spanned by $\{(q(x))_{q \in \mathcal{M}_p(d,n)} : x \in S\}$ is at least $|\mathcal{M}_p(d,r)|$.*

Though this is all that we require to prove Theorem 1.1, and can be shown to follow from [14, 5], it is a special case of a more general theorem, which we derive using [5], giving an exact extremal characterization of the dimension of the span of truncated Reed-Muller codes of all sizes[1] and generalizing a characterization for the case of $\mathbb{F}_2$ proved by Keevash and Sudakov [18].

Theorem 1.5. *Let $1 \leqslant m \leqslant p^r$ and let $n \geqslant r$. For $S \subseteq \mathbb{F}_p^n$ with $|S| = m$, the value of*

$$\dim\langle\{(q(x))_{q \in \mathcal{M}_p(d,n)} \ : \ x \in S\}\rangle$$

is minimized when $S = S_m$, the set of m lexicographically minimal vectors in $\mathbb{F}_p^r$.

Proof Overview Our basic approach is a generalization of the high level outline of [6] to odd prime fields, though parts of the argument are substantially more complex:

We begin by using a moment method, showing that that $\mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)|^t]$ is bounded for suitable t. Because we are dealing with odd prime fields rather than $\mathbb{F}_2$ we restrict ourselves to the case that t is even. For bounding these high moments, we reduce the problem to lower bounding the rank of certain random matrices (Lemma 2.4). This is the place where we can apply Lemma 1.4 to prove the bound.

For the case of $\mathbb{F}_2$ handled in [6], a similar property to Lemma 2.4 (Lemma 4 in [6]), which follows from an extremal characterization of $\mathbb{F}_2$ polynomial evaluations by Keevash and Sudakov [18], was independently shown to follow more simply via an algorithmic construction that avoids consideration of any subset size that is not a power of 2. Unfortunately, this simpler algorithmic construction seems to break down completely for the case of odd prime fields.

We instead consider the duality between the rank of sub-matrix of the generating matrix of the truncated Reed-Muller codes, and the maximal number of common zeros for a given number of polynomials over the finite field. The latter problem has been extensively studied in the context of generalized Hamming weights of Reed-Muller codes [14, 5]. Using the results of [5] on the generalized Hamming weights of Reed-Muller codes, we are able to characterize the desired rank property. We also derive an explicit recursive formula for the extremal rank, which may be of independent interest. This recursive formula is the analogue of the formula used by Keevash and Sudakov [18] to obtain their characterization over $\mathbb{F}_2$.

Discussion and Related Work Prior to our work, the main approach to analyzing the bias of polynomials over arbitrary prime fields has been to take a structural point of view. The general idea is to show that polynomials of large bias must have this bias because of some structural property. For polynomials of degree $d = 2$, a complete structural characterization has been known for more than a century ([9]). Green and Tao [12] initiated the modern study of the relationship between the bias and the structure of polynomials over finite fields. Kaufman, Lovett, and Porat [17] used this approach to obtain their bounds on bias over $\mathbb{F}_2$. Over general prime fields, Haramaty and Shpilka [13] gave sharper structural properties for polynomials

[1] In a preliminary version of this paper [2] we had a more complicated direct proof of this characterization.

of degrees $d = 3, 4$. In papers [8] for constant degree and [7] for large degree, Bhowmick and Lovett generalized the result of [17] to show that if a degree d polynomial f has large bias, then f can be expressed as a function of a constant number of polynomials of degree at most $d-1$. These bounds are sufficient to analyze the list-decoding properties of Reed-Muller codes. However, all of these structural results, except for the characterization of degree 2 polynomials, are too weak to obtain the bounds on sub-constant bias that we derive. Indeed, none is sufficient even to derive Corollary 1.2.

An open problem that remains from our work, as well as that of Ben-Eliezer, Hod, and Lovett [6] is whether the amount of the bias can be improved still further by removing the $1/d$ factor from the exponent in the bias in the statement of Theorem 1.1 for some range of values of d growing with n. Though Proposition 1.3 (and its analogue in [6]) show that a large number of polynomials have bias $p^{-O(n/d)}$, we would need to extend them to say that for *all* $c' > 0$ there is a $c'' > 0$ such that the conclusion of the proposition holds in order to rule out improving the bias in Theorem 1.1.

Organization The proof of Theorem 1.1, except for the proof of Lemma 1.4, is in Section 2. Section 2 also contains the proof of Proposition 1.3. In Section 3 we prove Theorem 1.5, which is a generalization of Lemma 1.4.

2 The bias of random polynomials over odd prime fields

In this section we prove Theorem 1.1. To provide tail bounds on the bias, we first characterize its high moments, focusing on even moments to ensure that they are real-valued.

Lemma 2.1. *Let p be an odd prime and $d \leqslant n$. For $t \in \mathbb{N}$, let $x^{(1)}, \cdots, x^{(t)}$ and $y^{(1)}, \cdots, y^{(t)}$ be chosen uniformly at random from $\mathbb{F}_p^n$. Then*

$$\mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\ |\mathrm{bias}_j(f)|^{2t}\ ] = \mathbf{Pr}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\ \forall q \in \mathcal{M}_p(d,n),\ \sum_{k=1}^{t} q(x^{(k)}) = \sum_{k=1}^{t} q(y^{(k)})\ ].$$

Proof. Note that $\overline{\mathrm{bias}_j(f)} = \overline{\mathbf{E}_x[\omega^{j\cdot f(x)}]} = \mathbf{E}_x[\overline{\omega^{j\cdot f(x)}}] = \mathbf{E}_x[\omega^{-j\cdot f(x)}] = \mathrm{bias}_{-j}(f)$, therefore $|\mathrm{bias}_j(f)|^2 = \mathrm{bias}_j(f) \cdot \mathrm{bias}_{-j}(f)$. So we have

$$\begin{aligned}
\mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)|^{2t}] &= \mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\mathrm{bias}_j(f)^t \cdot \mathrm{bias}_{-j}(f)^t] \\
&= \mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\prod_{k=1}^{t} \mathbf{E}_{x^{(k)}}[\omega^{j\cdot f(x^{(k)})}] \cdot \prod_{k=1}^{t} \mathbf{E}_{y^{(k)}}[\omega^{-j\cdot f(y^{(k)})}]] \\
&= \mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\mathbf{E}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\omega^{j\cdot(\sum_{k=1}^{t} f(x^{(k)}) - \sum_{k=1}^{t} f(y^{(k)}))}]] \\
&= \mathbf{E}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\omega^{j\cdot(\sum_{k=1}^{t} f(x^{(k)}) - \sum_{k=1}^{t} f(y^{(k)}))}]]
\end{aligned}$$

