Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 119 (2019)

On Hitting-Set Generators for Polynomials that Vanish Rarely

Dean Doron[*], Amnon Ta-Shma[†], and Roei Tell[‡]

January 6, 2020

Abstract

The problem of constructing hitting-set generators for polynomials of low degree is fundamental in complexity theory and has numerous well-known applications. We study the following question, which is a relaxation of this problem: Is it easier to construct a hitting-set generator for polynomials $p : \mathbb{F}^n \to \mathbb{F}$ of degree d if we are guaranteed that the polynomial vanishes on at most an $\varepsilon > 0$ fraction of its inputs? We will specifically be interested in tiny values of $\varepsilon \ll d/|\mathbb{F}|$. This question was first considered by Goldreich and Wigderson (STOC 2014), who studied a specific setting geared for a particular application, and another specific setting was later studied by the third author (CCC 2017).

In this work our main interest is a *systematic study of the relaxed problem*, in its general form, and we prove results that significantly improve and extend the two previously-known results. Our contributions are of two types:

- Over fields of size $2 \leq |\mathbb{F}| \leq \mathrm{poly}(n)$, we show that the seed length of any hitting-set generator for polynomials of degree $d \leq n^{.49}$ that vanish on at most $\varepsilon = |\mathbb{F}|^{-t}$ of their inputs is at least $\Omega\left((d/t) \cdot \log(n)\right)$.
- Over $\mathbb{F}_2$, we show that there exists a (non-explicit) hitting-set generator for polynomials of degree $d \leq n^{.99}$ that vanish on at most $\varepsilon = |\mathbb{F}|^{-t}$ of their inputs with seed length $O\left((d-t) \cdot \log(n)\right)$. We also show a polynomial-time computable hitting-set generator with seed length $O\left((d-t) \cdot \left(2^{d-t} + \log(n)\right)\right)$.

In addition, we prove that the problem we study is closely related to the following question: "Does there exist a small set $S \subseteq \mathbb{F}^n$ whose degree-d closure is very large?", where the degree-d closure of S is the variety induced by the set of degree-d polynomials that vanish on S.

[*] Department of Computer Science, Stanford University. Email: ddoron@stanford.edu.

[†] The Blavatnik School of Computer Science, Tel-Aviv University. Email: amnon@tau.ac.il.

[‡] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science. Email: roei.tell@weizmann.ac.il.

ISSN 1433-8092

Contents

1 Introduction

Let $\mathcal{P}_{n,q,d}$ denote the set of all polynomials $\mathbb{F}^n \to \mathbb{F}$ of total degree d over the field of size $q = |\mathbb{F}|$. We think of n as sufficiently large, and of the degree d and the field size q as functions of n. For simplicity, throughout the paper we assume that $d < n$.[1]

A fundamental problem in complexity theory is that of constructing *hitting-set generators for low-degree polynomials*. Recall that a Hitting-Set Generator (HSG) for $\mathcal{P}_{n,q,d}$ is a function $H\colon \{0,1\}^\ell \to \mathbb{F}^n$ such that for every non-zero polynomial $p \in \mathcal{P}_{n,q,d}$ there exists $s \in \{0,1\}^\ell$ satisfying $p(H(s)) \neq 0$ (see Definition 11); in other words, every non-zero polynomial $p \in \mathcal{P}_{n,q,d}$ does not vanish on at least one element in the *hitting-set* $S = \left\{H(s)\colon s \in \{0,1\}^\ell\right\}$. The two main measures of efficiency for HSGs are the seed length ℓ (equivalently, the size of the hitting-set S as a multiset) and the computational complexity of H as a function (i.e., the computational complexity of generating an element of the hitting-set S given its index s).

A standard linear-algebraic argument yields a lower bound of $\Omega\left(d \cdot \log\left(n/d\right)\right)$ on the seed length of any HSG for $\mathcal{P}_{n,q,d}$, and a standard probabilistic argument shows that there *exists* a HSG for $\mathcal{P}_{n,q,d}$ with matching seed length $O\left(d \cdot \log(n/d) + \log\log(q)\right)$ (see Facts 14 and 15). Naturally, the probabilistic upper-bound does not guarantee that the function H is *efficiently-computable*. Thus, the main open problem concerning HSGs for $\mathcal{P}_{n,q,d}$ is to construct efficiently-computable HSGs with seed length that matches the known lower bound. This well-known problem (as well as a variant that refers to *pseudorandom generators* as in Definition 13) has attracted a significant amount of attention over the years; see, e.g., [NN93; LVW93; LV98; KS01; Bog05; BV10; BHS08; Lov09; Vio09b; Lu12; CTS13; ST18], and the related survey by Viola [Vio09a].

Several years ago, Goldreich and Wigderson [GW14, Section 5] considered a *relaxed version* of the foregoing problem. In general terms, what they asked is the following:

> Does the HSG problem become easier if we are guaranteed that the polynomial *vanishes rarely* (i.e., has very few roots)?

Note that, intuitively, we expect that the relaxed problem will indeed be easier: This is both since there are less polynomials that vanish rarely (than arbitrary polynomials), and since for any such polynomial p, almost all inputs will "hit" p.

In their original paper, Goldreich and Wigderson considered a specific instance of this problem, geared for a particular application (see Section 1.2 for details). In this paper our goal is to *study the relaxed problem in and of itself, in a systematic and general way*. Our motivation for doing so is three-fold. First, this is a special (and potentially-easy) case of the classical HSG problem, and thus constitutes a potential path to make progress on the classical problem. Secondly, the relaxed question is of independent interest as part of the broad study of *quantified derandomization*, which was initiated in the original work of Goldreich and Wigderson [GW14] (see also, e.g., [Tel19; CT19;

[1] Most of our results also carry on to the setting of $d > n$, albeit with less "clean" parametrizations.

DMO+19]). And thirdly, as polynomial-based constructions are ubiquitous in complexity theory, any progress in our understanding of structured classes of polynomials or in related HSG constructions may be valuable for other explicit constructions.

To be more formal, denote by $\mathcal{P}_{n,q,d,\varepsilon}$ the set of polynomials $p \in \mathcal{P}_{n,q,d}$ such that $\Pr_{x\in\mathbb{F}^n}[p(x) = 0] \leq \varepsilon$; that is, $\mathcal{P}_{n,q,d,\varepsilon}$ is the set of degree-d polynomials that *vanish rarely*, where the notion of "rarely" is parametrized by the parameter ε. The two main questions we consider in this context are:

- **The combinatorial question:** What is the minimal size of a hitting-set for $\mathcal{P}_{n,q,d,\varepsilon}$? Equivalently, we ask what is the minimal seed length of any HSG for $\mathcal{P}_{n,q,d,\varepsilon}$. This question is combinatorial since it refers to the *existence* of a HSG, regardless of its computational complexity.

- **The computational question:** For which values of $\varepsilon > 0$ can we construct a HSG for $\mathcal{P}_{n,q,d,\varepsilon}$ with small seed length that will be *efficiently-computable*? In other words, can we simultaneously optimize not only the seed length but also the *computational complexity* of HSGs for $\mathcal{P}_{n,q,d,\varepsilon}$?

1.1 Context and previous work

Let us first delineate some trivial values for ε. To do so, first recall that we expect a random polynomial to vanish on q^{-1} of its inputs. Now, by the Schwartz-Zippel lemma, any non-zero $p \in \mathcal{P}_{n,q,d}$ has at most an $\varepsilon = d/q$ fraction of roots; this bound is quite good when q is large compared to d, and in general, for abitrary d and q, any non-zero polynomial vanishes on at most $1-\delta$ of its inputs, where $\delta \geq q^{-d/(q-1)}$ denotes the relative distance of the Reed-Muller code of degree d over $\mathbb{F}_q$. Therefore, the value $\varepsilon = 1-\delta$ represents the general case (i.e., the case of hitting *any* non-zero polynomial). Remarkably, we also have a minimal non-zero value that ε can have: By a theorem of Warning [War35], every polynomial in $\mathbb{F}_q^n \to \mathbb{F}_q$ of degree d that vanishes *somewhere* vanishes on at least a q^{-d} fraction of its inputs. Therefore, hitting polynomials that vanish on $\varepsilon < q^{-d}$ fraction of their inputs is trivial, since such polynomials have no zeroes. It will be useful to denote $\varepsilon = q^{-t}$ from now on.

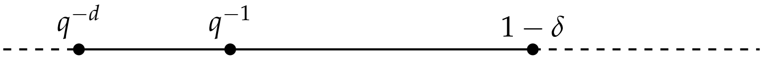

Figure 1: The two extremal values of ε (i.e., $\varepsilon = q^{-d}$ and $\varepsilon = 1-\delta$) and the expected $\varepsilon = q^{-1}$ for a random polynomial. (The parameter δ denotes the relative distance of the corresponding q-ary Reed-Muller code $RM(n,d)$.)

Referring to the combinatorial question, the standard probabilistic argument mentioned before shows there exists a HSG for $\mathcal{P}_{n,q,d,\varepsilon}$ with seed length $O(\log\log(|\mathcal{P}_{n,q,d,\varepsilon}|))$. Thus, the combinatorial question is intimately connected to the long-standing open problem of determining the *weight distribution of the Reed-Muller code*, i.e., *counting* the number of polynomials in $\mathcal{P}_{n,q,d}$ that vanish on precisely $\varepsilon > 0$ of their inputs, for every

$\varepsilon > 0$. The latter problem has been studied since the late 60's (see, e.g., [BS69; KT70]), but is currently settled only for $d = 2$ (see [SB70; McE69]). Only recently have general results been obtained for $d > 2$, and the bounds in these results are asymptotic (rather than precise bounds) and hold only over $\mathbb{F}_2$ (see [KLP12; ASW15]). More generally, this problem is a special case of the well-known problem of studying weight distributions of (classes of) linear codes, which is typically tackled using weight enumerator polynomials (for relevant background see, e.g., [MS77, Chapter 5]). Note, however, that the weight distribution problem is more general, since it refers to all non-trivial values of $\varepsilon > 0$, whereas in our setting we focus only on tiny values of ε.

Another related line of works focuses on structural properties of *biased polynomials*. Fixing a polynomial $p : \mathbb{F}^n \to \mathbb{F}$ and looking at the distribution over $\mathbb{F}$ that is obtained by evaluating p at a random point, we can ask whether this distribution is close to uniform, or whether it is far from uniform, in which case we call the polynomial biased. A sequence of works showed that biased polynomials are very "structured", in the sense that they can be determined by a relatively-small number of polynomials of lower degree (see [GT09; KL08; HS10; Bha14; BHT15; BBG16]). Our setting is much more specific than the setting in these works, since their assumption is only that the polynomial is *biased*, whereas our assumption is that the polynomial is biased in a very specific manner (i.e., one output-value has tiny weight $\varepsilon > 0$). Thus, the results in these works typically do not seem sufficiently strong to be useful in our more specific setting.[2]

Goldreich and Wigderson [GW14, Section 5], who were motivated by a specific application in circuit complexity (derandomization of $\mathcal{AC}^0[\oplus]$), constructed a polynomial-time computable HSG for the setting of $q = 2$ and $\varepsilon = 2^{-(d-O(1))} = O(2^{-d})$ (for details see Section 1.2). Thus, they gave an upper-bound for the *computational question*, which holds only for $\mathbb{F}_2$ polynomials with extremely few roots. In a subsequent work by the third author [Tel19], two combinatorial lower bounds were proved for the setting of $q = \text{poly}(n)$ and $\varepsilon = q^{-O(1)}$ (again, for details see Section 1.2). Thus, the subsequent work showed lower bounds for the *combinatorial question*, which hold only for polynomials over $\mathbb{F}_{\text{poly}(n)}$ with a relatively-large number of roots (i.e., only mildly less roots than the expected value of $\varepsilon = q^{-1}$). In both previous works, ad-hoc arguments were used to obtain the corresponding results.

1.2 Our main results

Our first main result is a general lower bound for the combinatorial problem. For context, in [Tel19] it was shown that when $q = \text{poly}(n)$, any HSG for $\mathcal{P}_{n,d,q,q^{-O(1)}}$ requires a seed of length $\Omega(d^{\Omega(1)} \cdot \log(n/d^{\Omega(1)}))$; and any HSG with constant density[3] for $\mathcal{P}_{n,d,q,q^{-1}}$ requires a seed of length $\Omega(d \cdot \log(n/d))$. Thus, both previous lower bounds referred to the setting of $q = \text{poly}(n)$ and of $\varepsilon = q^{-O(1)}$ (i.e., $t = O(1)$).

The following result shows a lower bound that is both significantly stronger, and –

[2] One exception is the field $\mathbb{F}_2$, in which the notions of bias and of "vanish rarely" converge. Indeed, the proofs of our results for $\mathbb{F}_2$ use insights developed in this sequence of works.

[3] A hitting-set S for a class $\mathcal{P}$ has density $\varepsilon > 0$ if for every $p \in \mathcal{P}$ it holds that $\Pr_{s \in S}[p(s) \neq 0] \geq \varepsilon$.

more importantly – applies to a far broader parameter setting. In particular, the following result applies to a general $q \leq \text{poly}(n)$ and to values of $\varepsilon = q^{-t}$ almost up to the extreme value of $\varepsilon = q^{-d}$, and gives a lower bound of $\Omega((d/t) \cdot \log(n))$:

Theorem 1 (lower bound over general fields) *For every constant $c > 1$ there exists a constant $\gamma > 0$ such that the following holds. For every $n, q, d, t \in \mathbb{N}$ such that $2 \leq q \leq n^c$ is a prime power, $d \leq n^{.49}$, and $t \leq \gamma \cdot d$, any HSG for $\mathcal{P}_{n,q,d,q^{-t}}$ requires a seed of length $\Omega\left((d/t) \cdot \log(n)\right)$.*

Let us parse the meaning of the lower bound in Theorem 1. For comparison, recall that there exists a HSG for all polynomials of degree $d \leq n^{.49}$ with seed length $O(d \cdot \log(n))$. Theorem 1 tells us that the relaxation of only requiring to "hit" polynomials that vanish with probability q^{-t} can "buy" a factor of at most $1/t$ in the seed length. In particular, there does not exist a significantly smaller hitting-set for polynomials that vanish with probability $q^{-O(1)}$. Perhaps surprisingly, this is also true for polynomials that vanish with probability $q^{-d^{o(1)}}$ (since the lower bound remains almost linear in $d \cdot \log(n)$). Only for polynomials that vanish with probability $q^{-d^{\Omega(1)}}$ does our lower bound imply that a significantly smaller hitting-set *might* exist; and at an "extreme" value of $q^{-\Omega(d)}$, our lower bound does not rule out a polynomial-sized hitting-set.

For technical statements that include various extensions and improvements of Theorem 1 (and in particular also hold for polynomials of higher degree $n^{.49} < d \leq \gamma \cdot n$), see the beginning of Section 6, and specifically Theorems 28, 33, and 34.[4]

Now, still referring to the combinatorial question, we observe that a result of Kaufmann, Lovett, and Porat [KLP12], which upper-bounds the *number* of biased $\mathbb{F}_2$ polynomials (i.e., analyzes the weight distribution of the Reed-Muller code over $\mathbb{F}_2$), yields a corresponding existential upper-bound. Specifically:

Theorem 2 (upper-bound over $\mathbb{F}_2$, following [KLP12]) *Let $n, d, t \in \mathbb{N}$ where $d > t$. Then, there exists a (non-explicit) hitting-set for $\mathcal{P}_{n,2,d,2^{-t}}$ with seed length $O\left((d-t) \cdot \log(\frac{n}{d-t})\right)$.*

Note that while the lower bound in Theorem 1 holds for any finite field, the upper bound in Theorem 2 holds only over $\mathbb{F}_2$. Nevertheless, comparing Theorems 1 and 2 (for $\mathbb{F} = \mathbb{F}_2$ and $d \leq n^{.49}$) reveals that there is still a *significant gap* between the upper-bound and the lower-bound: The lower bound is of the form $(d/t) \cdot \log(n)$, whereas the existential upper bound is of the form $(d-t) \cdot \log(n)$. For example, the lower bound indicates that there *might* exist a significantly smaller hitting-set for the relaxed problem when $t = d^{\Omega(1)}$, whereas the existential upper bound is significantly better than the one for the original problem only for $t = d - d^{\Omega(1)}$.

Our last main result is computational and shows an *explicit* construction of a HSG. As mentioned above, Goldreich and Wigderson [GW14] constructed a polynomial-time computable HSG with seed length $O(\log(n))$ that "hits" polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ of degree d that vanish on $O(2^{-d})$ of their inputs (for any $d \in \mathbb{N}$). We prove a significantly more general result, by constructing an explicit HSG for $\mathcal{P}_{n,2,d,2^{-t}}$ for any $t < d - O(1)$:

[4] In these technical results, the $\log(n)$ term in the lower bound in Theorem 1 is replaced by a more complicated term that depends on d and on t, for example $\log(n^{.99} \cdot (t/d))$.

Theorem 3 (explicit upper-bound over $\mathbb{F}_2$) *Let $n \in \mathbb{N}$ be sufficiently large, and let $d > t+4$ be integers. Then, there exists a polynomial-time computable HSG for $\mathcal{P}_{n,2,d,2^{-t}}$ with seed length $O\left((d-t)\cdot\left(2^{d-t}+\log(\frac{n}{d-t})\right)\right)$.*

Note that the original result from [GW14] is the special case of Theorem 3 when $t = d - O(1)$. Also note that the seed length of the explicit HSG from Theorem 3 depends exponentially on $d-t$, whereas the seed length of the non-explicit HSG from Theorem 2 depends linearly on $d-t$. We also comment that the result is actually slightly stronger, and asserts that for any $r \in \mathbb{N}$ there exists a polynomial-time computable HSG for $\bigcup_d \mathcal{P}_{n,2,d,q^{d-r}}$ with seed length $O(r\cdot(2^r+\log(n/r)))$; that is, for every r there is a *single* HSG that works for *all degrees* d with $t = d-r$.