For each $q \in \mathcal{M}_p(d,n)$ let $f_q \in \mathbb{F}_p$ denote the coefficient of q in f. We identify f with its vector

of coefficients $(f_q)_{q\in\mathcal{M}_p(d,n)}$ and choose f uniformly by choosing the f_q uniformly. Therefore

$$\begin{aligned}
\mathbf{E}_{f\in_R\mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)|^{2t}] &= \mathbf{E}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\mathbf{E}_{f\in_R\mathcal{P}_p(d,n)}[\omega^{j\cdot(\sum_{q\in\mathcal{M}_p(d,n)} f_q\cdot(\sum_{k=1}^t q(x^{(k)})-\sum_{k=1}^t q(y^{(k)})))}]] \\
&= \mathbf{E}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\prod_{q\in\mathcal{M}_p(d,n)} \mathbf{E}_{f_q\in_R\mathbb{F}_p}[\omega^{j\cdot f_q\cdot(\sum_{k=1}^t q(x^{(k)})-\sum_{k=1}^t q(y^{(k)}))}]] \\
&= \mathbf{E}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\mathbf{1}_{(\forall q\in\mathcal{M}_p(d,n),\ \sum_{k=1}^t q(x^{(k)})-\sum_{k=1}^t q(y^{(k)})=0)}] \\
&= \mathbf{Pr}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\forall q\in\mathcal{M}_p(d,n),\ \sum_{k=1}^t q(x^{(k)}) = \sum_{k=1}^t q(y^{(k)})]
\end{aligned}$$

where the second equality follows since $\mathbf{E}_{a\in_R\mathbb{F}_p}[\omega^{j\cdot a\cdot b}] = 0$ for all $b\in\mathbb{F}_p^*$. □

Now let us look at the probability

$$\mathbf{Pr}_{x^{(1)},\cdots,x^{(t)},y^{(1)},\cdots,y^{(t)}}[\forall q\in\mathcal{M}_p(d,n),\ \sum_{k=1}^t q(x^{(k)}) = \sum_{k=1}^t q(y^{(k)})].$$

We view $y^{(1)},\cdots,y^{(t)}$ as arbitrary fixed values and we will upper bound this probability following the analysis of a similar probability in [6]. That is, we will upper bound the probability that this holds by considering a special subset $\mathcal{M}'\subseteq\mathcal{M}_p(d,n)$ that allows us to derive a linear system whose rank will bound the probability that the constraints indexed by $\mathcal{M}'$ all hold.

We divide $[n]$ arbitrarily into two disjoint parts L and R with $|L| = \lfloor\frac{n}{d}\rfloor$. $\mathcal{M}'\subseteq\mathcal{M}_p(d,n)$ consists of all monomials of degree are most d that have degree 1 on L and degree at most $d-1$ on R.

We use the following properties of the $|\mathcal{M}_p(d,n)|$, whose proof we defer to later, to show that $\mathcal{M}'$ contains a significant fraction of all monomials in $\mathcal{M}_p(d,n)$.

Proposition 2.2. *If $2\leqslant d\leqslant\delta n$ for some $0<\delta<1$ then*

(a) there exists a constant $\gamma'=\gamma'(\delta)>0$ such that for sufficiently large n, if $n'\geqslant\max(d,(1-\frac{1}{d})n)$ then

$$|\mathcal{M}_p(d,n')|\geqslant\gamma'|\mathcal{M}_p(d,n)|.$$

(b) If $p\geqslant 3$ there exist constants $\rho_1,\rho_2>0$ that are independent of δ such that for sufficiently large n,

$$\rho_1|\mathcal{M}_p(d,n)|\leqslant\frac{n}{d}\cdot|\mathcal{M}_p(d-1,n)|\leqslant\rho_2|\mathcal{M}_p(d,n)|.$$

Corollary 2.3. *Let $p\geqslant 3$. If $d\leqslant\delta n$ for some $0<\delta<1$, then there exists a constant $\gamma=\gamma(\delta)>0$ such that for sufficiently large n,*

$$|\mathcal{M}'| = \lfloor\frac{n}{d}\rfloor\cdot|\mathcal{M}_p(d-1,n-\lfloor\frac{n}{d}\rfloor)|\geqslant\gamma\cdot|\mathcal{M}_p(d,n)|.$$

Proof. The equality follows immediately from the definition of $\mathcal{M}'$. Let $n' = n-\lfloor\frac{n}{d}\rfloor$. Then

$$\begin{aligned}
|\mathcal{M}'| &= \lfloor\frac{n}{d}\rfloor\cdot|\mathcal{M}_p(d-1,n')| \\
&\geqslant \frac{n'}{2d}|\mathcal{M}_p(d-1,n')| \qquad \text{since } d\leqslant n \\
&\geqslant \frac{\rho_1}{2}|\mathcal{M}_p(d,n')| \qquad \text{by Proposition 2.2(b)} \\
&\geqslant \frac{\rho_1\gamma'}{2}|\mathcal{M}_p(d,n)| \qquad \text{by Proposition 2.2(a)}
\end{aligned}$$

and setting $\gamma=\rho_1\gamma'/2$ yields the claim. □

Let $\mathcal{E}$ denote the event that $\sum_{k=1}^{t} q(x^{(k)}) = \sum_{k=1}^{t} q(y^{(k)})$ for all $q \in \mathcal{M}'$. To simply notation, since we think of $y^{(1)}, \ldots, y^{(k)}$ as fixed, for each $q \in \mathcal{M}'$ define $b_q \in \mathbb{F}_p$ by $b_q = \sum_{k=1}^{t} q(y^{(k)})$. Since any $q \in \mathcal{M}'$ is of the form $q = x_i \cdot q'$ for some $i \in L$ and q' a monomial of degree at most $d-1$ on R, $\mathcal{E}$ requires that

$$b_q = \sum_{k=1}^{t} q(x^{(k)}) = \sum_{k=1}^{t} q'(x_R^{(k)}) \cdot x_i^{(k)}.$$

where for $x \in \mathbb{F}_p^n$, we write x_R for x restricted to the coordinates in R. We view these constraints as a system of linear equations over the set of variables $x_i^{(k)}$ for $k \in [t]$ and $i \in L$ whose coefficients are given by the values of $q'(x_R^{(k)})$ for $x_R^{(k)} \in \mathbb{F}_p^R$ for all $k \in [t]$. Observe that for different values of $i \in L$ we get separate and independent subsystems of equations with precisely the same coefficients but potentially different constant terms b_q since q depends on both i and q'. Therefore the probability that $(x_i^{(k)})_{i \in L, k \in [t]}$ is a solution is the product of the probabilities for the individual choices of $i \in L$.

For each $\mathbf{x}_R = x_R^{(1)}, \ldots, x_R^{(t)}$, there is a $|\mathcal{M}_p(d-1, R)| \times t$ matrix $Q_{\mathbf{x}_R}$ for a system of linear equations on $(x_i^{(1)}, \ldots, x_i^{(t)})$ for each $i \in L$, having one constraint for each polynomial q' of degree at most $d-1$ on R. Observe that $Q_{\mathbf{x}_R}(q', k) = q'(x_R^{(k)})$.

In particular, it follows that

$$\mathbf{Pr}_{x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}}[\, \mathcal{E} \mid (x_R^{(1)}, \cdots, x_R^{(t)}) = \mathbf{x}_R] \leqslant p^{-rank(Q_{\mathbf{x}_R}) \cdot |L|}. \tag{1}$$

We now see that for almost all choices of $\mathbf{x}_R$, if t is at least a constant factor larger than $|\mathcal{M}_p(d-1, |R|)|$ then the rank of $Q_{\mathbf{x}_R}$ is large. This follows by replacing n by $|R|$, d by $d-1$, q' by q and $\mathbf{x}$ by $\mathbf{x}_R$ in the following lemma.

Lemma 2.4. *For any $0 < \delta \leqslant 1/2$ there is a constant $\gamma = \gamma(\delta) > 0$ such that there exist constants $c > 0$ and $\eta > 1$ such that for $d = \lfloor \delta n \rfloor$ and $t \geqslant \eta |\mathcal{M}_p(d, n)|$, if $\mathbf{x} = x^{(1)}, \cdots, x^{(t)}$ is chosen uniformly at random from $(\mathbb{F}_p^n)^t$, then the matrix $Q_{\mathbf{x}} \in \mathbb{F}_p^{\mathcal{M}_p(d,n) \times [t]}$ given by $Q_{\mathbf{x}}(q, k) = q(x^{(k)})$. then*

$$\mathbf{Pr}_{\mathbf{x}}[\, rank(Q_{\mathbf{x}}) \leqslant \gamma |\mathcal{M}_p(d, n)| \,] \leqslant p^{-c|\mathcal{M}_p(d+1, n)|}.$$

We first show how to use Lemma 2.4 to prove Theorem 1.1.

Proof of Theorem 1.1. Let $0 < \delta \leqslant 1/2$, and set $\gamma > 0$ and $\eta > 1$ and $c > 0$ as in Lemma 2.4. Let $t = \lceil \eta \mathcal{M}_p(d-1, n) \rceil$. We first bound the expected value of $|\mathrm{bias}_j(f)|^{2t}$. By Lemma 2.1 and the definition of event $\mathcal{E}$ we have

$$\mathbf{E}_{f \in_R \mathcal{P}_p(d,n)}[\, |\mathrm{bias}_j(f)|^{2t} \,] = \mathbf{Pr}_{x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}}[\, \mathcal{E} \,].$$

Let $n' = \lceil n(1 - 1/d) \rceil$ and $d' = d - 1$. Let $\mathcal{A}$ be the event that given $x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}$, $rank(Q_{\mathbf{x}_R}) \leqslant \gamma |\mathcal{M}_p(d', n')|$, and $\overline{\mathcal{A}}$ be the complement event of $\mathcal{A}$. Now we have,

$$\begin{aligned}
&\mathbf{Pr}_{x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}}[\, \mathcal{E} \,] \\
&\leqslant \mathbf{Pr}_{x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}}[\mathcal{A}] + \mathbf{Pr}_{x^{(1)}, \cdots, x^{(t)}, y^{(1)}, \cdot, y^{(t)}}[\, \mathcal{E} \mid \overline{\mathcal{A}} \,] \\
&\leqslant \mathbf{Pr}_{\mathbf{x_R}}[rank(Q_{\mathbf{x}_R}) \leqslant \gamma |\mathcal{M}_p(d', n')|] + p^{-\gamma |\mathcal{M}_p(d', n')| \cdot |L|}
\end{aligned}$$

where the last step follows from Eq. (1) and the condition $rank(Q_{\mathbf{x}_R}) > \gamma|\mathcal{M}_p(d', n')|$. Observe that $t \geqslant \eta|\mathcal{M}_p(d', n)| \geqslant \eta|\mathcal{M}_p(d', n')|$ so we can apply Lemma 2.4 with $(n', d' = d-1)$ in place of (n, d), and $\mathbf{x} = \mathbf{x}_R$ to derive that

$$\mathbf{Pr}_{\mathbf{x_R}}[rank(Q_{\mathbf{x}_R}) \leqslant \gamma|\mathcal{M}_p(d-1, n')|] \leqslant p^{-c|\mathcal{M}_p(d,n')|}.$$

Therefore,

$$\mathbf{E}_{f\in_R\mathcal{P}_p(d,n)}[\ |\mathrm{bias}_j(f)|^{2t}\ ] \leqslant p^{-c|\mathcal{M}_p(d,n')|} + p^{-\gamma|\mathcal{M}_p(d-1,n')|\cdot|L|}. \tag{2}$$

Now, for sufficiently large n, by Proposition 2.2(a), $|\mathcal{M}_p(d, n')| \geqslant \gamma'|\mathcal{M}_p(d, n)|$ and by Corollary 2.3 $|\mathcal{M}_p(d-1, n')| \cdot |L| \geqslant \gamma|\mathcal{M}_p(d, n)|$. Therefore,

$$\mathbf{E}_{f\in_R\mathcal{P}_p(d,n)}[\ |\mathrm{bias}_j(f)|^{2t}\ ] \leqslant p^{-c\gamma'|\mathcal{M}_p(d,n)|} + p^{-\gamma^2|\mathcal{M}_p(d,n)|} \geqslant p^{-c'|\mathcal{M}_p(d,n)|}$$

for some constant $c' > 0$. Now we can apply Markov's inequality to obtain that for any $c_1 > 0$.

$$\begin{aligned}\mathbf{Pr}_{f\in_R\mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)| > p^{-c_1 n/d}] &= \mathbf{Pr}_{f\in_R\mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)|^{2t} > p^{-2t\cdot c_1 n/d}] \\ &\leqslant \frac{p^{-c'|\mathcal{M}_p(d,n)|}}{p^{-2t\cdot c_1 n/d}} \\ &= p^{2t\cdot c_1 n/d - c'|\mathcal{M}_p(d,n)|}\end{aligned}$$

By definition, $t = \lceil\eta|\mathcal{M}_p(d-1, n)|\rceil \leqslant \eta'|\mathcal{M}_p(d-1, n)|$ for a fixed $\eta' > \eta$. Therefore, by Proposition 2.2(b), $2tn/d \leqslant 2\eta'\rho_2|\mathcal{M}_p(d, n)|$. By choosing $c_1 = c'/(4\eta'\rho_2)$, we obtain that $2t \cdot c_1 n/d - c'|\mathcal{M}_p(d, n)| \leqslant -c'|\mathcal{M}_p(d, n)|/2$ and setting $c_2 = c'/2$ we derive that

$$\mathbf{Pr}_{f\in_R\mathcal{P}_p(d,n)}[|\mathrm{bias}_j(f)| > p^{-c_1 n/d}] \leqslant p^{-c_2|\mathcal{M}_p(d,n)|}$$

as required. □

It remains to prove Lemma 2.4 and Proposition 2.2. We first prove Lemma 2.4 using Lemma 1.4, which follows from Theorem 1.5 whose proof is in the next section. Lemma 2.4 is a generalization of Claim 2.3 in [6], and its proof follows the same lines as that in [6].

Proof of Lemma 2.4 using Lemma 1.4. Let $d = \lfloor\delta n\rfloor$ for $0 < \delta \leqslant 1/2$ and let $\gamma > 0$ be the minimum of $\gamma'(\delta)$ from Proposition 2.2 and $\gamma(\delta)$ from Corollary 2.3. Fix $b = \lfloor\gamma \cdot |\mathcal{M}_p(d, n)|\rfloor$. We will first check the probability that an arbitrary fixed set of b columns spans the whole matrix, and then apply a union bound to obtain the final result.

Let V denote the linear space spanned by those b columns. Recall that each column of Q_{x_R} is the evaluation of all monomials of degree at most d at some point $\mathbb{F}_p^n$.