Below, in Table 1, we present an informal summary of the main results mentioned above, and compare them to previously-known results.

	Seed length	**Field Size**	ε
Lower bounds			
[Tel19]	$\Omega(d^{\Omega(1)}\cdot\log(n/d^{\Omega(1)}))$	$q = \mathrm{poly}(n)$	$q^{-O(1)}$
Thm 1	$\Omega((d/t)\cdot\log n)$ $(d \le n^{.49})$	$2 \le q \le \mathrm{poly}(n)$	q^{-t}
Thm 28	$\Omega((d/t)\cdot\log(n^{.99}\cdot t/d))$ $(d/t \lessapprox q\cdot n^{.01})$	$2 \le q \le \mathrm{poly}(n)$	q^{-t}
Upper bounds			
[GW14]	$O(\log n)$ (explicit)	$q = 2$	$2^{-d+O(1)}$
Thm 2	$O((d-t)\log(\frac{n}{d-t}))$ (non-explicit)	$q = 2$	2^{-t}
Thm 3	$O((d-t)\cdot(2^{d-t}+\log(\frac{n}{d-t})))$ (explicit)	$q = 2$	2^{-t}

Table 1: An informal summary of our results and comparison to previous results.

1.3 The connection to small sets with large degree-d closures

In addition to our lower-bounds and upper-bounds for the problem of HSGs for polynomials that vanish rarely, we also tie this problem to the study of a clean and elegant algebraic question; namely, to the study of the degree-d closure of a set $S \subseteq \mathbb{F}^n$, which was recently initiated by Nie and Wang [NW15].

Using terminology from algebraic geometry, the degree-d closure of a set $S \subseteq \mathbb{F}^n$ is a finite-degree analogue of the Zariski closure of S, and is defined as the variety induced by the set of degree-d polynomials $\mathbb{F}^n \to \mathbb{F}$ that vanish on S. In more detail, let us first define the degree-d ideal of S to be $\mathcal{I}^{(d)}(S) = \{p \in \mathcal{P}_d : \forall s \in S, p(s) = 0\}$, where $\mathcal{P}_d$ is the set of degree-d polynomials $\mathbb{F}^n \to \mathbb{F}$.[5] Then, the degree-$d$ closure of S is defined by:

$$\mathtt{Cl}^{(\mathtt{d})}(S) = \{x \in \mathbb{F}^n : \forall p \in \mathcal{I}^{(d)}(S), p(x) = 0\} .$$

As an example, observe that the degree-d closure of any $d+1$ points on a fixed line in $\mathbb{F}^n$ contains the entire line. As another example, recall that the closure of any Kakeya set in $\mathbb{F}_q^n$ with respect to homogeneous degree-$(q-1)$ polynomials is the entire domain $\mathbb{F}_q^n$ (this was proved by Dvir [Dvi09, Section 3] towards showing that any Kakeya set is necessarily of size at least $\binom{q+n-1}{n}$).

Following the latter example, it is natural to ask whether there exists a *very small* set $S \subseteq \mathbb{F}^n$ whose degree-d closure is *very large*. An initial observation towards answering this question is that a set $S \subseteq \mathbb{F}^n$ has *maximal* degree-d closure (i.e., $\mathtt{Cl}^{(\mathtt{d})}(S) = \mathbb{F}^n$) if and only if S is a hitting-set for degree-d polynomials. (This is since in both cases, the only degree-d polynomial that vanishes on S is the zero polynomial.)

Observation 4 *(maximal closure $\iff$ hitting-set). A set $S \subseteq \mathbb{F}^n$ is a hitting-set for (all) degree-d polynomials* if and only if $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| = q^n$.

Loosely speaking, the main result of Nie and Wang [NW15] extends Observation 4 by showing that that for any $S \subseteq \mathbb{F}^n$ it holds that $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| \leq \frac{|S|}{\binom{n+d}{d}} \cdot |\mathbb{F}|^n$. The meaning of this result is that, while there exist sets of size $|S| = \binom{n+d}{d}$ whose degree-d closure is $\mathbb{F}^n$, the degree-d closure of smaller sets decreases by a factor of at least $\frac{|S|}{\binom{n+d}{d}}$.[6]

We take another approach to extending Observation 4, by by establishing a connection between the study of small sets with large closures and the study of HSGs for polynomials that vanish rarely. Specifically, we show two-way implications between the statement that S is a hitting-set generator for polynomials *that vanish rarely*, and the statement that S has *large closure*. In more detail, we relate hitting-sets for polynomials that vanish with probability q^{-t} to sets with closure of size q^{n-t}:

Theorem 5 *(small sets with large closures versus hitting-sets for polynomials that vanish rarely). Let $\mathbb{F}$ be a field of size q, let $n \in \mathbb{N}$ and $t < d < n$, and let $S \subseteq \mathbb{F}^n$. Then,*

1. *If $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| > q^{n-t}$, then S is a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$.*

2. *If S is a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$, then $\left|\mathtt{Cl}^{(\mathtt{d/2(t+1)})}(S)\right| > \frac{1}{2} \cdot q^{n-t}$.*

Notice that Theorem 5 does not show a complete equivalence between the two notions, since in the second item the closure refers to degree $d/2t$ rather than to degree d.

[5] Note that $\mathcal{I}^{(d)}(S)$ is not an actual ideal in the ring of n-variate polynomials over $\mathbb{F}$, since multiplying $p \in \mathcal{I}^{(d)}(S)$ by another polynomial does not necessarily preserve the degree of p.

[6] Another result along these lines was recently proved by Beelen and Datta [BD18], who showed a tight upper-bound on the size of the variety induced by *any* subspace of degree-d polynomials (rather than only for varieties induced by a subspace of the form $\mathcal{I}^{(d)}(S)$ for some $S \subseteq \mathbb{F}^n$).

Thus, intuitively, Theorem 5 asserts that constructing a small set with a large degree-d closure is at least as hard as constructing a hitting-set for polynomials that vanish rarely; and while it also gives a converse reduction (in the second item), it is nevertheless possible that constructing a hitting-set for polynomials that vanish rarely is an easier problem. We also remark that the first item in Theorem 5 is almost immediate, whereas the second item requires more work (see Section 7 for details).

Lastly, we comment that one can obtain an upper-bound on the size of $\mathtt{Cl}^{(\mathtt{d})}(S)$ for small sets $S \subseteq \mathbb{F}^n$ by combining the first item in Theorem 5 with our lower bound from Theorem 1. (This is since the former asserts that sets with closure of size q^{n-t} are hitting-sets for $\mathcal{P}_{n,q,d,q^{-t}}$, whereas the latter asserts that any such hitting-set must be large.) However, the bounds obtained in this way are not stronger than the known bounds proved in [NW15]. For more details see Section 7.

2 Overview of our techniques

2.1 Combinatorial lower bounds from low-degree dispersers

The proofs of our lower bounds on HSGs for polynomials that vanish rarely rely on a *complexity-theoretic* approach, rather than on a direct algebraic analysis. Specifically, we reduce the problem of constructing HSGs for *arbitrary* polynomials to the problem of constructing HSGs for polynomials that *vanish rarely*; since we already know lower bounds for the former, we obtain lower bounds for the latter.

Specifically, given an arbitrary non-zero polynomial $p_0 \colon \mathbb{F}^m \to \mathbb{F}$, we will use a form of "error-reduction" for polynomials (akin to error-reduction for probabilistic algorithms; see below) to obtain another polynomial $p \colon \mathbb{F}^n \to \mathbb{F}$ such that:

1. The polynomial p vanishes rarely.
2. Any non-zero input for p can be mapped into a small list of inputs for p_0 that contains a non-zero input for p_0.

To define p, fix a (k,δ)-disperser $\mathsf{Disp} \colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$, for appropriate parameters k and δ that we will determine in a moment.[7] Then, p is the result of the following procedure: Given $z \in \mathbb{F}^n$, compute the 2^ℓ inputs $\{\mathsf{Disp}(z,i)\}_{i\in\{0,1\}^\ell}$, evaluate p_0 at each of these inputs, and output the disjunction of these evaluations; that is:

$$p(z) = \bigvee_{i\in\{0,1\}^\ell} p_0\left(\mathsf{Disp}(z,i)\right) .$$

The disperser Disp has the property that for every set $T \subseteq \mathbb{F}^m$ of density at least δ it holds that $\Pr_{z\in\mathbb{F}^n}[\forall i \ \mathsf{Disp}(z,i) \notin T] \le \varepsilon = 2^k/q^n$. We take T to be the set of elements in $\mathbb{F}^n$ on which p_0 does not vanish, and take δ to be the density of T (i.e., δ is the distance of the corresponding Reed-Muller code); we also let $k = (n-t)\cdot\log(q)$. Then, the

[7] A (k,δ)-disperser $\mathsf{Disp} \colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$ is a function such that for every $T \subseteq \mathbb{F}^m$ satisfying $|T|/|F|^m \ge \delta$, for all but at most 2^k of the inputs $z \in \mathbb{F}^n$ there exists $i \in \{0,1\}^\ell$ such that $\mathsf{Disp}(z,i) \in T$.

polynomial p vanishes on at most an $\varepsilon = 2^k/q^n = q^{-t}$ fraction of its inputs. Also, any non-zero input $z \in \mathbb{F}^n$ for p can be mapped to a list of 2^ℓ inputs $\{x_i = \mathsf{Disp}(z,i)\}_{i\in\{0,1\}^\ell}$ for p_0 such that for some $i \in \{0,1\}^\ell$ it holds that $p_0(x_i) \neq 0$, as we wanted.

The reduction above shows that if there exists a HSG with seed length s for polynomials $\mathbb{F}^n \to \mathbb{F}$ of degree $d = \deg(p)$ that vanish with probability ε, then there exists a corresponding HSG with seed length $s + \ell$ for all non-zero polynomials $\mathbb{F}^m \to \mathbb{F}$ of degree $d_0 = \deg(p_0)$. The known lower bound on the latter, which asserts that $s + \ell = \Omega(d_0 \cdot \log(m/d_0))$, yields a corresponding lower bound on the former.

While this is indeed our main idea, it unfortunately does not quite work as-is. The main challenge is that the reduction above incurs *significant overheads* that crucially deteriorate the lower bound. Most importantly, the *degree* of the polynomial increases (from $d_0 = \deg(p_0)$ to $d = \deg(p)$), and the number of variables also increases (from m to n); this affects us since we are interested in a lower bound as a function of n and d, whereas our lower bound is a function of m and d_0. Moreover, the lower bound deteriorates by an additive factor of ℓ, since each non-zero input $z \in \mathbb{F}^n$ for p yields 2^ℓ inputs for p_0, one of which is guaranteed to be non-zero. Thus, we want to modify the reduction above, in order to minimize the blowup in the degree and in the number of variables, and also minimize the seed length ℓ of the disperser.

A coding-theoretic perspective. One can view the procedure described above as amplifying the *weight* (i.e., the fraction of non-zero coordinates) of a codeword in the Reed-Muller code. At first glance, this task seems similar to the task of amplifying the *distance* of linear error-correcting codes; in particular, the disperser-based technique described above is technically reminiscent of the well-known distance amplification technique of Alon *et al.* [ABN+92].[8] However, the crucial difference is that we are interested in amplifying the weight to be much larger than $1 - 1/q$, and indeed our resulting subcode (of polynomials that vanish rarely) is a small and non-linear subcode of the Reed-Muller code. Moreover, as explained above, we will be particularly interested in the degree blow-up, which is a parameter specific to polynomial-based codes.

Warm-up: The setting of $d \ll q$. For simplicity, let us assume that $q = \text{poly}(n)$ and that $d \leq n^{.99}$. In this case the fraction δ of non-zeroes of p_0 is very close to one and we only need Disp to be a $(k, .99)$-disperser for $k = (n-t) \cdot \log(q)$.

Note that to compute p at an input $z \in \mathbb{F}^n$, we wish to compute $\mathsf{Disp}_i(z) = \mathsf{Disp}(z,i)$ as a function of z for each *fixed* value i of the seed. Since we want p to have degree as low as possible, we are interested in objects that we call low-degree dispersers: Informally, a disperser $\mathsf{Disp}\colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$ has low degree if for any $i \in \{0,1\}^\ell$ and $j \in [m]$, the polynomial $q_{i,j}(z) = \mathsf{Disp}(z,i)_j$ (i.e., $q_{i,j}(z)$ is the j^{th} output element of $\mathsf{Disp}(z,i)$ as a function of z) has low degree (see Definitions 16 and 17). Note that in our argument we only need the *existence* of a low-degree disperser (i.e., we do not need the low-degree

[8] The main differences are that we will use a specific disperser that is different from theirs, to minimize the degree blow-up; and that we handle alphabet reduction differently (using an OR function instead of code concatenation), since our target weight is much larger than $1 - 1/q$.

disperser to be efficiently computable); however, the dispersers that are obtained via naive probabilistic arguments do not have low degree.

Fortunately, in the current "warm-up" setting we can get a good (albeit non-optimal) lower bound even using the "naive disperser" that just performs uniform sampling: That is, the disperser that treats its input $z \in \mathbb{F}^n$ as n/m substrings of length m, and treats its seed as an index $i \in [n/m]$, and outputs the i^{th} substring of length m in z. Note that this disperser is *linear* (i.e., has degree one), since for a fixed seed, each output element is a projection of a corresponding input element.

We do encounter one other problem in implementing our idea in this setting, which is the degree blow-up that comes from the fact that p computes the OR function on the outputs of the disperser (recall that the OR function of 2^ℓ inputs has maximal degree $(q-1) \cdot 2^\ell$). To circumvent this problem, we replace the OR function with a multivalued OR function. Specifically, observe that in the reduction above it suffices that on any non-zero input $y \in \mathbb{F}^{2^\ell}$, the OR function will output *some* non-zero element (rather than map any non-zero y to $1 \in \mathbb{F}$). In contrast to the OR function, there exists a multivalued OR function of 2^ℓ elements with degree roughly 2^ℓ (see Proposition 10).

Working out the precise parameters, this approach transforms any p_0 of degree d_0 into a corresponding p of degree $d = d_0 \cdot 2^\ell = d_0 \cdot t \cdot \log(q)$, and for every $t \leq d/O(\log(q))$ implies a lower bound of $\Omega(d_0 \cdot \log(m/d_0)) - \ell = \Omega(d/t)$ on the seed length of HSGs for polynomials that vanish with probability q^{-t}. To improve this lower bound to match the bound stated in Theorem 1, we use a disperser that is better than the naive one, and utilize the techniques that are outlined below (see Section 6).

The more challenging setting of $d \gg q$. Observe that in the argument above we "paid" for the seed length ℓ of the disperser *twice*: One loss was a blow-up of 2^ℓ in the degree (since the multivalued OR function has degree 2^ℓ), and the other loss was that the lower bound on the seed length of the HSG decayed additively in ℓ (because our reduction maps any non-zero input for p to a list of 2^ℓ inputs for p_0). Also note that the first loss decreases the lower bound itself, whereas the second loss limits the values of t to which the lower bound applies (to ones for which $\ell \ll d_0 \cdot \log(m/d_0)$).

When $d \gg q$ these two losses may deteriorate our lower bound much more severely than in the "warm-up" setting. This is because when q was large we instantiated the disperser with the parameter $\delta = \Omega(1)$, and hence its seed length was relatively small, whereas in our current setting the value of $\delta = q^{-d_0/(q-1)}$ may be much smaller.[9]

Over prime fields this problem can be overcome by starting not from a lower bound for hitting all degree-d_0 polynomials, but rather from a lower bound for hitting a large subcode of the corresponding Reed-Muller code (i.e., a subcode with dimension linear in $\binom{m+d_0}{d_0}$) that still has distance $\Omega(1)$; see Appendix B for an explanation. To overcome the problem also over non-prime fields, we show a general method that, regardless of the disperser, *allows us to "pay" only an $O(t)$ factor in the degree blow-up*, instead of the 2^ℓ

[9] To demonstrate the problem, note that over fields of constant size, even a disperser with optimal parameters would yield a quadratic degree blow-up, regardless of t; that is, $d \geq 2^\ell \cdot d_0 \geq 2^{\log(t \cdot \log(q)/\delta)} \cdot d_0 = \Omega_q((d_0)^2 \cdot t)$, compared to the previous blow-up of $d = \Omega_q(d_0 \cdot t)$ when we had $\delta = \Omega(1)$.

factor. This method does not prevent the additive loss of ℓ in the seed length, and we will explain how this additive loss affects us in the end of the current section.

To explain this method, fix a disperser, and recall that our goal is to "hit" the set $\mathrm{G} \subseteq \mathbb{F}^n$ of inputs z such that for some $i \in \{0,1\}^\ell$ it holds that $p_0(\mathsf{Disp}(z,i)) \neq 0$ (since any $z \in \mathrm{G}$ maps to 2^ℓ inputs, one of which "hits" the original polynomial p_0). We think of the polynomial p above as a test of its input $z \in \mathbb{F}^n$ that distinguishes between G and $\mathbb{F}^n \setminus \mathrm{G}$ (i.e., p vanishes precisely on $\mathbb{F}^n \setminus \mathrm{G}$). Our initial approach to hit G was to construct a HSG for the test p, which would output some $z \in \mathrm{G}$.

The key observation is that constructing a HSG for p is an "overkill". Specifically, to hit G, *we can replace the test p by a distribution $\mathbf{p}$ over tests that distinguishes between* G *and* $\mathbb{F}^n \setminus \mathrm{G}$, *with high probability*, and still deduce that any HSG for the tests in the support of $\mathbf{p}$ outputs some $z \in \mathrm{G}$. That is, we replace the test p for G by a randomized test $\mathbf{p}$ for G such that the polynomials in the support of $\mathbf{p}$ have lower degree than p, and show that "hitting" the polynomials in the support of $\mathbf{p}$ still allows us to "hit" G. Moreover, since $\mathbf{p}$ "tests" a *dense* set G with small error, by an averaging argument almost all of the polynomials in the support of $\mathbf{p}$ *vanish rarely*; thus, it suffices to "hit" only the polynomials in the support of $\mathbf{p}$ that vanish rarely.