Let integer r be maximal such that there are at least p^r distinct elements of $\mathbb{F}_p^n$ with evaluations that are in V. Then by Lemma 1.4, we have $\dim(V) \geqslant |\mathcal{M}_p(d, r)|$. But since V can be spanned by b vectors, we have

$$\gamma|\mathcal{M}_p(d, n)| \geqslant b \geqslant \dim(V) \geqslant |\mathcal{M}_p(d, r)|$$

By Proposition 2.2(a), we have

$$|\mathcal{M}_p(d, \lceil n(1-1/d)\rceil)| \geqslant \gamma|\mathcal{M}_p(d, n)| \geqslant |\mathcal{M}_p(d, r)|$$

So $r \leqslant \lceil n(1-1/d)\rceil$. There are p^n distinct evaluations and fewer than p^{r+1} of them fall into V. So a uniform random evaluation is in V with probability $< \frac{p^{r+1}}{p^n} \leqslant p^{1-\lfloor n/d\rfloor}$. Since the

$t-b$ other columns of Q_{x_R} are chosen uniformly and independently, the probability that these b columns span the whole matrix is at most

$$(p^{1-\lfloor n/d \rfloor})^{t-b} \leqslant (p^{1-\lfloor n/d \rfloor})^{(\eta-\gamma)|\mathcal{M}_p(d,n)|}$$

since $t = \eta|\mathcal{M}_p(d,n)|$ for some $\eta > 1$ to be chosen later. Since $d \leqslant \delta n \leqslant n/2$, we have $1 - \lfloor n/d \rfloor \leqslant -n/(2d)$ and we can apply Proposition 2.2 to get that

$$(p^{1-\lfloor n/d \rfloor})^{t-b} \leqslant p^{-(\eta-\gamma)\frac{n}{d}\cdot|\mathcal{M}_p(d,n)|/2} \leqslant p^{-(\eta-\gamma)\rho_1|\mathcal{M}_p(d+1,n)|/2}$$

for some $\rho_1 > 0$. Therefore, by a union bound over all choices of b columns we have

$$\mathbf{Pr}_{x^{(1)},\cdots,x^{(t)}}[rank(Q_{x_R}) \leqslant \gamma|\mathcal{M}_p(d,n)|] \leqslant \binom{t}{b} \cdot p^{-(\eta-\gamma)\rho_1|\mathcal{M}_p(d+1,n)|/2}.$$

Note that $\binom{t}{b} \leqslant (\frac{te}{b})^b \leqslant (\frac{2e\eta}{\gamma})^{\gamma|\mathcal{M}_p(d,n)|} \leqslant (\frac{2e\eta}{\gamma})^{\gamma|\mathcal{M}_p(d+1,n)|}$, so we have

$$\mathbf{Pr}_{x^{(1)},\cdots,x^{(t)}}[rank(Q_{x_R}) \leqslant \gamma|\mathcal{M}_p(d,n)|] \leqslant p^{|\mathcal{M}_p(d+1,n)|(\gamma\log_p(\frac{2e\eta}{\gamma})-(\eta-\gamma)\rho_1/2)}$$

Note that for any constant $c' > 0$, $\gamma\log_p(c'\eta)$ is $o(\eta)$. Therefore, for fixed constant $\gamma > 0$, we can choose a sufficiently large $\eta > 1$ such that

$$\mathbf{Pr}_{x^{(1)},\cdots,x^{(t)}}[rank(Q_{x_R}) \leqslant \gamma|\mathcal{M}_p(d,n)|] \leqslant p^{-c|\mathcal{M}_p(d+1,n)|}$$

for some constant $c > 0$. □

2.1 Proof of Proposition 2.2

We first give basic inequalities regarding $|\mathcal{M}_p(d,n)|$ that are independent of the choice of p.

Proposition 2.5. *For $d \leqslant n$, $\sum_{i=0}^{d}\binom{n}{i} \leqslant |\mathcal{M}_p(d,n)| \leqslant \binom{n+d}{d}$.*

Proof. It is well known that there are $\binom{n+d}{d}$ non-negative integer solutions to the equation $\sum_{i=1}^{n} e_i \leqslant d$. Thus we have $|\mathcal{M}_p(d,n)| \leqslant \binom{n+d}{d}$. On the other hand, if we only consider multi-linear terms, we obtain $\sum_{i=0}^{d}\binom{n}{i} \leqslant |\mathcal{M}_p(d,n)|$. □

We now prove part (a): For $\mathbf{e} = (e_1,\cdots,e_k)$ where $1 \leqslant e_i \leqslant p-1$, let $\mathcal{M}_{\mathbf{e},n}$ denote the set of monomials of the form $\prod_{i=1}^{k} x_{h(i)}^{e_i}$, $1 \leqslant h(1) < h(2) < \cdots < h(k) \leqslant n$. Then we have $|\mathcal{M}_p(d,n)| = \sum_{\mathbf{e}:\sum_i e_i \leqslant d}|\mathcal{M}_{\mathbf{e},n}|$. Therefore

$$\frac{|\mathcal{M}_p(d,n')|}{|\mathcal{M}_p(d,n)|} \geqslant \min_{\mathbf{e}:\sum_i e_i \leqslant d}\frac{|\mathcal{M}_{\mathbf{e},n'}|}{|\mathcal{M}_{\mathbf{e},n}|}$$

For fixed $\mathbf{e} = (e_1,\cdots,e_k)$, We then argue that for all integer n, $|\mathcal{M}_{\mathbf{e},n}| = f_{\mathbf{e}}\cdot\binom{n}{k}$, for some $f_{\mathbf{e}} > 0$ that only depends on $\mathbf{e}$. This is because there are $\binom{n}{k}$ different ways to choose the k variables appearing in the monomial, and each of them are indistinguishable, hence $|\mathcal{M}_{\mathbf{e},n}|$ is proportional to $\binom{n}{k}$.

Therefore, we have

$$\frac{|\mathcal{M}_{\mathbf{e},n'}|}{|\mathcal{M}_{\mathbf{e},n}|} = \frac{\binom{n'}{k}}{\binom{n}{k}}$$

This quantity is a decreasing function of k. Hence we have

$$\frac{|\mathcal{M}_p(d,n')|}{|\mathcal{M}_p(d,n)|} \geqslant \frac{\binom{n'}{d}}{\binom{n}{d}} = \prod_{i=0}^{d-1} \frac{n'-i}{n-i}$$

From the well known inequality $\ln(1+x) \geqslant \frac{x}{1+x}$, we have

$$\begin{aligned}\ln \prod_{i=0}^{d-1} \frac{n'-i}{n-i} = \sum_{i=0}^{d-1} \ln \frac{n'-i}{n-i} \geqslant \sum_{i=0}^{d-1} \frac{n'-n}{n'-i} = -(n-n') \sum_{i=0}^{d-1} \frac{1}{n'-i} \\ \geqslant -(n-n') \int_{n'-d}^{n'} x^{-1} \mathbf{d}x = -(n-n') \ln \frac{n'}{n'-d} \geqslant -\frac{d(n-n')}{n'-d}\end{aligned}$$

By the condition $n' \geqslant (1-1/d)n$ and $d \leqslant \delta n$, we have

$$\begin{aligned}\prod_{i=0}^{d-1} \frac{n'-i}{n-i} &\geqslant \exp(-\frac{d(n-(1-1/d)n)}{n'-\delta n}) = \exp(-\frac{n}{n'-\delta n}) \\ &\geqslant \exp(-\frac{1}{1-1/d-\delta}) \geqslant \exp(-\frac{1}{1/2-\delta}) := \gamma'(\delta),\end{aligned}$$

which completes the proof of part (a).

We now prove part (b): Define $H := \{(q_1, q_2) : q_1 \in \mathcal{M}_p(d-1,n), q_2 \in \mathcal{M}_p(d,n), \exists i \in [n] \text{ s.t. } q_2 = x_i q_1\}$. We will obtain the inequalities by bounding $|H|$.

We first bound $|H|$ in terms of $|\mathcal{M}_p(d-1,n)|$. Clearly for each $q_1 \in \mathcal{M}_p(d-1,n)$, there are at most n choices of x_i to yield $q_2 = x_i q_1$, so $|H| \leqslant n|\mathcal{M}_p(d-1,n)|$. On the other hand, any x_i that does not have degree $p-1$ in q_1 can be chosen. There are at most $\frac{d-1}{p-1}$ variables in q_1 having degree $p-1$ so we can choose at least $n - \frac{d-1}{p-1} > n - \frac{n}{p-1} \geqslant \frac{n}{2}$ variables x_i since $p \geqslant 3$. This gives us $|H| > \frac{n}{2}|\mathcal{M}_p(d-1,n)|$.

We now bound $|H|$ in terms of $|\mathcal{M}_p(d,n)|$. Each $q \in \mathcal{M}_p(d,n)$ contains at most d distinct variables, hence $|H| \leqslant d|\mathcal{M}_p(d,n)|$.

It immediately follows that $\mathcal{M}_p(d,n) \geqslant |H|/d \geqslant \frac{n}{2d}|\mathcal{M}_p(d-1,|)$ and hence we can choose $\rho_2 = 2$.

To lower bound $|H|$ in terms of $|\mathcal{M}_p(d,n)|$, we show that a large portion of monomials contain many distinct variables and hence each $q_2 \in \mathcal{M}_p(d,n)$ can be associated with many different q_1. We first bound the number of monomials that have degree at most d and are composed of at most $k \leqslant d$ distinct variables. We can generate such monomials by first choosing k variables, then using these variables to form a monomial of degree $\leqslant d$ and so we can upper bound the number of such monomials by $\binom{n}{k}\binom{k+d}{d}$. For sufficiently small k we can argue that this is a small fraction of $\mathcal{M}_p(d,n)$: Suppose that $k \leqslant d/6$. Since by hypothesis, $d \leqslant \delta n \leqslant n/2$, we have $k+d \leqslant n-k$ and

$$\begin{aligned}\frac{\binom{n}{k}\binom{k+d}{d}}{|\mathcal{M}_p(d,n)|} &\leqslant \frac{\binom{n}{k}\binom{k+d}{d}}{\binom{n}{d}} \qquad \text{by Proposition 2.5} \\ &= \frac{(k+d)!}{(k!)^2(n-k)\cdots(n-d+1)} = \frac{(k+d)\cdots(2k+1)}{(n-k)\cdots(n-d+1)} \cdot \binom{2k}{k} \\ &\leqslant \left(\frac{k+d}{n-k}\right)^{d-k} \cdot 2^{2k} \leqslant (7/11)^{5k} \cdot 2^{2k} \leqslant (3/7)^k.\end{aligned}$$

Summing over all values of $k \leqslant d/6$ we obtain that a total fraction at most $3/4$ of all monomials in $\mathcal{M}_p(d,n)$ have at most $d/6$ distinct variables. Therefore, since at least $1/4$ of $\mathcal{M}_p(d,n)$

contain at least $d/6$ distinct variables, it must be the case that $|H| \geqslant \frac{d}{24} \cdot |\mathcal{M}_p(d,n)|$. Since $|H| \leqslant n \cdot |\mathcal{M}_p(d-1,n)|$, we obtain that $|\mathcal{M}_p(d,n)|/24 \leqslant \frac{n}{d}|\mathcal{M}_p(d-1,n)|$. Hence we derive (b) with $\rho_1 = 1/24$. This completes the proof of Proposition 2.2(b).

2.2 Lower bound on the likelihood of bias

We now prove Proposition 1.3, on the limits on the extent to which Theorem 1.1 can be improved. The argument is analogous to that of [6] for the case of $\mathbb{F}_2$.

Proof of Proposition 1.3. We follow the same division of variables $[n]$ into parts L and R with $|L| = \lfloor \frac{n}{d} \rfloor$ and $|R| = n' = \lceil n(1-1/d) \rceil$ and $d' = d-1$ that we used for the upper bound on the bias. Define $\mathcal{L}$ to be the set of all polynomials in $\mathcal{P}_p(d,n)$ whose monomials are from the set $\mathcal{M}' \subseteq \mathcal{M}_p(d,n)$ (defined earlier) that have degree 1 on L and degree at most $d-1$ on R. By Corollary 2.3, there is some constant $\gamma > 0$ such that for sufficiently large n, $|\mathcal{M}'| \geqslant \gamma \cdot |\mathcal{M}_p(d,n)|$ and hence $|\mathcal{L}| \geqslant p^{\gamma|\mathcal{M}_p(d,n)|}$. Therefore, we have $|\mathcal{L}|/|\mathcal{P}_p(d,n)| \geqslant p^{-(1-\gamma)|\mathcal{M}_p(d,n)|}$.

Now consider the expected bias of polynomials in $\mathcal{L}$: We can write f chosen uniformly from $\mathcal{L}$ uniquely as

$$f(x) = \sum_{i \in L} x_i \cdot g_i(x_R)$$

where the g_i are independently chosen polynomials over monomials $\mathcal{M}_p(d-1,n')$ on R.

For $j \in \mathbb{F}_p^*$,

$$\begin{aligned}
\mathbf{E}_{f \in_R \mathcal{L}} \operatorname{bias}_j(f) =& \mathbf{E}_{f \in_R \mathcal{L}} \mathbf{E}_{x \in_R \mathbb{F}_p^n} \omega^{j \cdot f(x)} \\
=& \mathbf{E}_{f \in_R \mathcal{L}} \mathbf{E}_{x_L \in_R \mathbb{F}_p^L} \mathbf{E}_{x_R \in_R \mathbb{F}_p^R} \omega^{j \cdot f(x)} \\
=& \mathbf{E}_{x_L \in_R \mathbb{F}_p^L} \mathbf{E}_{x_R \in_R \mathbb{F}_p^R} \mathbf{E}_{f \in_R \mathcal{L}} \omega^{j \cdot f(x)}.
\end{aligned}$$

Now with probability $p^{-|L|}$, all the x_i for $i \in L$ are 0 and every $f \in \mathcal{L}$ evaluates to 0, so $\mathbf{E}_{f \in_R \mathcal{L}} \omega^{j \cdot f(0^L, x_R)} = 1$. With the remaining probability, $x_L \neq 0^L$ and hence there is some $i \in L$ and $b_i \neq 0$ such that $x_i = b_i$. For f chosen at random from $\mathcal{L}$, for each fixed value of $x_L = b_L$ with $b_i \neq 0$, we have

$$f(b_L, x_R) = b_i g_{i0} + f'(x_R)$$

where g_{i0} is the constant term of the polynomial g_i and is chosen independently of f'. Since g_{i0} is uniformly chosen from $\mathbb{F}_p$ for random f in $\mathcal{L}$ and since $b_i \neq 0$, $b_i g_{i0}$ is also uniformly chosen from $\mathbb{F}_p$. Further, since g_{i0} is independent of f', for every fixed x_R, the value $\mathbf{E}_{f \in_R \mathcal{L}} \omega^{j \cdot f(b_L, x_R)} = 0$. Therefore, $\mathbf{E}_{f \in_R \mathcal{L}} \operatorname{bias}_j(f) = p^{-|L|}$. Now since $|\operatorname{bias}_j(f)| \leqslant 1$, we obtain

$$\mathbf{Pr}_{f \in_R \mathcal{L}}[\ |\operatorname{bias}_j(f)| \geqslant p^{-|L|}/2\ ] \geqslant p^{-|L|}/2.$$

Therefore

$$\mathbf{Pr}_{f \in_R \mathcal{P}_p(d,n)}[|\operatorname{bias}_j(f)| \geqslant p^{-|L|}/2] \geqslant \frac{|\mathcal{L}|}{|\mathcal{P}_p(d,n)|} \cdot p^{-|L|}/2 \geqslant p^{-c'|\mathcal{M}_p(d,n)|}$$

for some $c' < 1$ since $|L| \ll |\mathcal{M}_p(d,n)|$. Since $|L| = \lfloor n/d \rfloor \geqslant 2$, we obtain $p^{-|L|}/2 > p^{-c''n/d}$ for some constant $c'' > 0$ as required. □

3 An extremal property of truncated Reed-Muller codes

In this section we prove Theorem 1.5. It is closely related to the problem of maximizing the number of common zeros of a set of linearly independent polynomials, which has been studied by Heijnen and Pelikaan [14]. Here, we make use of more general recent results of Beelen and Datta [5] who extend the methods of [14].