More accurately, let us instantiate our disperser with $k = (n-2t)\cdot\log(q)$, instead of $k = (n-t)\cdot\log(q)$, such that the density of G is $1-q^{-2t}$ (this is to allow for some slackness in the parameters). Then, the following holds:

Lemma 6 (informal; see Section 4) *Assume there exists a distribution* $\mathbf{p}$ *over polynomials* $\mathbb{F}^n \to \mathbb{F}$ *such that for every* $z \in \mathrm{G}$ *it holds that* $\Pr[\mathbf{p}(z) \neq 0] \geq 1 - q^{-2t}$ *and for every* $z \notin \mathrm{G}$ *it holds that* $\Pr[\mathbf{p}(z) = 0] = 1$. *Further assume that every polynomial in the support of* $\mathbf{p}$ *has degree* $O(d \cdot t)$. *Then, any hitting-set for polynomials of degree* $O(d\cdot t)$ *that vanish on at most* $2q^{-t}$ *of their inputs contains some* $z \in \mathrm{G}$.

Our construction of the specific distribution $\mathbf{p}$ that we use is simple: Starting from the construction of p above, instead of taking an $\mathtt{OR}$ of the evaluations of p_0 on the entire output-set of the disperser (i.e., on all seeds), we *sample from the seeds of the disperser*. More accurately, to sample a polynomial $f \sim \mathbf{p}$, we uniformly sample $2t$ vectors $a^{(1)}, ..., a^{(2t)} \in \mathbb{F}^{2^\ell}$, and output the polynomial

$$f(z) = \mathtt{OR}_{j\in[2t]}\left(\sum_{i\in 2^\ell} a_i^{(j)} \cdot p_0(\mathsf{Disp}(z,i))\right) .$$

To see why this distribution works, observe that if $z \in \mathrm{G}$ then a random $\mathbb{F}$-linear sum of the elements $\{\mathsf{Disp}(x,i)\}_{i\in\{0,1\}^\ell}$ will be non-zero with probability $1-1/q$, whereas if $z \notin \mathrm{G}$ then such a sum will be zero, with probability one. Thus, a random polynomial in $\mathbf{p}$ computes the disjunction of $2t$ such random sums, and it is straightforward to see that its "error probability" is q^{-2t} and its degree is $O(d_0 \cdot t)$ (assuming that the disperser is linear). Using Lemma 6, any HSG for polynomials of degree $O(d_0\cdot t)$ that vanish on at most q^{-2t} of their inputs outputs some $z \in \mathrm{G}$. We therefore reduced the problem of constructing a HSG for p_0 to the problem of constructing a HSG for polynomials of degree $d = O(d_0 \cdot t)$ that vanish on at most q^{-2t} of their inputs.

The last missing piece is that we need a concrete disperser to instantiate the argument with, and the parameters of the disperser will determine the lower bound that we get. Furthermore, recall that we are losing an additive factor of ℓ in the lower bound, and thus any lower bound that we get using this approach applies only to values of t such that $\ell \ll d_0 \cdot \log(m/d_0)$. Specifically, the approach above gives the following lemma (for simplicity, we state it only for linear dispersers):

Lemma 7 (linear dispersers yield lower bounds on HSGs for polynomials that vanish rarely; informal, see Corollary 26) *Let $d_0 < m$ be integers, let $\mathbb{F}$ be a field of size q, and let $t \in \mathbb{N}$. Assume that for $k = (n-2t) \cdot \log(q)$ and $\delta = q^{-d_0/(q-1)}$ there exists a linear (k, δ)-disperser* Disp: $\mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$. *Then, for $d = 4d_0 \cdot t$, if $\ell \le \frac{d}{8t} \cdot \log(mt/d)$, then the seed length for any HSG for $\mathcal{P}_{n,q,d,2q^{-t}}$ is $\Omega\left((d/t) \cdot \log\left(mt/d\right)\right)$.*

Note that to get a good lower bound using Lemma 7 we want a *linear* disperser $\mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$ for large min-entropy $k = (n-2t) \cdot \log(q)$ that has small seed length ℓ and large output length m.[10] In particular, if there exists a *linear* disperser with *optimal* parameters, then a lower bound of $\Omega((d/t) \cdot \log(nt/d))$ would follow for essentially all settings of the parameters (see Corollary 27).

Our lower bounds (i.e., Theorem 1 and its extensions) will be proved by instantiating Lemma 7 with specific useful dispersers. To prove Theorem 1 and some of its extensions (i.e., Theorems 28 and 33), we use a linear disperser that we obtain by modifying the extractor by Shaltiel and Umans [SU05]; the original extractor works over the binary alphabet, and we modify it to a linear disperser over an arbitrary field $\mathbb{F}_q$ (see Section 6 for details). We prove another lower bound, which applies only to fields of constant size (see Theorem 34), using a linear disperser that is based on the recent construction of "linear 1-local expanders" by Goldreich [Gol16], following Viola and Wigderson [VW17] (see Section 6.3). More details are given in Section 6.

2.2 Explicit upper bound over $\mathbb{F}_2$

To construct the explicit HSG for polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish rarely in Theorem 3 we generalize a construction of [GW14], by extending a proof approach from [Tel19]. In high-level, we reduce the problem of constructing a HSG for polynomials that vanish rarely to the problem of constructing a PRG for arbitrary *low-degree polynomials*, and then use the explicit PRG of Viola [Vio09b] for low-degree polynomials.

In more detail, we say that a polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ is approximated by a distribution $\mathbf{h}$ over polynomials $h\colon \mathbb{F}_2^n \to \mathbb{F}_2$ if for every $x \in \mathbb{F}_2^n$ it holds that $\Pr_h[\mathbf{h}(x) = p(x)] \ge .99$. Our first step is to show that any polynomial $p \in \mathcal{P}_{n,2,d,q^{-t}}$ can be approximated by a distribution $\mathbf{h}$ over polynomials of degree $d-t$. To do so, let $\Delta_a(p)$ be the directional derivative of p in direction $a \in \mathbb{F}_2^n$ (i.e., the function $\Delta_a p(x) = p(x+a) + p(x)$). We sample $h \sim \mathbf{h}$ by uniformly sampling $\vec{a} = a^{(1)}, ..., a^{(k)} \in \mathbb{F}_2^n$, where $k = t - O(1)$,

[10] Moreover, since our error $\delta = q^{-d_0/(q-1)}$ might be large, we want good dependency of the parameters ℓ and m on the error δ.

and outputting the polynomial $h_{\vec{a}} = \Delta_{a^{(k)}}\Delta_{a^{(k-1)}}...\Delta_{a^{(1)}}(p) + 1$; that is, we derive p in k random directions, and "negate" the output.

Note that indeed $\deg(h_{\vec{a}}) = d - t + O(1)$. Now, for any fixed $x \in \mathbb{F}_2^n$ and non-empty $S \subseteq [k]$, the probability over $\vec{a}$ that $p\left(x + \sum_{i\in S} a^{(i)}\right) = 1$ is at least $1 - 2^{-t}$ (since p vanishes with probability at most 2^{-t}, and $x + \sum_{i\in S} a^{(i)}$ is uniform in $\mathbb{F}_2^n$). Thus, by a union bound, with probability at least .99 over the choice of $\vec{a}$, for every non-empty $S \subseteq [k]$ it holds that $p\left(x + \sum_{i\in S} a^{(i)}\right) = 1$. In this case, we have that $h_{\vec{a}}(x) = \sum_{S\subseteq[k]} p\left(x + \sum_{i\in S} a^{(i)}\right) + 1 = p(x) + (2^k - 1) + 1 = p(x)$. Hence, the distribution $\mathbf{h}$ also has the property that for every $x \in \mathbb{F}_2^n$ it holds that $\Pr[\mathbf{h}(x) = p(x)] \geq .99$.

Our next observation is similar to the "randomized tests" technique mentioned in Section 2.1: We show that if a distribution $\mathbf{h}$ over low-degree polynomials approximates p, then a pseudorandom generator for the polynomials in the support of $\mathbf{h}$ (with sufficiently small constant error) also "hits" p (for a proof see Section 4). Combining the two claims, we get a reduction from the problem of constructing a HSG for $\mathcal{P}_{n,2,d,q^{-t}}$ to the problem of constructing a PRG (with small constant error) for arbitrary polynomials of degree $d - t + O(1)$. Thus, the PRG of Viola [Vio09b] for such polynomials, which uses a seed of length $O((d-t)\cdot(2^{d-t} + \log(n)))$, is also a HSG for $\mathcal{P}_{n,2,d,2^{-t}}$.

On the tightness of the reduction above. Recall that there is a gap between the seed length of the explicit HSG above and the seed length of the *non-explicit* HSG from Theorem 2, which is $O\left((d-t)\cdot\log(\frac{n}{d-t})\right)$. We note that to close this gap, one does not need to improve the *reduction* detailed above, but only the *explicit PRG for arbitrary polynomials* (i.e., Viola's construction). Specifically, if there exists an explicit PRG for all polynomials of degree $d' = d - t + O(1)$ with seed length $O(d' \cdot \log(n/d'))$ (matching the non-explicit upper-bound for such PRGs), then the reduction above yields a HSG for $\mathcal{P}_{n,2,d,2^{-t}}$ with seed length $O((d-t)\cdot\log(n/(d-t)))$.

3 Preliminaries

We denote random variables by boldface. For an alphabet Σ and $n \in \mathbb{N}$, we denote the uniform distribution over Σ^n by $\mathbf{u}_n$, where Σ will be clear from context.

3.1 Polynomials over finite fields

We consider multivariate polynomials over a finite field. A polynomial $p\colon \mathbb{F}^n \to \mathbb{F}$ of degree d can be viewed as a codeword in the corresponding Reed-Muller code; thus, if p is non-zero, then the relative distance of the corresponding Reed-Muller code, which is stated below, lower bounds the fraction of inputs on which p does not vanish.

Theorem 8 (distance of the Reed-Muller code; see, e.g., [GRS19, Lemma 9.4.1]). For any $d, q \in \mathbb{N}$, let $a = \lfloor d/(q-1) \rfloor$ and $b = d \pmod{q-1}$. The relative distance of the Reed-Muller code of degree d over alphabet q is $\delta_{RM}(d,q) = q^{-a}\cdot(1 - b/q) \geq q^{-d/(q-1)}$.

The $\texttt{OR}\colon \mathbb{F}^k \to \mathbb{F}$ function maps any non-zero input $z \in \mathbb{F}^k \setminus \{0^k\}$ to $1 \in \mathbb{F}$, and maps 0^k to zero. We consider a generalization of this function, which we call *multivalued OR*; a multivalued OR function maps any non-zero $z \in \mathbb{F}^k \setminus \{0^k\}$ to *some* non-zero element (i.e., different non-zero inputs may yield different outputs), while still mapping 0^k to zero. That is:

Definition 9 (multivalued OR functions) *For any finite field $\mathbb{F}$, we say that a polynomial $\texttt{mvOR}\colon \mathbb{F}^k \to \mathbb{F}$ is a* multivalued OR function *if $\texttt{mvOR}(0^k) = 0$, but $\texttt{mvOR}(x) \neq 0$ for every $x \neq 0^k$.*

For a fixed field $\mathbb{F}$ there are many different k-variate multivalued OR functions. Indeed, the standard $\texttt{OR}$ function is a multivalued OR function, but it has maximal degree $k \cdot (q-1)$ as a polynomial. We will need k-variate multivalued OR functions that are of much lower degree (i.e., degree approximately k); such functions can be constructed relying on well-known techniques in algebraic geometry (see [Tel19, Proposition 7.3] for the construction, and see e.g. [CLO15, Exercise 8] for a reference to the well-known underlying techniques):

Proposition 10 (low-degree multivalued OR function) *Let $\mathbb{F}$ be a finite field and let $k \in \mathbb{N}$. Then, there exists a multivalued OR function $\texttt{mvOR}\colon \mathbb{F}^k \to \mathbb{F}$ that is computable by a polynomial of degree less than $2k$.*

3.2 Hitting-set Generators

We recall the standard definitions of hitting-set generators (HSGs), of hitting-set generators and of pseudorandom generators (PRGs). Recall that HSGs for a class of polynomials need to produce a set of inputs such that any polynomial from the class evaluates to *non-zero* on some input in the set. That is:

Definition 11 (hitting-set generator) *Fix a field $\mathbb{F}$, and let $d, n \in \mathbb{N}$. A function $H\colon \{0,1\}^\ell \to \mathbb{F}^n$ is a* hitting-set generator *for a set of functions $\mathcal{P} \subseteq \{\mathbb{F}^n \to \mathbb{F}\}$ if for every non-zero function $p \in \mathcal{P}$ there exists $s \in \{0,1\}^\ell$ satisfying $p(H(s)) \neq 0$. In this case, the set $S = \{H(s) : s \in \{0,1\}^\ell\}$ is called a* hitting-set *for $\mathcal{P}$.*

Definition 12 (explicit hitting-set generators) *Let $\ell, q, d\colon \mathbb{N} \to \mathbb{N}$, let $\{\mathbb{F}_{q(n)}\}_{n \in \mathbb{N}}$ such that for every $n \in \mathbb{N}$ it holds that $\mathbb{F}_{q(n)}$ is a field of size $q(n)$, and let $H = \{H_n\colon \{0,1\}^{\ell(n)} \to \mathbb{F}^n_{q(n)}\}$ such that for every $n \in \mathbb{N}$ it holds that H_n is a hitting-set generator for polynomials of degree $d(n)$. We say that H is* polynomial-time computable *if there exists an algorithm that gets as input $s \in \{0,1\}^\ell$ and outputs $H_n(s)$ in time $\mathrm{poly}(\ell, \log(q), n)$.*

The standard definition of PRGs for polynomials in $p\colon \mathbb{F}^n \to \mathbb{F}$ that we will use is as follows. Consider the distribution over $\mathbb{F}$ that is obtained by uniformly choosing $x \in \mathbb{F}^n$ and outputting $p(x)$, and the distribution over $\mathbb{F}$ that is obtained by choosing a seed s for a PRG G and outputting $p(G(s))$. We require that the statistical distance between the two distributions is small. That is:

Definition 13 (pseudorandom generator) *Fix a field $\mathbb{F}$, let $d, n \in \mathbb{N}$, and let $\rho > 0$. A function $G\colon \{0,1\}^\ell \to \mathbb{F}^n$ is a* pseudorandom generator with error ρ *for polynomials of degree d if for every polynomial $p\colon \mathbb{F}^n \to \mathbb{F}$ of degree at most d it holds that*

$$\sum_{\sigma \in \mathbb{F}} \left| \Pr_{s \in \{0,1\}^\ell}[p(G(s)) = \sigma] - \Pr_{x \in \mathbb{F}^n}[p(x) = \sigma] \right| \leq \rho \, .$$

An alternative standard definition of PRGs for polynomials requires that the "character distance" $\left|\mathbb{E}_{x \in \mathbb{F}^n}[\mathsf{e}^{p(x)}] - \mathbb{E}_x[\mathsf{e}^{p(G(s))}]\right|$ will be small, where e is any (fixed, non-trivial) character of $\mathbb{F}$. The "character distance" and the statistical distance are equivalent, up to a multiplicative factor of $\sqrt{q-1}$ (see [Lov09, Lemma 2.4]).

Lastly, we recall the standard lower bound on the size of hitting-sets for polynomials of degree d (for completeness, we include its proof) and state the complementary upper-bound that is obtained by a standard probabilistic argument.

Fact 14 (lower bound on the size of hitting-sets for linear subspaces) *Let $\mathbb{F}$ be a finite field, let $n \in \mathbb{N}$, and let $\mathcal{C} \subseteq \{\mathbb{F}^n \to \mathbb{F}\}$ be a linear subspace of dimension $D = \dim(\mathcal{C})$. Then, any hitting-set for $\mathcal{C}$ has at least D points. In particular, for any $d < n$, any hitting-set for degree-d polynomials $\mathbb{F}^n \to \mathbb{F}$ has size at least $\binom{n+d}{d}$, and correspondingly the seed length of any hitting-set generator for such polynomials is at least $d \cdot \log(n/d)$.*

Proof: Assume towards a contradiction that $S \subseteq \mathbb{F}^n$ is a hitting-set for $\mathcal{C}$ with $D-1$ points. Consider a generator matrix M for the linear subspace $\mathcal{C}$, which is a full-rank $D \times |\mathbb{F}|^n$ matrix over $\mathbb{F}$ whose D rows span $\mathcal{C}$. Let M_S be the projection of M to the $D-1$ columns corresponding to the points in S.

Since M_S is a $D \times (D-1)$ matrix, there exists a non-trivial linear combination of the rows of M_S that yields the all-zero row. Now, since S is a hitting-set for $\mathcal{C}$, the only function in $\mathcal{C}$ that vanishes on all of S is the all-zero function; in particular, any non-trivial linear combination of the rows of M_S that yields the all-zero row (which induces a corresponding function in $\mathcal{C}$ that vanishes on S) also yields the all-zero row in M. Thus, we obtain a non-trivial linear combination of the rows of M that yields the all-zero row, contradicting the hypothesis that M is full-rank.

The "in particular" part follows since the dimension of the corresponding Reed-Muller code (which is a linear subspace of $\mathbb{F}^n$) is $D = \binom{n+d}{d} > \binom{n}{d} > (n/d)^d$, where we used the hypothesis that $d < n$. ■

Fact 15 (upper bound on the size of hitting-sets) *Let $\mathbb{F}$ be a finite field, let $n \in \mathbb{N}$, and let $d < n$. Then, there exists a (non-explicit) hitting-set generator for polynomials $\mathbb{F}^n \to \mathbb{F}$ of degree d with seed length $O(d \cdot \log(n/d) + \log\log(q))$.*

Proof: The number of degree-d polynomials is at most $q^{\binom{n+d}{d}}$, and each of them vanishes on at most $1 - \delta$ of its inputs, where $\delta \geq q^{-d/(q-1)}$ is the distance of the corresponding Reed-Muller code. Thus, if we randomly choose

$$O\left((1/\delta) \cdot \binom{n+d}{d} \cdot \log(q)\right) < O\left(q^{d/(q-1)} \cdot \binom{2n}{d} \cdot \log(q)\right)$$

elements in $\mathbb{F}^n$, with high probability we obtain a hitting-set for degree-d polynomials. The number of bits that we need to sample an element from this hitting-set is

$$O\left(\frac{d}{q-1}\cdot\log(q)+d\cdot\log(n/d)+\log\log(q)\right)<O\left(d\cdot\log(n/d)+\log\log(q)\right)\ .$$

■

3.3 Dispersers and extractors

We recall the standard definition of dispersers $\mathsf{Disp}:[N]\times\{0,1\}^\ell\to[M]$, where we identify the domain N with the vector space $\mathbb{F}^n$ and the range M with the vector space $\mathbb{F}^m$.

Definition 16 (disperser) *Let $\mathbb{F}$ be a finite field of size $q=|\mathbb{F}|$. A function $\mathsf{Disp}\colon\mathbb{F}^n\times\{0,1\}^\ell\to\mathbb{F}^m$ is a (k,δ)-disperser if for every $T\subseteq\mathbb{F}^m$ of size $|T|\geq\delta\cdot q^m$, the probability over $x\in\mathbb{F}^n$ that for all $i\in\{0,1\}^\ell$ it holds that $\mathsf{Disp}(x,i)\notin T$ is less than $2^k/q^n$. The value ℓ is the* seed length *of the disperser.*

In this work we are interested in dispersers that can be computed by low-degree polynomials. Specifically, we require that for each fixed seed $s\in\{0,1\}^\ell$ and output index $i\in[m]$, the function that maps any $z\in\mathbb{F}^n$ to the i^{th} output of Disp at z with seed s (i.e., $z\mapsto\mathsf{Disp}(z,s)_i$) has low degree as a polynomial $\mathbb{F}^n\to\mathbb{F}$.

Definition 17 (degree of a disperser) *We say that a disperser $\mathsf{Disp}:\mathbb{F}^n\times\{0,1\}^\ell\to\mathbb{F}^m$ has* degree *d if for every fixed $s\in\{0,1\}^\ell$ and $i\in[m]$, the polynomial $p_{s,i}\colon\mathbb{F}^n\to\mathbb{F}$ defined by $p_{s,i}(z)=\mathsf{Disp}(z,s)_i$ is of degree at most d. If $d=1$, then we say the disperser is* linear*.*

Recall that there are two standard dispersers that are linear: The naive disperser, which treats its input $z\in\mathbb{F}^n$ as a list of samples from $\mathbb{F}^m$ and its seed as an index of a sample in this list; and the subspace sampler, which treats its input as the description of an affine subspace in $\mathbb{F}^m$ and its seed as an index of an element in the subspace. Nevertheless, these dispersers have disadvantages (small output length and large seed length, respectively), and in our results we will use more sophisticated linear dispersers (see Section 6 for details).

Alternatively, one can verify that Definition 16 is equivalent to the following definition: Disp is a (k,δ)-disperser if for any random variable $\mathbf{x}\sim\mathbb{F}^n$ with min-entropy[11] k, the support of $\mathsf{Disp}(\mathbf{x},\mathbf{u}_\ell)$ covers at least $(1-\delta)q^m$ elements from $\mathbb{F}^m$. Although dispersers will be our main pseudorandom object, we will sometimes work with the stronger notion of an *extractor*. While in dispersers we only care about covering almost all of $\mathbb{F}^m$, in extractors we want to do it *uniformly*, i.e., we require $\mathsf{Ext}(\mathbf{x},\mathbf{u}_\ell)$ to be δ-close to the uniform distribution $\mathbf{u}_m$ over $\mathbb{F}^m$. Formally:

Definition 18 (extractor) *Let $\mathbb{F}$ be a finite field of size $q=|\mathbb{F}|$. A function $\mathsf{Ext}\colon\mathbb{F}^n\times\{0,1\}^\ell\to\mathbb{F}^m$ is a (k,δ)-extractor if for every random variable $\mathbf{x}\sim\mathbb{F}^n$ with min-entropy k it holds that $\mathsf{Ext}(\mathbf{x},\mathbf{u}_\ell)$ is δ-close to $\mathbf{u}_m$. The value ℓ is the* seed length *of the extractor.*

[11] A random variable $\mathbf{x}$ has min-entropy k if for every $x\in\mathsf{supp}(\mathbf{x})$ is holds that $\Pr[\mathbf{x}=x]\leq 2^{-k}$.

As the support size of a distribution which is δ-close to $\mathbf{u}_m$ is at least $(1-\delta)q^m$, any (k,δ)-extractor is readily a (k,δ)-disperser.

4 Randomized tests

The proofs of both our upper bounds and of our lower bounds will rely on a general observation that we now explain. The observation is essentially from [Tel19, Sections 2.1 & 4], following a proof idea from [BV10].

Assume that we want to deterministically *find* an element in a set $G \subseteq \mathbb{F}^n$. A standard way to do so is to show that G can by decided by a simple algorithm p (e.g., p is a low-degree polynomial), which we think of as a *simple test*. Then, a hitting-set generator for p outputs an element in G. Our goal now is to find an element in G using a hitting-set generator for tests that are *simpler* than p. The basic observation is that if G can be decided, with high probability, by a *distribution* $\mathbf{p}$ *over simple tests*, then a hitting-set generator with small density for the tests in the *support* of $\mathbf{p}$ outputs an element in G (see [Tel19, Observation 2.1]). The advantage is that instead of constructing a deterministic test p we can now construct a *randomized test* $\mathbf{p}$, whose complexity is potentially lower than that of p; that is, the complexity of the tests in the support of the distribution $\mathbf{p}$ may be lower than the complexity of the deterministic test p.

The observation above can be extended in various ways (see [Tel19] for details), and we will apply it in two specific settings. In the first setting, the set G is dense (i.e., $\Pr_{x\in\mathbb{F}^n}[x \in G] \geq .99$), and can be decided by a distribution $\mathbf{p}$ over polynomials with small "one-sided" error (i.e., every $x \in G$ is accepted with high probability, and every $x \notin G$ is rejected with probability one). We show that in this case, any hitting-set generator for the polynomials in the support of $\mathbf{p}$ that *vanish rarely* outputs an element in G (and this holds without any density requirement from the HSG).

Lemma 19 (randomized tests, a special case) *Let $\varepsilon, \rho > 0$ such that $\varepsilon + \rho < 1$, and let $G \subseteq \mathbb{F}^n$ be such that $\Pr_{x\in\mathbb{F}^n}[x \in G] \geq 1-\varepsilon$. Assume that there exists a distribution $\mathbf{p}$ over polynomials $p \colon \mathbb{F}^n \to \mathbb{F}$ such that:*

1. *For every fixed $x \in G$ it holds that $\Pr[\mathbf{p}(x) \neq 0] \geq 1-\rho$.*
2. *For every fixed $x \notin G$ it holds that $\Pr[\mathbf{p}(x) = 0] = 1$.*

Let $\mathbf{w}$ be a distribution over $\mathbb{F}^n$ such that for every $p \colon \mathbb{F}^n \to \mathbb{F}$ in the support of $\mathbf{p}$ that vanishes on at most a $\sqrt{\rho+\varepsilon}$ fraction of its inputs there exists $w \sim \mathbf{w}$ such that $p(w) \neq 0$. Then, there exists $w \sim \mathbf{w}$ such that $w \in G$.

Proof: Consider the behavior of a random polynomial $p \sim \mathbf{p}$ on a pseudorandom input $w \sim \mathbf{w}$. On the one hand, we have that

$$\Pr[\mathbf{p}(\mathbf{w}) = 0] \geq \Pr[\mathbf{w} \notin G] \cdot \min_{x\notin G}\left\{\Pr[\mathbf{p}(x) = 0]\right\} = \Pr[\mathbf{w} \notin G] \,. \tag{4.1}$$

On the other hand, denote by P the set of polynomials in the support of $\mathbf{p}$ that vanish on at most $\sqrt{\rho+\varepsilon}$ of the inputs $x \in \mathbb{F}^n$; then, we have that

$$\Pr[\mathbf{p}(\mathbf{w}) \neq 0] \geq \Pr[\mathbf{p} \in P] \cdot \min_{p \in P} \{\Pr[p(\mathbf{w}) \neq 0]\} \ . \tag{4.2}$$

Next, note that

$$\Pr_{x \in \mathbb{F}^n}[\mathbf{p}(x) \neq 0] \geq \Pr_{x \in \mathbb{F}^n}[x \in G] \cdot \min_{x \in G}\{\Pr[\mathbf{p}(x) \neq 0]\} > 1-(\rho+\varepsilon) \ ,$$

so by using Markov's inequality we get that $\Pr[\mathbf{p} \in P] > 1-\sqrt{\rho+\varepsilon} > 0$. Plugging this into Eq. (4.2), we deduce that $\Pr[\mathbf{p}(\mathbf{w}) \neq 0] > 0$, or equivalently that $\Pr[\mathbf{p}(\mathbf{w}) = 0] < 1$. Thus, using Eq. (4.1) we deduce that $\Pr[\mathbf{w} \notin G] < 1$. ■

In the second setting that we will be interested in, we want to "fool" a polynomial $p \colon \mathbb{F}^n \to \mathbb{F}$ using a pseudorandom generator for polynomials that are simpler than p (e.g., they are of lower degree). This is indeed possible if there is a distribution $\mathbf{h}$ over polynomials that are simpler than p such that for every fixed $x \in \mathbb{F}^n \to \mathbb{F}$ it holds that $\Pr[\mathbf{h}(x) = p(x)]$ is high. In the following statement, it is useful to think of $\zeta \colon \mathbb{F} \to \mathbb{C}$ as a non-trivial character, which implies that $\max_{v,w \in \mathbb{F}}\{|\zeta(v)-\zeta(w)|\} = 2$.

Lemma 20 (randomized tests, a PRG version for polynomials; see Lemma 4.4 in [Tel19], extending Lemma 23 in [BV10]) *Let $n \in \mathbb{N}$, let $\mathbb{F}$ be any finite field, let $\varepsilon > 0$. Also, let $\zeta \colon \mathbb{F} \to \mathbb{C}$, and let $\delta = \max_{v,w \in \mathbb{F}}\{|\zeta(v)-\zeta(w)|\}$. Let $p \colon \mathbb{F}^n \to \mathbb{F}$, and assume that there exists a distribution $\mathbf{h}$ over polynomials $\mathbb{F}^n \to \mathbb{F}$ such that for every fixed $x \in \mathbb{F}^n$ it holds that $\Pr[\mathbf{h}(x) = p(x)] \geq 1-\varepsilon$. Finally, let $\mathbf{w}$ be a distribution over $\mathbb{F}^n$ such that for every polynomial h in the support of $\mathbf{h}$ it holds that $\left|\mathbb{E}_{x \in \mathbb{F}^n}[\zeta(h(x))] - \mathbb{E}[\zeta(h(\mathbf{w}))]\right| \leq \varepsilon$. Then,*

$$\left|\mathbb{E}_{x \in \mathbb{F}^n}[\zeta(p(x))] - \mathbb{E}[\zeta(p(\mathbf{w}))]\right| \leq (2\delta+1) \cdot \varepsilon \ .$$

Since the proof of Lemma 20 is simple, we include it for completeness.

Proof of Lemma 20: Let $\mathbf{u}_n$ be the uniform distribution over $\mathbb{F}^n$. For simplicity of notation, define $p' = \zeta \circ p \colon \mathbb{F}^n \to \mathbb{C}$, and for every h in the support of $\mathbf{h}$, define $h' = \zeta \circ h \colon \mathbb{F}^n \to \mathbb{C}$. Also denote by $\mathbf{h}'$ the distribution that is obtained by sampling $h \sim \mathbf{h}$ and outputting $h' = \zeta \circ h$. By the triangle inequality,

$$\begin{aligned}\left|\mathbb{E}[p'(\mathbf{u}_n)] - \mathbb{E}[p'(\mathbf{w})]\right| \leq & \left|\mathbb{E}[p'(\mathbf{u}_n)] - \mathbb{E}[\mathbf{h}'(\mathbf{u}_n)]\right| \\ & + \left|\mathbb{E}[\mathbf{h}'(\mathbf{u}_n)] - \mathbb{E}[\mathbf{h}'(\mathbf{w})]\right| \\ & + \left|\mathbb{E}[\mathbf{h}'(\mathbf{w})] - \mathbb{E}[p'(\mathbf{w})]\right| \ . \end{aligned} \tag{4.3}$$

To upper bound the first item in Equation (4.3), note that

$$\begin{aligned}\left|\mathbb{E}[p'(\mathbf{u}_n)] - \mathbb{E}[\mathbf{h}'(\mathbf{u}_n)]\right| &\le \mathbb{E}_{x\sim\mathbb{F}^n, h\sim\mathbf{h}}\left[\left|p'(x) - h'(x)\right|\right] \\ &\le \mathbb{E}_{x\in\mathbb{F}^n}\left[\Pr_{h\sim\mathbf{h}}[h(x) \neq p(x)] \cdot \max_{v,w\in\mathbb{F}}\{|\zeta(v) - \zeta(w)|\}\right] \\ &\le \delta \cdot \varepsilon\,,\end{aligned}$$

where the last inequality holds because for every fixed $x \in \mathbb{F}^n$ we have that $\Pr_{h\sim\mathbf{h}}[h(x) \neq p(x)] \le \varepsilon$. The third item in Equation (4.3) is similarly upper bounded by $\delta \cdot \varepsilon$, by replacing the uniform choice of $x \in \mathbb{F}^n$ with a choice of $x \sim \mathbf{w}$.

To upper bound the second item in Equation (4.3), note that

$$\left|\mathbb{E}[\mathbf{h}'(\mathbf{u}_n)] - \mathbb{E}[\mathbf{h}'(\mathbf{w})]\right| \le \mathbb{E}_{h\sim\mathbf{h}}\left[\left|\mathbb{E}[h'(\mathbf{u}_n)] - \mathbb{E}[h'(\mathbf{w})]\right|\right] \le \varepsilon\,,$$

where we used the hypothesis that for every polynomial h in the support of $\mathbf{h}$ it holds that $\left|\mathbb{E}_{x\in\mathbb{F}^n}[\zeta(h(x))] - \mathbb{E}[\zeta(h(\mathbf{w}))]\right| \le \varepsilon$. ■

Applying Lemma 20 to the special case of $\mathbb{F} = \mathbb{F}_2$ with $\zeta(x) = (-1)^x$, we obtain the following useful corollary:

Corollary 21 (randomized tests applied to PRGs for $\mathbb{F}_2$ polynomials) *Let $n \in \mathbb{N}$ and let $\varepsilon > 0$. Let $p\colon \mathbb{F}_2^n \to \mathbb{F}_2$, and assume that there exists a distribution $\mathbf{h}$ over polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ such that for every fixed $x \in \mathbb{F}_2^n$ it holds that $\Pr[\mathbf{h}(x) = p(x)] \ge 1 - \varepsilon$. Finally, let $\mathbf{w}$ be a distribution over $\mathbb{F}_2^n$ such that for every polynomial h in the support of $\mathbf{h}$ it holds that $\left|\Pr_{x\in\mathbb{F}_2^n}[h(x) = 1] - \Pr[h(\mathbf{w}) = 1]\right| \le \varepsilon$. Then,*

$$\left|\Pr_{x\in\mathbb{F}_2^n}[p(x) = 1] - \Pr[p(\mathbf{w})]\right| \le 5\varepsilon\,.$$

5 Upper bounds over $\mathbb{F}_2$

In this section we prove Theorems 2 and 3; that is, we construct explicit and non-explicit hitting-set generators for polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish rarely.

We define the weight of a polynomial $p\colon \mathbb{F}^n \to \mathbb{F}$ to be $wt(p) = \Pr_{x\in\mathbb{F}^n}[p(x) \neq 0]$. Indeed, in this paper we are interested in polynomials with very high weight. Kaufman, Lovett, and Porat [KLP12] proved a near-tight upper-bound on the number of polynomials with very low weight when $\mathbb{F} = \mathbb{F}_2$; as a consequence, we get the following non-explicit hitting-set generator on polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish rarely:

Theorem 22 (non-explicit HSGs for $\mathbb{F}_2$ polynomials that vanish rarely, following [KLP12]) *Let $n, d, t \in \mathbb{N}$ where $t < d \le n$. Then, the number of polynomials in $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish with probability at most 2^{t-d} is at most $2^{O(d^2\cdot t/(d-t+1)!\cdot n^{d-t+1})}$. In particular, there exists a hitting-set generator for this set of polynomials with seed length $O\left((d-t)\cdot\log\left(\frac{n}{d-t}\right)\right)$.*

Proof: We define an injective mapping $\Phi\colon \{\mathbb{F}_2^n \to \mathbb{F}_2\} \to \{\mathbb{F}_2^n \to \mathbb{F}_2\}$ that maps every degree-d polynomial p that vanishes on at most 2^{-t} of its inputs to a degree-d polynomial $\Phi(p)$ whose weight is at most 2^{-t}. Indeed, the mapping is simply $\Phi(p) = p + 1$ (i.e., for every $x \in \mathbb{F}_2^n$ it holds that $\Phi(p)(x) = p(x) + 1$). By [KLP12, Theorem 14] (using the parameter values $k = d - t + 1$ and $\varepsilon = 2^{-t}$), the number of polynomials with weight at most 2^{-t} is at most $2^{O(d^2 \cdot t/(d-t+1)! \cdot n^{d-t+1})}$. Since Φ is injective, the number of polynomials that vanish on at most 2^{-t} of their inputs is also at most $N = 2^{O(d^2 \cdot t/(d-t+1)! \cdot n^{d-t+1})}$.