We start with some useful notations.

Definition 3.1. *Fix integers p, n. Let $F = \{0, \cdots, p-1\}^n$. For $d \leqslant n(p-1)$, define*

$$F_{\leqslant d} = \{a \in F : \sum_{i=1}^{n} a_i \leqslant d\}.$$

Define the ascending lexicographic order $\leqslant$ *over F as for $a, b \in F$, $b \leqslant a$ if and only if $\exists i \in [n]$ so that $b_j = a_j$ for all $j < i$ and $b_i < a_i$. Define the* descending lexicographic order $\geqslant$ *as $a \geqslant b$ if and only if $b \leqslant a$.*

Proposition 3.2 (Proposition 4.3 in [5])**.** *Recall that $\mathcal{P}_p(d, n)$ is the set of polynomials of degree at most d in n variables over $\mathbb{F}_p$. Fix integer r. Let $f_1, \cdots, f_r \in \mathcal{P}_p(d, n)$ be linear independent over $\mathbb{F}_p$. Let $Z(f_1, \cdots, f_r)$ denote the number of common zeros between $f_1, \cdots, f_r$. Then*

$$Z(f_1, \cdots, f_r) \leqslant \sum_{i=1}^{n} a_{r,i} p^{n-i},$$

where $a_r \in F_{\leqslant d}$ is the r-th element in $F_{\leqslant d}$ in descending lexicographic order.

Proposition 3.3 (Lemma 4.2, Lemma 4.3 and Proposition 4.5 in [5])**.** *For $b = (b_1, \cdots, b_n) \in F_{\leqslant d}$, define $f_b \in \mathcal{P}_p(d, n)$ as*

$$f_b(x) := \prod_{i=1}^{n} \prod_{j=1}^{b_i} (x_i - (j-1)).$$

Let $a_1, \cdots, a_r$ be the first r elements in $F_{\leqslant d}$ in descending lexicographic order. Then

$$Z(f_{a_1}, \cdots, f_{a_r}) = \sum_{i=1}^{n} a_{r,i} p^{n-i} := h_d(n, r).$$

Moreover, the common zeros of $f_{a_1}, \cdots, f_{a_r}$ are the first $h_d(n, r)$ elements in F^n in ascending lexicographic order.

Proposition 3.4. *$\{f_b\}_{b \in F_{\leqslant d}}$ forms a basis of $\mathcal{P}_p(d, n)$.*

Proof. The elements $b \in F_{\leqslant d}$ are the exponent vectors of the monomials $x^b := \prod_{i=1}^{n} x^{b_i} \in \mathcal{M}_p(d, n)$ which form a basis of $\mathcal{P}_p(d, n)$. The proposition follows by observing that the leading term of each f_b is the monomial x^b. □

Now we come to the quantities we focus on in this paper.

Definition 3.5. *For integer m, let $S_m \subset \mathbb{F}_p^n$ be the subset that contains the smallest m elements.*

Theorem 3.6 (Duality)**.** *Fix integers n, d, p. For any subset $S \subset \mathbb{F}_p^n$ with $|S| = m$, we have*

$$rank(M_S^{(d)}) \geqslant rank(M_{S_m}^{(d)}).$$

Proof. Let $V_S \subset \mathbb{F}_p^{|\mathcal{M}_p(d,n)|}$ be the linear subspace spanned by the rows in $M^{(d)}$ indexed by S. Then $rank(M_S^{(d)}) = k$ if and only if $\dim(V_S) = k$, which is equivalent to $\dim(V_S^\top) = |\mathcal{M}_p(d,n)| - k$.

On the other hand, we can interpret elements in $\mathbb{F}_p^{|\mathcal{M}_p(d,n)|}$ as polynomials by considering each entry as the coefficient in the corresponding monomial. Formally, this is the bijection $\phi : \mathbb{F}_p^{|\mathcal{M}_p(d,n)|} \to \mathcal{P}_p(d,n)$ defined as

$$\phi((v_\alpha)_{\alpha \in \mathcal{M}_p(d,n)}) = \sum_{\alpha \in \mathcal{M}_p(d,n)} v_\alpha \cdot \alpha.$$

From this view, $f \in \mathcal{P}_p(d,n)$ is contained in $\dim(V_S^\top)$ if and only if $\langle \phi^{-1}(f), M_{\{x\}}^{(d)} \rangle = 0$ for all $x \in S$. But $\langle \phi^{-1}(f), M_{\{x\}}^{(d)} \rangle$ is just $f(x)$, which means that S is the set of common zeros for all f so that $\phi(f) \in \dim(V_S^\top)$.

Let r be the largest integer so that $h_d(n,r) \geqslant m$, where $h_d(n,r)$ is defined in Proposition 3.3.

We first claim that $rank(M_S^{(d)}) \geqslant |\mathcal{M}_p(d,n)| - r$. If this is not the case, then $\dim(V_S^\top) = |\mathcal{M}_p(d,n)| - rank(M_S^{(d)}) > r$; namely, we can find $r+1$ linearly independent polynomials $f_1, \cdots, f_{r+1}$ so that they evaluate to 0 on S. Therefore, $r+1$ linearly independent polynomials can have m common zeros. By Proposition 3.2 and Proposition 3.3, this means $h_d(n, r+1) \geqslant m$, which violates our definition of r.

We then argue that $rank(M_{S_m}^{(d)}) \leqslant |\mathcal{M}_p(d,n)| - r$. This follows because we can choose $a_1, \cdots, a_r$ as in Proposition 3.3, and construct $V_r \subset \mathbb{F}_p^{|\mathcal{M}_p(d,n)|}$ as the linear span of $\{\phi^{-1}(f_{a_i})\}_{i=1}^r$. By Proposition 3.4, $\{f_{a_i}\}_{i=1}^r$ are linearly independent. Since ϕ is linear, $\{\phi^{-1}(f_{a_i})\}_{i=1}^r$ are also linearly independent, so $\dim(V_r) = r$. By Proposition 3.3, S_m are the common zeros of $f_{a_1}, \cdots, f_{a_r}$, hence V_{S_m} is contained in $V_r^\top$. Therefore

$$rank(M_{S_m}^{(d)}) \leqslant \dim(V_{S_m}) \leqslant \dim(V_r^\top) = |\mathcal{M}_p(d,n)| - r,$$

which completes the proof. □

Theorem 3.6 is sufficient to prove Theorem 1.5. However, Proposition 3.2 and Proposition 3.3 does not give the explicit expression of $rank(M_{S_m}^{(d)})$. Here we give an explicit recursive expression.