Thus, a set of $O(\log(N)) = O(d^2 \cdot t/(d-t+1)! \cdot n^{d-t+1})$ uniformly-chosen elements in $\mathbb{F}_2^n$ "hits", with high probability, every polynomial that vanishes on at most 2^{-t} of its inputs. The seed length required to sample from such a set is

$$\begin{aligned} &O\Big((d-t+1)\cdot\log(n) + \log(d\cdot t) - (d-t)\cdot\log(d-t)\Big) \\ &\quad = O\Big((d-t+1)\cdot\log(n) - (d-t)\cdot\log(d-t)\Big) && (d\cdot t \le n^2) \\ &\quad = O\Big((d-t)\cdot\log(n/(d-t))\Big) . \end{aligned}$$

■

We mention that Abbe, Shpilka, and Wigderson [ASW15] proved a tighter upper-bound on the number of polynomials with low weight, which replaces the d^2 term in the result in [KLP12, Theorem 14] by a smaller term. It is still an open problem to replace this term by some universal constant (such a result would match a lower bound from [KLP12, Lemma 15]). However, even a solution to this open problem would not improve the result in Theorem 22.[12]

To construct an *explicit* (i.e., polynomial-time computable) hitting-set generator for polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish rarely, we generalize results from previous works [GW14; Tel19]. For the construction we will need the pseudorandom generator of Viola [Vio09b] for low-degree polynomials.

Theorem 23 (Viola's PRG for low-degree polynomials [Vio09b]) *For $n, d' \in \mathbb{N}$ and $\varepsilon > 0$, there exists a polynomial-time computable pseudorandom generator for polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ of degree d' with seed length $d' \cdot \log(n) + O(d' \cdot 2^{d'} \cdot \log(1/\varepsilon))$.*

Theorem 24 (explicit hitting-set generator for $\mathbb{F}_2$ polynomials that vanish rarely) *For every $n, d, t \in \mathbb{N}$ such that $d > t + 4$ there exists a polynomial-time computable hitting-set generator with seed length $O\left((d-t)\cdot\left(2^{d-t} + \log(n)\right)\right)$ for the set of polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ of degree d that vanish on at most 2^{-t} of their inputs.*

Proof: We show that for every polynomial $p\colon \mathbb{F}_2^n \to \mathbb{F}_2$ of degree d that vanishes on at most 2^{-t} of its inputs there exists a distribution $\mathbf{h}$ over polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ of degree

[12] This is because in our application we refer to the seed length, in which case the term d^2 only "contributes" the term $\log(d \cdot t) < 2 \cdot \log(n)$, which is dominated by the term $O((d-t)\cdot\log(n))$.

$(d-t)+4$ such that for every $x \in \mathbb{F}_2^n$ it holds that $\Pr[p(x) = \mathbf{h}(x)] \geq 15/16$. Then, we use Corollary 21 to deduce that any pseudorandom generator with error $1/16$ for polynomials of degree $(d-t)+4$ is also a pseudorandom generator for p with error less than $1/2$ (and is thus a hitting-set generator for p, which vanishes on at most half of its inputs). In particular, we use the pseudorandom generator from Theorem 23 for polynomials of degree $d-t+4$, which has seed length $O\left((d-t)\cdot\left(2^{d-t}+\log(n)\right)\right)$.

To define the distribution $\mathbf{h}$, recall that the discrete directional derivative operator on polynomials $p\colon \mathbb{F}_2^n \to \mathbb{F}_2$ for direction $a \in \mathbb{F}_2^n$ is defined by $\Delta_a(p) = p(x+a)+p(x)$. The iterated operator for $\vec{a} = a^{(1)}, \ldots, a^{(k)} \in \mathbb{F}_2^{n\cdot t}$ is defined in the natural way, and $\Delta_{\vec{a}}(p) = \sum_{S\subseteq[k]} p\left(x + \sum_{i\in S} a^{(i)}\right)$. For $k = t-4$, sampling $h \sim \mathbf{h}$ is done by uniformly and independently choosing $\vec{a} = a^{(1)}, \ldots, a^{(k)} \in \mathbb{F}_2^n$, and outputting the polynomial

$$h = h_{\vec{a}} = \Delta_{\vec{a}}(p) + 1.$$

Note that h is of degree $d-k = (d-t)+4$, and that for every $x \in \mathbb{F}_2^n$, the probability that $\mathbf{h}(x) = p(x)$ is at least $15/16$. This is the case since for every fixed $x \in \mathbb{F}_2^n$, if for every non-empty $S \subseteq [k]$ it holds that $p(x + \sum_{i\in S} a^{(i)}) = 1$ then $h(x) = p(x) + (2^k - 1) + 1 = p(x)$; and for every non-empty $S \subseteq [k]$, the probability over the choice of h that $p(x+\sum_{i\in S} a^{(i)}) = 1$ is at least $1-2^{-t}$. ■

6 Lower bounds over general finite fields

In this section we prove our lower bounds on the seed length of HSGs for polynomials that vanish rarely. First, in Section 6.1 we give the general framework for deriving lower bounds from low-degree dispersers, corresponding to the high-level description in Section 2.1 (i.e., we prove Lemma 7). Then, we prove three incomparable lower bounds, by instantiating this framework with specific dispersers that are suitable for the corresponding parameter settings.

Our first and main lower bound, which is presented in Section 6.2, is a generalization of Theorem 1. This lower bound is of the form $\Omega((d/t)\cdot\log(n^{1-\Omega(1)}t/d))$, and holds under complicated conditions on the degree d and on t; in particular, for $d \leq n^{.49}$ as in Theorem 1, it holds for all values of t up to $\Omega(d)$. (See Theorem 28.)

Then, in Section 6.3 we prove two additional lower bounds, which hold in two more specific settings but have advantages over the foregoing bound. The first lower bound holds only when $d \leq q$ (i.e., when the corresponding Reed-Muller code has distance $\Omega(1)$); this lower bound is of the same form as in Theorem 28, but holds for higher degrees up to $d \leq n^{1-\Omega(1)}$ without complicated conditions on d and t (see Theorem 33). The second lower bound holds only over fields of constant size; this lower bound is of the stronger form $\Omega((d/t)\cdot\log(nt/d))$,[13] and holds for degrees d up to $\Omega(n)$, but only for value of $t \lesssim \sqrt{d}$ (see Theorem 34).

[13] Recall, from Corollary 27, that this is the lower is that would be obtained if there exists a linear disperser with optimal parameters.

6.1 Sampling from the seeds of a disperser

In this section we prove general results that use low-degree dispersers to reduce hitting arbitrary polynomials to hitting polynomials that vanish rarely (and thus deduce lower bounds for the latter); this follows the high-level explanations that were presented in Section 2.1. The following proposition specifies the reduction itself, and the subsequent corollary specifies the lower bounds that we can obtain using the reduction.

Proposition 25 (reducing hitting polynomials to hitting polynomials that vanish rarely by sampling from the seeds of a disperser) *Let $m, d_0 \in \mathbb{N}$, let $\mathbb{F}$ be a field of size q, and let $\delta = \delta_{RM}(d_0, q)$. For $k < \log(q^n)$, let $\varepsilon = 2^k/q^n$, let $\rho < 1 - \varepsilon$, and let $r = \log_q(1/\rho)$. Assume that:*

1. *There exists a (k, δ)-disperser $\mathsf{Disp}\colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$ of degree $d_{\mathsf{Disp}} \in \mathbb{N}$.*
2. *There exists a hitting-set $W \subseteq \mathbb{F}^n$ for polynomials $\mathbb{F}^n \to \mathbb{F}$ of degree $d = 2d_0 \cdot r \cdot d_{\mathsf{Disp}}$ that vanish on at most $\sqrt{\rho + \varepsilon}$ of their inputs.*

Then, there exists a hitting-set $W_0 \subseteq \mathbb{F}^m$ for polynomials $\mathbb{F}^m \to \mathbb{F}$ of degree d_0 such that $|W_0| \leq |W| \cdot 2^\ell$.

Proof: For $L = 2^\ell$, let $W_0 = \{\mathsf{Disp}(z,i) : z \in W, i \in [L]\}$. We will prove that W_0 is a hitting-set for polynomials $\mathbb{F}^m \to \mathbb{F}$ of degree d_0.

To do so, fix any non-zero polynomial $f\colon \mathbb{F}^m \to \mathbb{F}$ of degree d_0. Let $V = \{x \in \mathbb{F}^m : f(x) = 0\}$ be the set of points on which f vanishes, and let $\mathrm{G} = \{z \in \mathbb{F}^n : \exists i \in [L], \mathsf{Disp}(z,i) \notin V\}$ be the set of inputs $z \in \mathbb{F}^n$ for Disp such that for some $i \in [L]$ it holds that f does not vanish on $\mathsf{Disp}(z,i)$. Note that G has density at least $1 - \varepsilon$; this is the case since $|V|/q^m \leq 1 - \delta$ (and recall that δ is the distance of the corresponding Reed-Muller code and f is non-zero), and since Disp is a (k, δ)-disperser.

Note that W_0 is a hitting-set for f if and only if $\Pr_{z \in W}[z \in \mathrm{G}] > 0$. We will prove that $\Pr_{z \in W}[z \in \mathrm{G}] > 0$ using Lemma 19. To construct the distribution $\mathbf{p}$ over polynomials in $\mathbb{F}^n \to \mathbb{F}$ needed for the hypothesis of the lemma, fix a multivalued OR polynomial $\mathtt{mvOR}\colon \mathbb{F}^r \to \mathbb{F}$ of degree less than $2r$ as in Proposition 10. Then, sampling $p \sim \mathbf{p}$ is equivalent to the following random process:

> Uniformly and independently choose $\alpha^{(1)}, ..., \alpha^{(r)} \in \mathbb{F}^L$, and output the polynomial $p(z) = \mathtt{mvOR}\left(\sum_{i \in [L]} \alpha_i^{(1)} \cdot f(\mathsf{Disp}(z,i)), ..., \sum_{i \in [L]} \alpha_i^{(r)} \cdot f(\mathsf{Disp}(z,i))\right)$.

Note that each $p \sim \mathbf{p}$ has degree less than $d = d_{\mathsf{Disp}} \cdot d_0 \cdot 2r$. Also note that for any $z \notin \mathrm{G}$ we have that $\Pr[\mathbf{p}(z) = 0] = 1$, whereas for any $z \in \mathrm{G}$ we have that $\Pr[\mathbf{p}(z) \neq 0] \geq 1 - q^{-r} = 1 - \rho$. Using Lemma 19 and the hypothesis that W is a hitting-set for polynomials that vanish on at most $\sqrt{\rho + \varepsilon}$ of their inputs, we deduce that $\Pr_{z \in W}[z \in \mathrm{G}] > 0$, as we wanted. ■

Using the reduction from Proposition 25, and relying on the unconditional lower bound from Fact 14, we obtain the following result, which uses low-degree dispersers to deduce lower bounds on HSGs for polynomials that vanish rarely:

Corollary 26 (a lower bound by sampling from the seeds of a disperser) *Let $m, d_0 \in \mathbb{N}$ such that $d_0 < m$, let $\mathbb{F}$ be a field of size q, and let $\delta = \delta_{RM}(d_0, q)$. For $t \in \mathbb{N}$ and $k = (n-2t) \cdot \log(q)$, assume that there exists a linear (k, δ)-disperser $\mathsf{Disp}\colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$. Then, any hitting-set $W \subseteq \mathbb{F}^n$ for polynomials in $\mathbb{F}^n \to \mathbb{F}$ of degree $d = 4d_0 \cdot t$ that vanish on at most $\sqrt{2} \cdot q^{-t}$ of their inputs has size at least $\binom{m+d_0}{d_0} \cdot 2^{-\ell}$. In particular, the seed length for any such hitting-set is at least*

$$\Omega\left(\frac{d}{t} \cdot \log\left(\frac{m \cdot t}{d}\right)\right) ,$$

provided that $t \leq \frac{\log(mt/d)}{8\ell} \cdot d$.

Proof: We use Proposition 25 with the parameter values $\varepsilon = \rho = q^{-2t} \leq 1/4$ (such that $r = 2t$) and $d_{\mathsf{Disp}} = 1$, and rely on the fact that any hitting-set $W_0 \subseteq \mathbb{F}^m$ for all polynomials $\mathbb{F}^m \to \mathbb{F}$ of degree d_0 has size at least $\binom{m+d_0}{d_0}$ (i.e., on Fact 14). The seed length (in bits) for sampling from the hitting-set is thus at least $d_0 \cdot \log(m/d_0) - \ell = \frac{d}{4t} \cdot \log(4mt/d) - \ell \geq \Omega((d/t) \cdot \log(mt/d)$, where the last inequality is due to the hypothesis that $\frac{d}{4t} \cdot \log(mt/d) \geq 2\ell$. ■

Finally, note that if there exists a linear (k, δ)-disperser $\mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$ with optimal parameters, then we get a lower bound of $\Omega((d/t) \cdot \log(nt/d))$ for essentially all settings of the parameters. That is:

Corollary 27 (lower bounds assuming an optimal linear disperser) *Assume that for every $n, q, k \in \mathbb{N}$ and $\delta > 0$ there exists a* linear *(k, δ)-disperser $\mathsf{Disp}\colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$ where $\ell = \log(n \cdot \log(q) - k) + \log(1/\delta) + O(1)$ and $m \cdot \log(q) = k + \ell - \log\log(1/\delta) - O(1)$. Then, for every constant $c > 1$ there exists a constant $\gamma > 0$ such that the following holds.*

Let $n, q, d, t \in \mathbb{N}$ such that $q \leq 2^{n^c}$, and $d < n/2$, and $t \leq \gamma \cdot n$, and $\frac{q-1}{\log(q)} \cdot \log(nt/d) \geq 1/\gamma$. Then, the seed length of any HSG for $\mathcal{P}_{n,q,d,\sqrt{2}\cdot q^{-t}}$ is at least $\Omega\left(\frac{d}{t} \cdot \log\left(\frac{n \cdot t}{d}\right)\right)$.

Proof: Let $d_0 = d/4t$, and let $a = d_0/(q-1)$ such that $\delta = \delta_{RM}(d_0, q) \geq q^{-a}$. When instantiating the hypothesized linear disperser with parameters n and $k = (n-2t) \cdot \log(q)$ and $\delta = q^{-a}$, it has seed length $\ell = O(\log(t \cdot \log(q)) + (d/4t) \cdot (\log(q)/(q-1)))$ and output length $m = \Omega(n)$. Relying on Corollary 26, we get a lower bound of $\Omega\left((d/t) \cdot \log(n \cdot (t/d))\right)$, assuming that $d_0 < m$ (which holds since we assumed that $d < n/2$) and that $t \leq \frac{\log(nt/d)}{8\ell} \cdot d$. Thus, we just need to verify the latter condition.

We verify the condition by a case analysis. The first case is when $t \geq \sqrt{d/4(q-1)}$, which implies that the seed length is $\ell = O(\log(t \cdot \log(q)))$. The condition in this case holds since $\log(nt/d) = \Omega(\log(n))$ and $q \leq 2^{\mathrm{poly}(n)}$, which implies that $\frac{\log(nt/d)}{8\ell} = \Omega(1)$. The second case is when $t < \sqrt{d/4(q-1)}$, which implies that the seed length is $\ell = O((d/t) \cdot \log(q)/(q-1))$. The condition in this case holds if and only if $\frac{q-1}{\log(q)} \cdot \log(nt/d)$ is larger than a sufficiently large constant, which is our hypothesis. ■

6.2 The main lower bound: Proof of Theorem 1

In this section we prove lower bounds that hold also when the degree is much larger than the field size (i.e., $d \gg q$). Specifically, we will prove the following, more general version of Theorem 1:

Theorem 28 (a lower bound using the Shaltiel-Umans linear disperser; a more general version of Theorem 1) *For any two constants $\gamma > 0$ and $\gamma' > 0$ there exists a constant $\gamma'' > 0$ such that the following holds. Let $n, d, t, q \in \mathbb{N}$ such that $q \le n^{1/\gamma'}$ is a prime power, $d \le n/4$, and:*

- *(essentially all values of $\varepsilon = q^{-t}$) $t \le \gamma'' \cdot \frac{\log(nt/d)}{\log(n)} \cdot d$.*
- *(auxiliary condition that holds for typical settings) $\frac{q-1}{\log(q)} \cdot \log(nt/d) \ge 1/\gamma''$.*
- *(main condition: d/t is upper-bounded) $d/t \le \gamma'' \cdot \min\left\{\frac{q-1}{\log(q)} \cdot n^{\gamma}, n^{1-(\gamma+\gamma')}\right\}$.*

Then, the seed length of any HSG for $\mathcal{P}_{n,q,d,\sqrt{2}\cdot q^{-t}}$ is at least $\Omega\left(\frac{d}{t} \cdot \log\left(\frac{n^{1-(\gamma+\gamma')}\cdot t}{d}\right)\right)$.

To deduce Theorem 1 from Theorem 28, note that if we are willing to assume that $d \le n^{.49}$, then we can choose $\gamma = .499$ and $\gamma' > 0$ that is sufficiently small, and the three conditions in Theorem 28 hold for every $q \le n^{1/\gamma'}$ and $t \le \gamma'' \cdot d$.

To prove Theorem 28 we will instantiate Corollary 26 with a linear disperser that we will construct relying on the extractor of Shaltiel and Umans [SU05]. Recall that [SU05] constructed an extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ by first constructing what they called a q-ary extractor, whose output lies in a field of size $\mathrm{poly}(n)$ and only satisfies a relatively-weak unpredictability requirement, and then transforming the q-ary extractor to a standard extractor over the binary alphabet (the transformation follows an idea of Ta-Shma, Zuckerman, and Safra [TSZS06]).

We want to construct a low-degree disperser $\mathsf{Disp}\colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q$ where the field $\mathbb{F}_q$ is of size much smaller than $\mathrm{poly}(n)$ (i.e., $q \le n^{\gamma'}$ for some small constant $\gamma' > 0$). To do so, we take as a starting-point their construction of a q_0-ary extractor from [SU05], where $q_0 = \mathrm{poly}(n)$, and then generalize their transformation of q_0-ary extractors to standard extractors (and in particular dispersers) such that the resulting extractor is both over the field $\mathbb{F}_q$, rather than over a binary alphabet, and linear.

Towards presenting the construction, let us first recall the definition of q_0-ary extractors and the main construction of such objects from [SU05].

Definition 29 (q_0-ary extractor) *For $n, k, m, \ell \in \mathbb{N}$ and $\rho > 0$, and a prime power $q_0 \in \mathbb{N}$, we say that $\mathsf{Ext}_0 : \mathbb{F}_{q_0}^n \times \{0,1\}^\ell \to \mathbb{F}_{q_0}^m$ is a (k, ρ) q_0-ary extractor if for every random variable $\mathbf{x}$ over $\mathbb{F}_{q_0}^n$ with min-entropy at least k, and every $i \in [m]$, and every function $P : \mathbb{F}_{q_0}^{i-1} \to \mathbb{F}_{q_0}^{\rho^{-2}}$, it holds that $\Pr_{x\sim\mathbf{x},u\sim\mathbf{u}_\ell}[P(\mathsf{Ext}_0(x,u)_1, ..., \mathsf{Ext}_0(x,u)_{i-1}) \ni \mathsf{Ext}_0(x,u)_i] \le \rho$.*

Theorem 30 ([SU05, Theorem 4.5, Item 1]) *There exists a universal constant $c > 1$ such that the following holds. Let $n_0, q_0, k, m, r, h \in \mathbb{N}$ and $\rho > 0$ such that q_0 is a prime power, and the following inequalities hold:*

1. *(Sufficiently large auxiliary parameters h and r)* $n_0 \leq \binom{h+r-1}{r}$.

2. *(Sufficiently large field)* $q_0 \geq c \cdot \frac{(h \cdot r)^2}{\rho^4}$.

3. *(Sufficiently small output length)* $m \leq \frac{k - \log(1/\rho)}{c \cdot h \cdot r \cdot \log(q_0)}$.

Then, there exists an $r \times r$ matrix A over $\mathbb{F}_{q_0}$ such that the following holds. Let $\mathsf{Ext}_0 : \mathbb{F}_{q_0}^{n_0} \times \{0,1\}^{r \cdot \log(q_0)} \to \mathbb{F}_{q_0}^m$ be defined by $\mathsf{Ext}_0(x, v) = p_x(A^1 \cdot v) \circ p_x(A^2 \cdot v) \circ ... \circ p_x(A^m \cdot v)$, where v is interpreted as an element in $\mathbb{F}_{q_0}^r$, and $p_x : \mathbb{F}_{q_0}^r \to \mathbb{F}_{q_0}$ is the r-variate polynomial of total degree $h-1$ whose coefficients are specified by x. Then, Ext_0 is a (k, ρ) q_0-ary extractor.

Note that in [SU05] the input of the extractor is represented in binary and interpreted as n_0 elements in $\mathbb{F}_q$, whereas in Theorem 30 we considered the input as n_0 elements in $\mathbb{F}_q$. The two formulations are equivalent, since a random variable over $\mathbb{F}_{q_0}^{n_0}$ has min-entropy k if and only if the corresponding random variable over $\{0,1\}^{n_0 \cdot \log(q_0)}$ has min-entropy k. Also note that [SU05, Lemma 4.4] showed that A can be constructed in time $q_0^{O(r)}$ (by an exhaustive search over the field $\mathbb{F}_{(q_0)^r}$), and deduced that the extractor is efficiently computable; however, we will not use this property of the extractor.

We now present the transformation of q_0-ary extractors to standard extractors whose inputs and outputs are vectors over $\mathbb{F}_q$, where $q \ll q_0$; as mentioned above, the proof generalizes an idea from [TSZS06]. The intuition for this transformation is the following. Consider the output distribution of a q_0-ary extractor as consisting of blocks of elements from $\mathbb{F}_q$, where each block represents a single element from $\mathbb{F}_{q_0}$; by definition, the output distribution of a q_0-ary extractor is "next-element unpredicatable", and hence the distribution of elements from $\mathbb{F}_q$ is a *block source* (see, e.g., [Vad12, Section 6.3.1]). Following Nisan and Zuckerman [NZ96], we compose the q_0-ary extractor with a strong extractor over $\mathbb{F}_q$ that outputs a single element (and maps each block to a single element) and obtain an extractor over $\mathbb{F}_q$. We will specifically use a single-output extractor that is obtained from a *linear list-decodable code* (see, e.g., [TSZ04, Claim 4.1]), relying on well-known constructions of such codes.[14]

Proposition 31 (transforming a q_0-ary extractor into a standard extractor over $\mathbb{F}_q$) *Let $\rho > 0$, let q be a prime power, let $q_0 = q^\Delta$ for some $\Delta \in \mathbb{N}$, and let $\mathfrak{C} : \mathbb{F}_q^\Delta \to \mathbb{F}_q^{\bar{\Delta}}$ be a $(1 - 1/q - \rho, \rho^{-2})$-list-decodable code. Assume that $\mathsf{Ext}_0 : \mathbb{F}_{q_0}^{n_0} \times \{0,1\}^{\ell_0} \to \mathbb{F}_{q_0}^m$ is a (k, ρ) q_0-ary extractor. Let $\mathsf{Ext} : \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$, where $n = n_0 \cdot \Delta$ and $\ell = \ell_0 + \log(\bar{\Delta})$, be defined by*

$$\mathsf{Ext}(x, (y, j)) = \mathfrak{C}(\mathsf{Ext}_0(\hat{x}, y)_1)_j \circ ... \circ \mathfrak{C}(\mathsf{Ext}_0(\hat{x}, y)_m)_j \,,$$

where $\hat{x} \in \mathbb{F}_{q_0}^{n_0}$ is the vector that is represented by $x \in \mathbb{F}_q^{n_0 \cdot \Delta}$. Then, Ext is a $(k, 2qm \cdot \rho)$-extractor.

[14] In fact, since in our case the output of the q_0-ary extractor is not only unpredictable but also unpredictable by predictors that output a *list* of elements, we use a simpler proof that does not go through the notion of strong extractors.

Proof: Assuming towards a contradiction that Ext is not a $(k, 2qm \cdot \rho)$-extractor, we will show that Ext_0 is not a (k, ρ) q_0-ary extractor. For simplicity, throughout the argument we do not distinguish between $x \in \mathbb{F}_q^{n_0 \cdot \bar{\Delta}}$ and $\hat{x} \in \mathbb{F}_{q_0}^{n_0}$.

Since Ext is not a $(k, 2qm \cdot \rho)$-extractor, there exists a random variable $\mathbf{x}$ over $\mathbb{F}_q^n$ with min-entropy at least k such that $\mathsf{Ext}(\mathbf{x}, \mathbf{u}_\ell)$ is $(2qm \cdot \rho)$-far from the uniform distribution over $\mathbb{F}_q^m$. By a standard argument showing that next-element unpredictability of a distribution implies that the distribution is close to uniform (see Appendix A), there exists an index $i \in [m]$ and a function $f : \mathbb{F}_q^{i-1} \to \mathbb{F}_q$ such that

$$\Pr_{x \sim \mathbf{x}, (y,j) \sim \mathbf{u}_\ell} [f(\mathsf{Ext}(x, (y,j))_1, ..., \mathsf{Ext}(x, (y,j))_{i-1}) = \mathsf{Ext}(x, (y,j))_i] > 1/q + 2\rho \,. \tag{6.1}$$

For any fixed $(x, y) \in \mathbb{F}_q^n \times \{0,1\}^{\ell_0}$, let $c_{x,y}$ be the string that is obtained by encoding each of the first $i-1$ output elements of $\mathsf{Ext}_0(x, y)$ by the code $\mathfrak{C}$; that is, $c_{x,y} = \mathfrak{C}(\mathsf{Ext}_0(x,y)_1), ..., \mathfrak{C}(\mathsf{Ext}_0(x,y)_{i-1}) \in (\mathbb{F}_q^{\bar{\Delta}})^{i-1}$. Also, for any $j \in [\bar{\Delta}]$, let $c_{x,y}^{(j)} \in \mathbb{F}_q^{i-1}$ be the string that is obtained by projecting each of the $i-1$ symbols of $c_{x,y}$ into its j^{th} coordinate; that is, $c_{x,y}^{(j)} = \mathfrak{C}(\mathsf{Ext}_0(x,y)_1)_j, ..., \mathfrak{C}(\mathsf{Ext}_0(x,y)_{i-1})_j$. Note that

$$c_{x,y}^{(j)} = \mathsf{Ext}(x, (y,j))_1, ..., \mathsf{Ext}(x, (y,j))_{i-1}.$$

It follows from Equation (6.1) by an averaging argument that for at least a ρ-fraction of the pairs $(x, y) \in \mathbb{F}_q^n \times \{0,1\}^\ell$ it holds that

$$\begin{aligned} 1/q + \rho &< \Pr_{j \in [\bar{\Delta}]} [f(\mathsf{Ext}(x, (y,j))_1, ..., \mathsf{Ext}(x, (y,j))_{i-1}) = \mathsf{Ext}(x, (y,j))_i] \\ &= \Pr_{j \in [\bar{\Delta}]} [f(c_{x,y}^{(j)}) = \mathfrak{C}(\mathsf{Ext}_0(x,y)_i)_j] \,; \end{aligned}$$

in other words, with probability at least ρ over choice of (x, y), for more than a $1/q + \rho$ fraction of the coordinates $j \in [\bar{\Delta}]$ it holds that $f(c_{x,y}^{(j)})$ correctly outputs the j^{th} coordinate of $\mathfrak{C}(\mathsf{Ext}_0(x,y)_i)$.

Let us now construct a predictor $P \colon \mathbb{F}_{q_0}^{i-1} \to \mathbb{F}_{q_0}^{\rho^{-2}}$ for Ext_0 that succeeds with probability more than ρ. The predictor P gets $i-1$ inputs $\mathsf{Ext}_0(x,y)_1, ..., \mathsf{Ext}_0(x,y)_{i-1}$, and computes $r = f\left(c_{x,y}^{(1)}\right), ..., f\left(c_{x,y}^{(\bar{\Delta})}\right) \in \mathbb{F}_q^{\bar{\Delta}}$. We think of r as a possibly-corrupt codeword in the code $\mathfrak{C}$. Since $\mathfrak{C}$ is $(1 - 1/q - \rho, \rho^{-2})$-list-decodable, there are at most ρ^{-2} messages whose encoding is of distance at most $1 - 1/q - \rho$ from r; the predictor outputs this list. By the argument above, with probability at least ρ over choice of (x, y) it holds that r will be of distance less than $1 - 1/q - \rho$ from $\mathfrak{C}(\mathsf{Ext}_0(x,y)_i)$. For every such (x, y), the list that P outputs will contain $\mathsf{Ext}_0(x,y)_i$. ■

We now combine Theorem 30 and Proposition 31 to obtain a linear (k, δ)-disperser $\mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$ with output length $m = k/n^{\Omega(1)}$ and seed length $\ell = O(\log(n/\delta))$.

Theorem 32 (an adaptation of the Shaltiel-Umans extractor to a linear disperser over general finite fields) *For any two constants $\gamma, \gamma' > 0$ the following holds. Let $n, k, q \in \mathbb{N}$ such that $k \geq n^{\gamma+\gamma'}$ and $q \leq n^{1/\gamma'}$, and let $\delta \geq 2^{-n^{\gamma}+\log(2qn)}$. Then, there exists a linear (k, δ)-disperser $\mathsf{Disp} : \mathbb{F}_q^n \times \{0,1\}^{\ell} \to \mathbb{F}_q^m$, where $\ell = O_{\gamma'}(\log(n/\delta))$ and $m = \Omega_{\gamma'}\left(k/n^{\gamma+\gamma'}\right)$.*

Proof: For a sufficiently large universal constant $c \in \mathbb{N}$, we choose q_0 to be a power of q in the interval $[(nq/\delta)^c, (nq/\delta)^{2c}]$, denote $\Delta = \log_q(q_0) = O(\log(n/\delta))$, and let $n_0 = n/\Delta$. We also let $h = \left\lceil n^{\gamma'} \right\rceil$, let $r = O(1)$ be a sufficiently large constant, let $m = c_{\gamma'} \cdot k/n^{\gamma+\gamma'}$, where $c_{\gamma'} > 0$ is a sufficiently small constant that depends on γ', and let $\rho = \delta/2qm$. We instantiate Theorem 30 with the foregoing parameters, to obtain a q_0-ary (k, ρ)-extractor $\mathsf{Ext}_0 \colon \mathbb{F}_{q_0}^{n_0} \times \{0,1\}^{O(\log(n))} \to \mathbb{F}_{q_0}^m$. (The conditions of Theorem 30 hold due to our hypothesized lower bounds for k and for δ.)

We now want to use Proposition 31 to transform Ext_0 into a standard extractor. As a list-decodable code we use the concatenation of the Reed-Solomon code with the Hadamard code over $\mathbb{F}_q$, which yields a linear code $\mathbb{F}_q^{\Delta} \to \mathbb{F}_q^{\bar{\Delta}}$ with relative distance $1 - 1/q - \rho^2$ such that $\bar{\Delta} = O(\Delta/\rho^2)^2$.[15] By an appropriate version of the Johnson bound (see, e.g., [GS01, Theorem 1]), the code is $(1 - 1/q - \rho, \rho^{-2})$-list-decodable. Using Proposition 31 with this code, we obtain a (k, δ)-extractor $\mathsf{Ext} \colon \mathbb{F}_q^n \times \{0,1\}^{\ell} \to \mathbb{F}_q^m$, where $\ell = O(\log(n)) + \log(\bar{\Delta}) = O(\log(n/\delta))$.

Finally, let us verify that Ext is linear. Recall that for any fixed seed $(y, j) \in \{0,1\}^{r \cdot \log(q_0) + \log(\bar{\Delta})}$ and output location $i \in [m]$, we want to show that the function that outputs the i^{th} output element of $\mathsf{Ext}(x, (y, j))$ is linear. To see this, note that the i^{th} output element of $\mathsf{Ext}(x, (y, j))$ can be computed from $x \in \mathbb{F}_q^n$ by first computing a predetermined output element of $\mathsf{Ext}_0(x, y)$, which we denote by $z_{y,i}(x) \in \mathbb{F}_q^{\Delta}$, and then computing the j^{th} output element of $\mathfrak{C}(z_{y,i}(x))$, where $\mathfrak{C} : \mathbb{F}_q^{\Delta} \to \mathbb{F}_q^{\bar{\Delta}}$ is a linear code. Thus, it suffices to show that the mapping of $x \in \mathbb{F}_q^n$ to $z_{y,i} \in \mathbb{F}_q^{\Delta}$ is $\mathbb{F}_q$-linear; this is indeed the case since $z_{y,i}(x)$ is the evaluation of the multivariate polynomial p_x over $\mathbb{F}_{q_0}$ whose coefficients are described in x (i.e., each block of Δ elements in x describes a coefficient of p_x) at the fixed point in $\mathbb{F}_{q_0}^r$ described by y. ■

Finally, we deduce our lower bound from Theorem 28 using Corollary 26 with the linear disperser from Theorem 32.

Proof of Theorem 28: Let $d_0 = d/4t$, and let $a = d_0/(q-1)$ such that $\delta = \delta_{RM}(d_0, q) \geq q^{-a}$. We instantiate the linear disperser from Theorem 32 with parameters n and $k = (n - 2t) \cdot \log(q)$ and $\delta = q^{-a} \geq 2^{-n^{\gamma}+\log(2qn)}$, and with the parameters $\gamma > 0$ and $\gamma' > 0$. The conditions of Theorem 32 hold due to our hypotheses that $d/t \leq \gamma'' \cdot \frac{q-1}{\log(q)} \cdot n^{\gamma}$ (which implies that $\delta \geq 2^{-n^{\gamma}+\log(2qn)}$) and that $d \leq n/4$ (which implies that $k = \Omega(n)$). For these parameters, the disperser has seed length $\ell = O(\log(n/\delta)) = O(\log(n) + (d/4t) \cdot (\log(q)/(q-1)))$ and output length $m = \Omega(n^{1-(\gamma+\gamma')})$.

[15] We use this specific code merely for simplicity, and since its sub-optimal parameters do not significantly affect the final parameters of the construction.

Relying on Corollary 26, we get a lower bound of $\Omega\left((d/t)\cdot\log(n^{1-(\gamma+\gamma')}\cdot(t/d))\right)$, assuming that $d_0 < m$ (which holds since $d/4t < \gamma''\cdot n^{1-(\gamma+\gamma')}$) and that $t \leq \frac{\log(nt/d)}{8\ell}\cdot d$. Thus, we just need to verify the latter condition.

We verify the condition by a case analysis. The first case is when $\log(n) > \frac{d\log(q)}{4t(q-1)}$, which implies that the seed length is $\ell = O(\log(n))$; then, the condition that we want holds due to our hypothesis $t \leq \gamma''\cdot\frac{\log(nt/d)}{\log(n)}\cdot d$. In the second case we have that $\frac{d\log(q)}{4t(q-1)} \geq \log(n)$, which implies that the seed length is $\ell = O\left(\frac{d\log(q)}{t(q-1)}\right)$; then, the condition holds since we assumed that $\frac{q-1}{\log(q)}\cdot\log(nt/d) \geq 1/\gamma''$. ■

6.3 Improved lower bounds in two special cases

In this section we extend Theorem 28 by proving the two additional lower bounds that were described in the beginning of Section 6. Recall that these lower bounds have advantages over the lower bound in Theorem 28 but hold only in two specific settings.