Definition 3.7. *Define $g_d(m)$ as the rank of $M_{S_m}^{(d)}$. For the completeness of definition, we set $g_d(m) = 0$ when $d < 0$ or $m = 0$.*

The g_d function is easy to compute when the input m is a power of p.

Lemma 3.8. *For integer $r \geqslant 0$, $g_d(p^r) = |\mathcal{M}_p(d,r)|$.*

Proof. The rows in S_{p^r} correspond to assignments that fix the first $n-r$ variables to 0 and have all possible values for the remaining r variables. Therefore, $M_{S_{p^k}}^{(d)}$ is 0 except on the columns indexed by monomials on these last r variables and has full column rank on the remaining $|\mathcal{M}_p(d,r)|$ columns since no non-zero polynomial using these monomials can be identically 0 over $\mathbb{F}_p^r$. □

More generally, we have the following recursive formulation of the g_d function.

Theorem 3.9. *Let r be the unique integer so that $p^r \leqslant m < p^{r+1}$. Let $m = k \cdot p^r + c$. Then*

$$g_d(m) = \sum_{i=0}^{k-1} g_{d-i}(p^r) + g_{d-k}(c).$$

As a special case, when $c = 0$, we have $g_d(k \cdot p^r) = \sum_{i=0}^{k-1} g_{d-i}(p^r)$.

The g_d function can be analyzed using the LU decomposition of the $(n+1)$-dimension Vandermonde matrix $V^{(n)}$ on variables $x_0, x_1, \ldots x_n$ (i.e., $V^{(n)}_{ij} = x_i^j$) given by Oruç and Phillips [21].

Proposition 3.10 (Theorem 2.1 in [21])**.** *$V^{(n)}$ has LU-decomposition $V^{(n)} = L^{(n)}U^{(n)}$ where lower triangular matrix $L^{(n)} \in \mathbb{R}^{(n+1)\times(n+1)}$ has all diagonal entries 1, and upper triangular matrix $U^{(n)} \in \mathbb{R}^{(n+1)\times(n+1)}$ has $U^{(n)}_{i,i} = \prod_{j<i}(x_i - x_j)$ for all $i \in \{0, 1 \ldots, n\}$.*

Proof of Theorem 3.9. For the sake of convenience, let $\mathcal{M}_p^{=d}(r)$ be the set of monomials over the last r variables whose degree equals d, and $\mathcal{M}_p^{\leqslant d}(r)$ be the set of monomials over the last r variables whose degree is at most d.

Consider the block structure of the matrix. For $i = 0, 1, \cdots, p-1$, let A_i be the submatrix with S_{p^r} as rows and $\mathcal{M}_p^{=d-p+i+1}(r)$ as columns. Let $A_{\leqslant i}$ be the submatrix $(A_0, \cdots, A_i)$. Then its columns are given by $\mathcal{M}_p^{\leqslant d-p+i+1}(r)$.

Now, let us consider the rows for $R_t := \{a \in \mathbb{F}_p^n \mid a_1 = \cdots = a_{n-r-1} = 0,\ a_{n-r} = t\}$. The non-zero parts correspond to monomials that only depend on $x_{n-r}, x_{n-r+1}, \cdots, x_n$. If we group all the monomials by their degree on x_{n-r} then, for $t \leqslant k-1$, the row will be of the form

$$A_{\leqslant p-1},\ t \cdot A_{\leqslant p-2},\ t^2 \cdot A_{\leqslant p-3},\ \cdots,\ t^{p-1} \cdot A_{\leqslant 0}.$$

Things are a little different for $t = k$, since $|R_k| < p^r$. In this case, we define $A'_{\leqslant i}$ to be the first c rows of $A_{\leqslant i}$ and it is easy to check that the row is of the form

$$A'_{\leqslant p-1},\ k \cdot A'_{\leqslant p-2},\ k^2 \cdot A'_{\leqslant p-3},\ \cdots,\ k^{p-1} \cdot A'_{\leqslant 0}.$$

Therefore the matrix $M^{(d)}_{S_m}$ is of the form:

	$\mathcal{M}_p^{\leqslant d}(r)$	$x_{n-r} \cdot \mathcal{M}_p^{\leqslant d-1}(r)$	$x_{n-r}^2 \cdot \mathcal{M}_p^{\leqslant d-2}(r)$	$\cdots$	$x_{n-r}^{p-1} \cdot \mathcal{M}_p^{\leqslant d-p+1}(r)$
$\{a_{n-r} = 0\}$	$A_{\leqslant p-1}$	0	0	$\cdots$	0
$\{a_{n-r} = 1\}$	$A_{\leqslant p-1}$	$A_{\leqslant p-2}$	$A_{\leqslant p-3}$	$\cdots$	$A_{\leqslant 0}$
$\{a_{n-r} = 2\}$	$A_{\leqslant p-1}$	$2 \cdot A_{\leqslant p-2}$	$4 \cdot A_{\leqslant p-3}$	$\cdots$	$2^{p-1} \cdot A_{\leqslant 0}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\{a_{n-r} = k-1\}$	$A_{\leqslant p-1}$	$(k-1) \cdot A_{\leqslant p-2}$	$(k-1)^2 \cdot A_{\leqslant p-3}$	$\cdots$	$(k-1)^{p-1} \cdot A_{\leqslant 0}$
$\{a_{n-r} = k\}$	$A'_{\leqslant p-1}$	$k \cdot A'_{\leqslant p-2}$	$k^2 \cdot A'_{\leqslant p-3}$	$\cdots$	$k^{p-1} \cdot A'_{\leqslant 0}$

We observe the following:

- $A_{\leqslant p-i-1}$ is the first $|\mathcal{M}_p^{\leqslant d-i-1}(r)|$ columns of $A_{\leqslant p-i}$, and
- $A'_{\leqslant i}$ is the first c rows of $A_{\leqslant i}$.

$M^{(d)}_{S_m}$ is closely related to the the Vandermonde matrix $V^{(p-1)}[0, 1, \cdots, p-1]$. To see this, notice that $M^{(d)}_{S_m}$ is a submatrix of $V = V^{(p-1)}[0, 1, \cdots, p-1] \otimes A_{\leqslant p-1} \in \mathbb{R}^{p \cdot p^r \times p \cdot |\mathcal{M}_p^{\leqslant d}(r)|}$. By Proposition 3.10, $V^{(p-1)}[0, 1, \cdots, p-1] = LU$ where lower triangular matrix $L \in \mathbb{R}^{p\times p}$ and upper triangular matrix $U \in \mathbb{R}^{p\times p}$ satisfy $L_{i,i} = 1$ and $U_{i,i} \neq 0$ for $i = 0, \cdots, p-1$. Therefore, if

we set $L' = L \otimes I_{p^r} \in \mathbb{R}^{p\cdot p^r \times p\cdot p^r}$, $C = I_p \otimes A_{\leqslant p-1} \in \mathbb{R}^{p\cdot p^r \times p\cdot|\mathcal{M}_p^{\leqslant d}(r)|}$ and $U' = U \otimes I_{|\mathcal{M}_p^{\leqslant d}(r)|} \in \mathbb{R}^{p\cdot|\mathcal{M}_p^{\leqslant d}(r)|\times p\cdot|\mathcal{M}_p^{\leqslant d}(r)|}$, we have $V = L'CU'$. Notice that L' is lower-triangular, because L is lower-triangular and I_{p^r} is diagonal. Similarly U' is upper-triangular.