The first lower bound is for the setting of $d \leq q$. Recall, from Section 2, that this setting is relatively easier to handle, since the corresponding Reed-Muller code has constant relative distance. To prove the lower bound we will instantiate Corollary 26 with the disperser from Theorem 30 used with the error parameter $\delta = \Omega(1)$.[16]

Theorem 33 (a lower bound when $d \leq q$) *For any constant $\eta > 0$ there exists a constant $\eta' > 0$ such that following holds. Let $n, q, d, t \in \mathbb{N}$ such that q is a prime power, and $d/t \leq \min\{3q, \eta'\cdot n^{1-2\eta}\}$, and $t \leq \eta'\cdot d$. Then, the seed length of any HSG for $\mathcal{P}_{n,q,d,\sqrt{2}\cdot q^{-t}}$ is at least $\Omega\left(\frac{d}{t}\cdot\log\left(\frac{n^{1-\eta}\cdot t}{d}\right)\right)$.*

Proof: Let $d_0 = d/4t$, and note that $d_0 \leq (3/4)\cdot q$, which implies that $\delta = \delta_{RM}(d_0, q) \geq 1/4$. We instantiate the disperser from Theorem 30 with parameters n and $k = (n-2t)\cdot\log(q)$ and $\delta = 1/4$, and with $\gamma = \gamma' = \eta/2$. For such parameters, this disperser has seed length $\ell = O(\log(n))$ and output length $m = \Omega(n^{1-\eta})$. The statement follows using Corollary 26 with the parameters m, q, d_0, t and with this disperser; the requirement that $d_0 < m$ is satisfied since $d/t \leq \eta'\cdot n^{1-2\eta} < m$, and the requirement that $t \leq \frac{\log(mt/d)}{8\ell}\cdot d$ is satisfied since $\log(mt/d) = \Omega(\log(n))$, relying on the hypothesis that $t/d \leq n^{1-2\eta}$. ■

The second lower bound holds only over fields of constant size. Recall that this lower bound is of the stronger form $\Omega((d/t)\cdot\log(nt/d))$ (as in Corollary 27), and holds even for high degrees up to $\Omega(n)$, and for every $t \lessapprox \sqrt{d}$. More accurately:

Theorem 34 (a lower bound using the local-expander disperser) *For every constant prime power q there exists a constant $\alpha_q > 0$ such that the following holds. Let $n, d, t \in \mathbb{N}$ such that*

[16] Additional lower bounds for this setting, which admit different trade-offs between the lower bound itself and the requirements on d/t, can be proved by instantiating Corollary 26 with other dispersers (e.g., with the naive disperser or with the subspace sampler). For simplicity, we omit these statements.

$2 \cdot (q-1) \leq d \leq n/2^{2(q-1)}$ *and* $t \leq \alpha_q \cdot \sqrt{\log(nt/d)} \cdot \sqrt{d}$. *Then, the seed length of any HSG for* $\mathcal{P}_{n,q,d,\sqrt{2}\cdot q^{-t}}$ *is at least* $\Omega\left(\frac{d}{t} \cdot \log\left(\frac{n \cdot t}{d}\right)\right)$.

To prove Theorem 34, we will instantiate Corollary 26 with linear dispersers that can be obtained from the recent construction of linear 1-local expanders over a constant-sized alphabet by Goldreich [Gol16], following Viola and Wigderson [VW17]. Let us first recall the definition of linear 1-local functions and Goldreich's result:

Definition 35 (linear local functions) *We say that a function* $f \colon \mathbb{F}_q^n \to \mathbb{F}_q^m$ *is* linear 1-local *if each output bit of* f *is an* $\mathbb{F}_q$*-linear function of a single input bit of* f.

Note that the composition of linear 1-local functions is linear 1-local. Then, Goldreich [Gol16], proved that there exist expanders over $\mathbb{F}_q^n$ whose neighbor functions are 1-local $\mathbb{F}_q$-linear functions. Specifically:

Theorem 36 (local expanders [Gol16]) *Let* $\mathbb{F}_q$ *be a field of constant size. Then, for any sufficiently large* $n \in \mathbb{N}$ *there exists an expander (i.e., a graph with a constant spectral gap)* $G = ([q^n], E)$ *of degree* $\Delta = O_q(1)$ *that satisfies the following. For each* $i \in [\Delta]$, *the* i^{th} *neighbor function* $f_i \colon [q^n] \to [q^n]$ *of the graph is a linear* 1*-local function.*

We now use a standard transformation of expanders to extractors: The input to the extractor is a name of a vertex, the seed specifies the directions in a walk of suitable length, and the output is the final vertex in the corresponding walk (that starts from the input vertex and proceeds according to the seed). The crucial point is that for every fixed seed, the output of the extractor is obtained by applying fixed neighbor functions (which correspond to the walk specified in the seed) to the input; in particular, since the neighbor functions are linear, the resulting disperser is also linear.

Theorem 37 (expanders yield good extractors; see, e.g., Theorem 6.22 in [Vad12]) *For any* $q, n \in \mathbb{N}$, *let* $G = ([q^n], E)$ *be an expander (i.e., a graph with a constant spectral gap) of degree* $\Delta = O(1)$. *For* $k < n \cdot \log(q)$ *and* $\delta > 0$, *let* $\mathsf{Disp} \colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^n$, *where* $\ell = r \cdot \log(\Delta)$ *and* $r = O(n \cdot \log(q) - k + \log(1/\delta))$, *be defined as follows. For every* $x \in \mathbb{F}_q^m$ *and* $w \in \{0,1\}^\ell$, *consider the* r*-long walk on* G *that starts from* x, *and let* $\mathsf{Disp}(x, w)$ *be the final vertex in this walk. Then,* Disp *is a* (k, δ)*-disperser.*

Theorem 38 (a linear disperser from a local expander) *Let* $\mathbb{F}_q$ *be a field of constant size, let* $n \in \mathbb{N}$ *be sufficiently large, and for* $a, t \in \mathbb{N}$ *let* $k = (n - 2t) \cdot \log(q)$ *and* $\delta = q^{-a}$. *Then, there exists a linear* (k, δ)*-disperser* $\mathsf{Disp} \colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^n$, *where* $\ell = O_q(t + a)$. *Moreover, the function that maps* x *to* $\{\mathsf{Disp}(x, w)\}_{w \in \{0,1\}^\ell}$ *is linear* 1*-local.*

Proof: We use the disperser from Theorem 37, instantiated with the expander from Theorem 36, and with error parameter $\delta = q^{-a}$ and with $k = (n - 2t) \cdot \log(q)$.

To show that the mapping $x \mapsto \{\mathsf{Disp}(x, w)\}_{w \in \{0,1\}^\ell}$ is linear 1-local, fix any $w \in [2^\ell]$, and let us focus on the w^{th} output element of Disp. Recall that the w^{th} output element is the final vertex in a walk of length r that starts at the input $x \in \mathbb{F}_q^n$ to Disp and whose steps are described by w. In particular, let $f_1, ..., f_\Delta$ be the neighbor functions of G, and let $(i_1, ..., i_r) \in [\Delta]^r$ be the r steps taken in the fixed walk w; then, $\mathsf{Disp}(x)_w = f_{i_r}(f_{i_{r-1}}(...(f_{i_1}(x))...))$. Since each of the neighbor functions is a linear 1-local function, their composition is also linear 1-local. Hence, for every $w \in \{0,1\}^\ell$ it holds that $\mathsf{Disp}(\cdot)_w$ is a linear 1-local function. ■

We now prove Theorem 34 by instantiating Corollary 26 with the linear disperser from Theorem 38:

Proof of Theorem 34: Let $d_0 = d/4t$, and let $a = d_0/(q-1)$ such that $\delta = \delta_{RM}(d_0, q) \geq q^{-a}$. When instantiating the disperser from Theorem 38 with parameters n and $k = (n-2t) \cdot \log(q)$ and $\delta = q^{-a}$, it has seed length $\ell = O_q(t+a)$. Relying on Corollary 26, we get a lower bound of $\Omega\left((d/t) \cdot \log(nt/d)\right)$, assuming that $t \leq \frac{\log(nt/d)}{8\ell} \cdot d$. Thus, we just need to verify the latter condition.

Note that $t \leq \frac{\log(nt/d)}{8\ell} \cdot d$ if and only if $t \cdot (t+a) \leq c_q \cdot \log(nt/d) \cdot d$, where c_q is a constant that depends only on q. Since $t \cdot (t+a) = t^2 + d/4(q-1)$, it suffices to prove that

$$t^2 + d/4(q-1) \leq c_q \cdot \log(nt/d) \cdot d \iff$$
$$t \leq \sqrt{c_q} \cdot \sqrt{(\log(nt/d) - 1/4(q-1))} \cdot \sqrt{d} \,.$$

Finally, since $d \leq n/2^{2(q-1)}$ we have that $\log(nt/d) - 1/4(q-1) \geq \frac{1}{2} \cdot \log(nt/d)$. Hence, it suffices that $t \leq (\sqrt{c_q/2}) \cdot \sqrt{\log(nt/d)} \cdot \sqrt{d}$, which holds due to our hypothesis (using $\alpha_q = \sqrt{c_q/2}$). ■

7 Small sets with a large degree-d closure

In this section we establish a connection between the study of HSGs for polynomials that vanish rarely, and the study of small sets with large degree-d closures, which was recently initiated by Nie and Wang [NW15]. To do so let us first define the degree-d closure of a set $S \subseteq \mathbb{F}^n$:

Definition 39 (degree-d closure) *Let $\mathbb{F}$ be a finite field, and let $n, d \in \mathbb{N}$. Then, for any $S \subseteq \mathbb{F}^n$, we define the* degree-d closure of S, *denoted $\mathtt{Cl}^{(d)}(S)$, by $\mathtt{Cl}^{(d)} = \{x \in \mathbb{F}^n : \forall p \in \mathcal{I}(S), p(x) = 0\}$, where $\mathcal{I}(S) = \{p : \mathbb{F}^n \to \mathbb{F} : \deg(p) = d \wedge \forall s \in S, p(s) = 0\}$.*

We now prove Theorem 5, which shows two reductions. Loosely speaking, we show that any set with degree-d closure of size q^{n-t} is a hitting-set for polynomials that vanish with probability at most q^{-t}; and we show that any hitting-set for polynomials that vanish with probability at most q^{-t} has degree-d' closure of size $q^{n-t}/2$, for d' that is not much smaller than d.

Theorem 40 (small sets with large closures are equivalent to hitting-sets for polynomials that vanish rarely; Theorem 5, restated) *Let $\mathbb{F}$ be a field of size q, let $n \in \mathbb{N}$ and $t < d < n$, and let $S \subseteq \mathbb{F}^n$. Then,* *1. If $\left|\texttt{Cl}^{(\texttt{d})}(S)\right| > q^{n-t}$, then S is a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$.*

2. If S is a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$, then $\left|\texttt{Cl}^{(\texttt{d/2(t+1)})}(S)\right| > \frac{1}{2} \cdot q^{n-t}$.

Proof: For the first statement, let $S \subseteq \mathbb{F}^n$ be such that $\left|\texttt{Cl}^{(\texttt{d})}(S)\right| > q^{n-t}$. Then, every degree-$d$ polynomial that vanishes on S also vanishes on more than q^{n-t} of the inputs. It follows that S is a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$.

For the second statement, for $d' = d/2(t+1)$, assuming that $\left|\texttt{Cl}^{(\texttt{d}')}(S)\right| \leq \frac{1}{2} \cdot q^{n-t}$, we construct a degree-$d$ polynomial that vanishes on S and that vanishes on at most q^{n-t} inputs in $\mathbb{F}^n$ (and it follows that S is not a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$).

To construct the polynomial, let $T_1 = \mathbb{F}^n \setminus \texttt{Cl}^{(\texttt{d}')}(S)$. Note that for every $x \in T_1$ there exists a degree-d' polynomial p_x that vanishes on S, but does not vanish at x. We can thus construct a collection $\mathcal{P}_1$ of degree-d' polynomials such that for every $x \in T_1$ there exists a corresponding $p_x \in \mathcal{P}_1$ satisfying $p_x(x) \neq 0$. (Indeed, a single polynomial might "cover" two distinct inputs, i.e. $p_x = p_y$ for $x \neq y$.)

Now, consider the distribution $\mathbf{p}_1$ over polynomials $\mathbb{F}^n \to \mathbb{F}$ that is defined by

$$\mathbf{p}_1(z) = \sum_{x \in T_1} \mathbf{c}_x \cdot p_x(z) ,$$

where the coefficients $\mathbf{c}_x$ are uniformly and independently chosen in $\mathbb{F}$. Note that $\mathbf{p}_1$ is supported by polynomials of degree d' that vanish on S. Also note that for any fixed $z \in T_1$ we have that

$$\begin{aligned}\Pr[\mathbf{p}_1(z) = 0] &= \Pr\left[\sum_{x \in T_1} \mathbf{c}_x \cdot p_x(z) = 0\right] \\ &= \mathbb{E}_{\{c_x\}_{x \in T_1 \setminus \{z\}}}\left[\Pr\left[\mathbf{c}_z \cdot p_z(z) = -\sum_{x \in T_1 \setminus \{z\}} c_x \cdot p_x(z)\right]\right] ,\end{aligned}$$

which equals $1/q$ since $p_z(z) \neq 0$. Therefore, there exists a fixed polynomial p of degree d' that vanishes on S and on at most $1/q$ of the inputs in T_1.

We now repeat this step t additional times, while maintaining the invariant that for every $x \in T_i$ there exists a polynomial $p_x \in \mathcal{P}_i$ such that $p_x(x) \neq 0$. Specifically, for $i = 2, ..., t+1$, we let $T_i = T_{i-1} \cap \{x \in T_i : p_{i-1}(x) = 0\}$ and $\mathcal{P}_i = \mathcal{P}_{i-1} \setminus \{p_{i-1}\}$. Note that $|T_i| \leq |T_{i-1}|/q$, and that for every $x \in T_i$ there exists $p_x \in \mathcal{P}_i$ such that $p_x(x) \neq 0$. We again define a distribution $\mathbf{p}_i(z) = \sum_{x \in T_i} \mathbf{c}_x \cdot p_x(z)$, and using the same argument as above, we deduce that there exists a fixed polynomial p_i of degree d' that vanishes on S and on at most $1/q$ of the inputs in T_i.

After $t+1$ steps we obtain $t+1$ polynomials $p_1, ..., p_{t+1}$ of degree d' that vanish on S such that $\left|\{x \notin \texttt{Cl}^{(\texttt{d})}(S) : \forall i \in [t], p_i(x) = 0\}\right| \leq |T_1|/q^{t+1} \leq \frac{1}{2} \cdot q^{-t}$. Let $p : \mathbb{F}^n \to \mathbb{F}$ be the multivalued OR of $p_1, ..., p_{t+1}$, defined by $p(x) = \texttt{mvOR}(p_1(x), ..., p_t(x))$.

Note that $\deg(p) < 2(t+1) \cdot d' = d$, and that p vanishes on S. Thus, denoting $\delta = \left|\mathtt{Cl}^{(\mathtt{d}')}(S)\right| / q^n \leq \frac{1}{2} \cdot q^{-t}$, we have that

$$\Pr_{x \in \mathbb{F}^n}[p(x) = 0] = \delta + (1-\delta) \cdot q^{-(t+1)} < q^{-t} ,$$

which implies that $p \in \mathcal{P}_{n,q,d,q^{-t}}$. Hence, S is not a hitting-set for $\mathcal{P}_{n,q,d,q^{-t}}$. ■

As mentioned in Section 1.3, we can obtain an upper-bound on the size of $\mathtt{Cl}^{(\mathtt{d})}(S)$ for any sufficiently-small set S, by combining Theorem 28 and the first item of Theorem 40. Specifically, we can deduce that for every $2 \leq q \leq \text{poly}(n)$ and $d \leq n^{.49}$ and $t \leq \gamma \cdot d$ (where $\gamma > 0$ is a sufficiently small constant), any set S of size $|S| \leq n^{\gamma \cdot (d/t)}$ satisfies $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| \leq q^{n-t}$. However, this corollary is superseded by the upper-bound of [NW15], who showed that for any $S \subseteq \mathbb{F}^n$ it holds that $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| \leq \frac{|S|}{\binom{n+d}{d}} \cdot q^n$.

Indeed, since the problem of constructing small sets with large degree-d closures is at least as hard as the problem of constructing HSGs for polynomials that vanish rarely (due to the first item of Theorem 40), it might be inherent that a direct lower bound on the former problem is stronger than a lower bound that is obtained via a reduction from the latter problem.

Acknowledgements

We are grateful to an exceptionally helpful anonymous reviewer, who pointed us to the work of Nie and Wang [NW15] and to follow-up works, suggested the alternative proof strategy that appears in Appendix B, and helped improve our initial results.

This work was partially done while Dean Doron was at Tel-Aviv University (supported by the Israel science Foundation grants no. 994/14 and 952/18 and by Len Blavatnik and the Blavatnik Family foundation) and at the University of Texas at Austin (supported by NSF Grant CCF-1705028). Amnon Ta-Shma is supported by ISF grant 18/952 and by Len Blavatnik and the Blavatnik Family foundation. Roei Tell is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

References

[ABN+92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. M. Roth. "Construction of Asymptotically Good Low-rate Error-correcting Codes Through Pseudo-random Graphs". In: *IEEE Transactions on Information Theory* 38.2 (1992), pp. 509–516.

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. "Reed-Muller codes for random erasures and errors". In: *IEEE Transactions on Information Theory* 61.10 (2015), pp. 5229–5252.

[BBG16] Arnab Bhattacharyya, Abhishek Bhowmick, and Chetan Gupta. "On higher-order Fourier analysis over non-prime fields". In: *Proc. 20th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2016, Art. No. 23, 29.

[BD18] Peter Beelen and Mrinmoy Datta. "Generalized Hamming weights of affine Cartesian codes". In: *Finite Fields and their Applications* 51 (2018), pp. 130–145.

[BGY18] Paul Beame, Shayan Oveis Gharan, and Xin Yang. *On the Bias of Reed-Muller Codes over Odd Prime Fields*. 2018. eprint: arXiv:1806.06973.

[Bha14] Arnab Bhattacharyya. "Polynomial decompositions in polynomial time". In: *Proc. 22nd European Symposia on Algorithms*. 2014, pp. 125–136.

[BHL12] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. "Random low-degree polynomials are hard to approximate". In: *Computational Complexity* 21.1 (2012), pp. 63–81.