Let $S_L \subset [p \cdot p^r]$ be the set of indices of the rows of $M_{S_m}^{(d)}$ and $S_U \subset [p \cdot |\mathcal{M}_p^{\leqslant d}(r)|]$ be the set of indices of the columns of $M_{S_m}^{(d)}$. Then S_L is simply S_m; S_U is more complicated. We can group the columns of V into p groups, each of size $|\mathcal{M}_p^{\leqslant d}(r)|$. By the first observation, for $i = 0, \cdots, p-1$, we simply take the first $|\mathcal{M}_p^{\leqslant d-i}(r)|$ columns in each block.

Let P be the m by $p \cdot p^r$ matrix defined as $P_{ij} = 1$ if $i = j$ and $P_{ij} = 0$ otherwise. Let Q be the $p \cdot |\mathcal{M}_p^{\leqslant d}(r)|$ by $\sum_{i=0}^{p-1} |\mathcal{M}_p^{\leqslant d-i}(r)|$ matrix defined as $Q_{ij} = 1$ if $i = j \in S_U$, and $Q_{ij} = 0$ otherwise. Then clearly we have $M_{S_m}^{(d)} = PL'CU'Q$.

Write L' as $\begin{pmatrix} L_1 & 0 \\ L_2 & L_3 \end{pmatrix}$ where $L_1 \in \mathbb{R}^{m\times m}$. Then we have $PL' = \begin{pmatrix} L_1 & 0 \end{pmatrix} = L_1 P$. Similarly, let U_1 be the principal submatrix of U indexed by S_U. Then we have $UQ = QU_1$. Since L', U' are triangular matrices with non-zero diagonal entries, L_1, U_1 are also triangular matrices with non-zero diagonal entries, and hence invertible.

Now we have $M_{S_m}^{(d)} = PL'CU'Q = L_1 PCQU_1$. Let $C_1 = PCQ$. Then C_1 is of the form

	$\mathcal{M}_p^{\leqslant d}(r)$	$x_{n-r}\cdot\mathcal{M}_p^{\leqslant d-1}(r)$	$x_{n-r}^2\cdot\mathcal{M}_p^{\leqslant d-2}(r)$	$\cdots$	$x_{n-r}^{k-1}\cdot\mathcal{M}_p^{\leqslant d-k+1}(r)$	$x_{n-r}^{k}\cdot\mathcal{M}_p^{\leqslant d-k}(r)$
$\{a_{n-r}=0\}$	$A_{\leqslant p-1}$	0	0	$\cdots$	0	0
$\{a_{n-r}=1\}$	0	$A_{\leqslant p-2}$	0	$\cdots$	0	0
$\{a_{n-r}=2\}$	0	0	$A_{\leqslant p-3}$	$\cdots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\{a_{n-r}=k-1\}$	0	0	0	$\cdots$	$A_{\leqslant p-k}$	0
$\{a_{n-r}=k\}$	0	0	0	$\cdots$	0	$A'_{\leqslant p-k-1}$

Since L_1, U_1 are invertible, we have $rank(M_{S_m}^{(d)}) = rank(C_1)$. By the definition of the g_d function, the rank of $A_{\leqslant p-i}$ is $g_{d-i}(p^r)$ and that of $A'_{\leqslant p-k-1}$ is $g_{d-k}(c)$. Hence

$$g_d(m) = g_d(k\cdot p^r + c) = \sum_{i=0}^{k-1} g_{d-i}(p^r) + g_{d-k}(c).$$

□

Acknowledgements

We thank the anonymous referees for suggestions that have helped to simplify and significantly improve the presentation of our results. We also thank Srikanth Srinivasan for pointing us to the reference [5].

References

[1] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed-Muller codes for random erasures and errors. *IEEE Transactions on Information Theory*, 61(10):5229–5252, 2015.

[2] Paul Beame, Shayan Oveis Gharan, and Xin Yang. On the bias of Reed-Muller codes over odd prime fields. *arXiv preprint arXiv:1806.06973v1*, 2018.

[3] Paul Beame, Shayan Oveis Gharan, and Xin Yang. Time-space tradeoffs for learning finite functions from random evaluations, with applications to polynomials. Technical Report TR18-114, Electronic Colloquium on Computational Complexity (ECCC), 2018.

[4] Paul Beame, Shayan Oveis Gharan, and Xin Yang. Time-space tradeoffs for learning finite functions from random evaluations, with applications to polynomials. In *Proceedings of the 31st Conference on Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, 2018.

[5] Peter Beelen and Mrinmoy Datta. Generalized Hamming weights of affine Cartesian codes. *Finite Fields and Their Applications*, 51:130–145, 2018.

[6] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Computational Complexity*, 21(1):63–81, 2012.

[7] Abhishek Bhowmick and Shachar Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *CoRR*, abs/1506.02047, 2015.

[8] Abhishek Bhowmick and Shachar Lovett. The list decoding radius for Reed-Muller codes over small fields. *IEEE Trans. Information Theory*, 64(6):4382–4391, 2018.

[9] Leonard E. Dickson. *Linear Groups with an Exposition of the Galois Field Theory*. B.G. Trubner, Leipzig, 1901.

[10] Sumegha Garg, Ran Raz, and Avishay Tal. Extractor-based time-space lower bounds for learning. In *Proceedings of the Fiftieth Annual ACM on Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA*, 2018. To appear.

[11] Parikshit Gopalan, Adam R Klivans, and David Zuckerman. List-decoding Reed-Muller codes over small fields. In *Proceedings of the Fortieth Annual ACM symposium on Theory of Computing*, pages 265–274. ACM, 2008.

[12] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *arXiv preprint arXiv:0711.3191*, 2007.

[13] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the Forty-Second Annual ACM symposium on Theory of computing*, pages 331–340. ACM, 2010.

[14] Petra Heijnen and Ruud Pellikaan. Generalized Hamming weights of q-ary Reed-Muller codes. *IEEE Transactions on Information Theory*, 44(1):181–196, 1998.

[15] Tadao Kasami and Nobuki Tokura. On the weight structure of Reed-Muller codes. *IEEE Transactions on Information Theory*, 16(6):752–759, 1970.

[16] Tadao Kasami, Nobuki Tokura, and Saburo Azumi. On the weight enumeration of weights less than 2.5d of Reed-Muller codes. *Information and Control*, 30(4):380–395, 1976.

[17] Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed-Muller codes. *IEEE Trans. Information Theory*, 58(5):2689–2696, 2012.

[18] Peter Keevash and Benny Sudakov. Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. *SIAM Journal on Discrete Mathematics*, 18(4):713–727, 2005.

[19] Robert James McEliece. *Linear recurring sequences over finite fields*. PhD thesis, California Institute of Technology, 1967.

[20] David E Muller. Application of Boolean algebra to switching circuit design and to error detection. *Transactions of the IRE Professional Group on Electronic Computers*, EC-3(3):6–12, 1954.

[21] Halil Oruç and George M Phillips. Explicit factorization of the Vandermonde matrix. *Linear Algebra and its Applications*, 315(1-3):113–123, 2000.

[22] Irving S Reed. A class of multiple-error-correcting codes and the decoding scheme. Technical report, MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1953.

[23] Neil J. A. Sloane and Elwyn R. Berlekamp. Weight enumerator for second-order Reed-Muller codes. *IEEE Trans. Information Theory*, 16(6):745–751, 1970.