[BHS08] Markus Bläser, Moritz Hardt, and David Steurer. "Asymptotically Optimal Hitting Sets Against Polynomials". In: *Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part I*. Proc. 35th International Colloquium on Automata, Languages and Programming (ICALP). 2008, pp. 345–356.

[BHT15] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. "Algorithmic regularity for polynomials and applications". In: *Proc. 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2015, pp. 1870–1889.

[Bog05] Andrej Bogdanov. "Pseudorandom generators for low degree polynomials". In: *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC)*. 2005, pp. 21–30.

[BS69] E. R. Berlekamp and N. J. A. Sloane. "Restrictions on weight distribution of Reed-Muller codes". In: *Information and Control* 14 (1969), pp. 442–456.

[BV10] Andrej Bogdanov and Emanuele Viola. "Pseudorandom bits for polynomials". In: *SIAM Journal of Computing* 39.6 (2010), pp. 2464–2486.

[CLO15] David A. Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms*. Fourth. Undergraduate Texts in Mathematics. Springer, Cham, 2015.

[CT19] Lijie Chen and Roei Tell. "Bootstrapping results for threshold circuits "just beyond" known lower bounds". In: *Proc. 51st Annual ACM Symposium on Theory of Computing (STOC)*. 2019, pp. 34–41.

[CTS13] Gil Cohen and Amnon Ta-Shma. "Pseudorandom Generators for Low Degree Polynomials from Algebraic Geometry Codes". In: *Electronic Colloquium on Computational Complexity: ECCC* 20 (2013), p. 155.

[DMO+19] Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. "Nearly Optimal Pseudorandomness From Hardness". In: *Electronic Colloquium on Computational Complexity: ECCC* 26 (2019), p. 99.

[Dvi09] Zeev Dvir. "On the size of Kakeya sets in finite fields". In: *Journal of the American Mathematical Society* 22.4 (2009), pp. 1093–1097.

[Gol16] Oded Goldreich. "Deconstructing 1-local expanders". In: *Electronic Colloquium on Computational Complexity: ECCC* 23 (2016), p. 152.

[GRS19] Venkatesan Guruswami, Atri Rudra1, and Madhu Sudan. *Essential Coding Theory*. Accessed at `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`. 2019.

[GS01] Venkatesan Guruswami and Madhu Sudan. "Extensions to the Johnson Bound". Manuscript. 2001.

[GT09] Ben Green and Terence Tao. "The distribution of polynomials over finite fields, with applications to the Gowers norms". In: *Contributions to Discrete Mathematics* 4.2 (2009), pp. 1–36.

[GW14] Oded Goldreich and Avi Widgerson. "On derandomizing algorithms that err extremely rarely". In: *Proc. 46th Annual ACM Symposium on Theory of Computing (STOC)*. Full version available online at *Electronic Colloquium on Computational Complexity: ECCC*, 20:152 (Rev. 2), 2013. 2014, pp. 109–118.

[HS10] Elad Haramaty and Amir Shpilka. "On the structure of cubic and quartic polynomials". In: *Proc. 42nd Annual ACM Symposium on Theory of Computing (STOC)*. 2010, pp. 331–340.

[KL08] Tali Kaufman and Shachar Lovett. "Worst Case to Average Case Reductions for Polynomials". In: *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2008, pp. 166–175.

[KLP12] Tali Kaufman, Shachar Lovett, and Ely Porat. "Weight distribution and list-decoding size of Reed-Muller codes". In: *IEEE Transactions on Information Theory* 58.5 (2012), pp. 2689–2696.

[KS01] Adam R. Klivans and Daniel Spielman. "Randomness efficient identity testing of multivariate polynomials". In: *Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC)*. 2001, pp. 216–223.

[KT70] Tadao Kasami and Nobuki Tokura. "On the weight structure of Reed-Muller codes". In: *IEEE Transactions on Information Theory* IT-16 (1970), pp. 752–759.

[Lov09] Shachar Lovett. "Unconditional pseudorandom generators for low-degree polynomials". In: *Theory of Computing* 5 (2009), pp. 69–82.

[Lu12] Chi-Jen Lu. "Hitting set generators for sparse polynomials over any finite fields". In: *Proc. 27th Annual IEEE Conference on Computational Complexity (CCC)*. 2012, pp. 280–286.

[LV98] Daniel Lewin and Salil Vadhan. "Checking polynomial identities over any field: towards a derandomization?" In: *Proc. 30th Annual ACM Symposium on Theory of Computing (STOC)*. 1998, pp. 438–447.

[LVW93] M. Luby, B. Velickovic, and A. Wigderson. "Deterministic approximate counting of depth-2 circuits". In: *Proc. 2nd Israel Symposium on Theory and Computing Systems*. 1993, pp. 18–24.

[McE69] R. J. McEliece. "Quadratic forms over finite fields and second- order Reed-Muller codes". In: *Space Program Summary* 3.37–58 (1969), pp. 28–33.

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[NN93] Joseph Naor and Moni Naor. "Small-bias probability spaces: efficient constructions and applications". In: *SIAM Journal of Computing* 22.4 (1993), pp. 838–856.

[NW15] Zipei Nie and Anthony Y. Wang. "Hilbert functions and the finite degree Zariski closure in finite field combinatorial geometry". In: *Journal of Combinatorial Theory. Series A* 134 (2015), pp. 196–220.

[NZ96] Noam Nisan and David Zuckerman. "Randomness is Linear in Space". In: *Journal of Computer and System Sciences* 52.1 (1996), pp. 43–52.

[SB70] Neil J. A. Sloane and Elwyn R. Berlekamp. "Weight enumerator for second-order Reed-Muller codes". In: *IEEE Transactions on Information Theory* IT-16 (1970), pp. 745–751.

[ST18] Rocco A. Servedio and Li-Yang Tan. "Luby-Veličković-Wigderson revisited: improved correlation bounds and pseudorandom generators for depth-two circuits". In: *Proc. 22nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. Vol. 116. 2018, Art. No. 56, 20.

[SU05] Ronen Shaltiel and Christopher Umans. "Simple extractors for all min-entropies and a new pseudorandom generator". In: *Journal of the ACM* 52.2 (2005), pp. 172–216.

[Tel19] Roei Tell. "Improved Bounds for Quantified Derandomization of Constant-Depth Circuits and Polynomials". In: *Computational Complexity*. 2019.

[TSZ04] Amnon Ta-Shma and David Zuckerman. "Extractor codes". In: *IEEE Transactions on Information Theory* 50.12 (2004), pp. 3015–3025.

[TSZS06] Amnon Ta-Shma, David Zuckerman, and Shmuel Safra. "Extractors from Reed-Muller codes". In: *Journal of Computer and System Sciences* 72.5 (2006), pp. 786–812.

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2012.

[Vio09a] Emanuele Viola. "Guest Column: correlation bounds for polynomials over 0 1." In: *SIGACT News* 40 (Feb. 2009), pp. 27–44.

[Vio09b] Emanuele Viola. "The sum of d small-bias generators fools polynomials of degree d". In: *Computational Complexity* 18.2 (2009), pp. 209–217.

[VW17] Emanuele Viola and Avi Wigderson. "Local Expanders". In: *Computational Complexity* (2017).

[War35] Ewald Warning. "Bemerkung zur vorstehenden Arbeit von Herrn Chevalley". In: *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* 11 (1935), pp. 76–83.

[Yao82] Andrew C. Yao. "Theory and Application of Trapdoor Functions". In: *Proc. 23rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1982, pp. 80–91.

Appendix A Next-element unpredictability over large alphabets

Recall that, as proved by Yao [Yao82], if a distribution $\mathbf{w}$ over $\{0,1\}^m$ is next-bit unpredictable, then $\mathbf{w}$ is close to the uniform distribution. In this appendix we prove a generalized version of this claim that applies also to distributions over Σ^m where Σ is an alphabet of arbitrary size.

Proposition 41 (next-element unpredictability implies closeness to uniform, over arbitrary alphabets) *Let Σ be a set of size $q = |\Sigma|$, let $\mathbf{w}$ be a distribution over Σ^m, and assume that the statistical distance between $\mathbf{w}$ and the uniform distribution on Σ^m, denoted $\mathbf{u}_m$, is at least $\rho > 0$. Then, there exists $i \in [m]$ and a function $P : \Sigma^{i-1} \to \Sigma$ such that* $\Pr[\mathbf{w}_i = P(\mathbf{w}_1, ..., \mathbf{w}_{i-1})] > 1/q + \rho/qm$,.

Proof: Let $\mathbf{h}^{(0)} = \mathbf{u}_n$, and for $i \in [m]$ let $\mathbf{h}^{(i)}$ be the distribution over Σ^m such that its first i elements are sampled from $\mathbf{w}$ and its last $m - i$ elements are sampled uniformly and independently. By a standard hybrid argument, for some $i \in [m]$ it holds that the statistical distance between $\mathbf{h}^{(i-1)}$ and $\mathbf{h}^{(i)}$ is at least ρ/m. Hence, there exists $T : \Sigma^i \to \{0,1\}$ such that

$$\Pr[T(\mathbf{h}^{(i)}_{1,...i}) = 1] - \Pr[T(\mathbf{h}^{(i-1)}_{1,...,i}) = 1] > \rho/m \ .$$

Now, for any $w_1, ..., w_{i-1} \in \Sigma^{i-1}$, let

$$P(w_1, ..., w_{i-1}) = \text{argmax}_{z\in\Sigma} \left\{\Pr\left[\mathbf{w}_i = z | \mathbf{w}_{1,...,i-1} = w_{1,...,i-1}\right]\right\} \ .$$

Denote $\Pr_{w\sim\mathbf{w}}[w_i = P(w_{1,...,i-1})] \overset{\text{def}}{=\!=} (1/q + \delta)$, where $\delta \in [0,1]$. Our goal is to prove that $\delta > \rho/qm$. By the definition of P, for every $z \in \Sigma$ and $w_{1,...,i-1} \in \Sigma^{i-1}$ we have that

$$\mathbb{E}_{w\sim\mathbf{w}}\left[\Pr[\mathbf{w}_i = z | \mathbf{w}_{1,...,i-1} = w_{1,...,i-1}]\right] \leq 1/q + \delta \ .$$

Thus, we have that

$$\begin{aligned}
&\Pr[T(\mathbf{h}^{(i)}_{1,...,i}) = 1] - \Pr[T(\mathbf{h}^{(i-1)}_{1,...,i}) = 1] \\
&\quad = \mathbb{E}_{u_{i+1,...,n}\sim\mathbf{u}_n, w_{1,...,i-1}\sim\mathbf{w}}\Big[\sum_{z\in\Sigma}\Pr[\mathbf{w}_i = z|\mathbf{w}_{1,...,i-1} = w_{1,...,i-1}]\cdot T(w_1,...,w_{i-1},z,u_{i+1},...,u_n) \\
&\qquad\qquad - \frac{1}{q}\cdot\sum_{z\in\Sigma}T(w_1,...,w_{i-1},z,u_{i+1},...,u_n)\Big] \\
&\quad \leq \mathbb{E}_{u_{i+1,...,n}\sim\mathbf{u}_n, w_{1,...,i-1}\sim\mathbf{w}}\Big[\sum_{z\in\Sigma}(1/q+\delta)\cdot T(w_1,...,w_{i-1},z,u_{i+1},...,u_n) \\
&\qquad\qquad - \frac{1}{q}\cdot T(w_1,...,w_{i-1},z,u_{i+1},...,u_n)\Big] \\
&\quad \leq q\cdot\delta\,,
\end{aligned}$$

which implies that $\delta > \rho/qm$. ■

Appendix B An alternative argument for lower bounds

In this appendix we describe an alternative argument for proving a lower bound on the size of hitting-sets for polynomials that vanish rarely; this argument was suggested to us by an anonymous reviewer. We note in advance that this argument is known to work only for prime fields (for reasons that will be explained below), and that our main reason for presenting it is since it is simple and elegant. For simplicity, we first present the argument only for the field $\mathbb{F}_2$.

Recall, from the "warm-up" in Section 2.1, that a lower bound on the seed length of any hitting-set generator for $\mathcal{P}_{n,q,d,t}$ can be proved quite easily (i.e., with the naive disperser and without "randomized tests") when the corresponding Reed-Muller code has constant relative distance. The main technical ingredient underlying the alternative argument is the existence of a *large subcode* of the Reed-Muller code that has constant relative distance; the existence of such a subcode can be deduced using the following lemma by Ben-Eliezer, Hod, and Lovett [BHL12]. Towards stating the lemma, we define the bias of a function $f\colon \mathbb{F}_2^n \to \mathbb{F}_2$ to be $\mathrm{bias}(f) = \Pr_{x\in\mathbb{F}_2^n}[f(x) = 0] - \Pr_{x\in\mathbb{F}_2^n}[f(x) = 1] = 2\Pr_x[f(x) = 0] - 1$. The following is showed in [BHL12]:

Lemma 42 (a random $\mathbb{F}_2$-polynomial is unbiased [BHL12, Lemma 2]) *For every constant $\varepsilon > 0$ there exist constants $\alpha, \beta > 0$ such that the following holds. For $n \in \mathbb{N}$ and $d \leq (1-\varepsilon)\cdot n$, let $\mathbf{p}$ be a uniformly-chosen degree-d polynomial in $\mathbb{F}_2^n \to \mathbb{F}_2$. Then, it holds that*

$$\Pr\left[\left|\mathrm{bias}(\mathbf{p})\right| > 2^{-\alpha\cdot(n/d)}\right] \leq 2^{-\beta\cdot\binom{n+d}{d}}.$$

Loosely speaking, the fact that a random degree-d polynomial is unbiased implies that the difference between two random degree-d polynomials is unbiased, or in other words that two random degree-d polynomials disagree on $\approx 1/2$ of their inputs. Hence, when independently choosing many random degree-d polynomials, with high probability the subcode spanned by them has distance close to $1/2$; that is, there exists a large subcode of the Reed-Muller code with relative distance close to $1/2$. In more detail:

Corollary 43 (a large subcode of the Reed-Muller code with constant relative distance) *For every $\varepsilon > 0$ there exists $\gamma > 0$ such that the following holds. For every sufficiently large $n \in \mathbb{N}$ and $d \leq n^{.99}$ there exists a linear subcode of the $[n, d]$ Reed-Muller code over $\mathbb{F}_2$ that has dimension at least $\gamma \cdot \binom{n+d}{d}$ and relative distance at least $1/2 - \varepsilon$.*

Proof: Fix $\varepsilon > 0$, and let $\gamma > 0$ be sufficiently small. For two polynomials $f_1, f_2 \colon \mathbb{F}_2^n \to \mathbb{F}_2$, let $\mathtt{agr}(f_1, f_2) = \Pr_{x \in \mathbb{F}_2^n}[f_1(x) = f_2(x)] = \Pr_x[f_1(x) - f_2(x) = 0] = \frac{1}{2} + \text{bias}[f_1 - f_2]/2$. Denoting a uniformly-chosen degree-d polynomial $\mathbb{F}_2^n \to \mathbb{F}_2$ by $\mathbf{p}$, we choose $D' = \gamma \cdot \binom{n+d}{d}$ polynomials $\mathbf{b}_1, ..., \mathbf{b}_{D'} \sim \mathbf{p}$, and denote the subcode of the Reed-Muller code spanned by these polynomials by $\mathcal{C}' = \{\mathbf{f}_1, ..., \mathbf{f}_T\}$, for $T \leq 2^{D'}$.

First note that with high probability $T = 2^{D'}$, or in other words the $\mathbf{b}_i$-s are linearly independent. This is the case since if we choose the $\mathbf{b}_i$-s sequentially, then at each iteration $i \in [D']$, the probability that $\mathbf{b}_i$ lies in the subspace spanned by the $i-1$ previously-chosen polynomials is at most $2^{(i-1)-\binom{n+d}{d}} < 2^{D'-\binom{n+d}{d}} = o(1/D')$.

Now, conditioned on the event that $T = 2^{D'}$, note that for every fixed $i \in [2^{D'}]$ it holds that $\mathbf{f}_i$ is uniformly distributed (i.e., its marginal distribution is $\mathbf{p}$). Thus, for every fixed $i \in [2^{D'}]$ we have that

$$\begin{aligned}
\Pr[\exists j \neq i : \mathtt{agr}(\mathbf{f}_i, \mathbf{f}_j) > 1/2 + \varepsilon] &= \Pr[\exists j \neq i : \text{bias}(\mathbf{f}_i - \mathbf{f}_j) > 2\varepsilon] \\
&= \sum_{p \in \mathtt{supp}(\mathbf{p})} \Pr[\mathbf{f}_i = p] \cdot \Pr[\exists j \neq i : \text{bias}(p - \mathbf{f}_j) > 2\varepsilon] \\
&< 2^{D'} \cdot \Pr[\text{bias}(\mathbf{p}) > 2\varepsilon] \cdot \sum_{p \in \mathtt{supp}(\mathbf{p})} \Pr[\mathbf{f}_i = p] \\
&\leq 2^{(\gamma - \beta) \cdot \binom{n+d}{d}}, \qquad \text{(Lemma 42)}
\end{aligned}$$

where $\beta = \beta(\varepsilon)$ and we used the fact that $2^{-\alpha \cdot (n/d)} \leq 2\varepsilon'$ for every constant $\alpha = \alpha(\varepsilon)$ and large enough n. Taking $\gamma < \beta$ to be a sufficiently small constnat, the above is $o(1)$. Therefore, with high probability over choice of $\mathbf{b}_1, ..., \mathbf{b}_{D'}$, the linear subcode induced by our choice has dimension D' and relative distance at least $1/2 - \varepsilon'$. ■

We now prove the lower bound. Loosely speaking, using the simple reduction that was described in the "warm-up" in Section 2.1, we show that if there exists a small hitting-set for degree-d polynomials $\mathbb{F}_2^n \to \mathbb{F}_2$ that vanish rarely then there exists a small hitting-set for the large subcode from Corollary 43.

Theorem 44 (a lower bound using subcodes of the Reed-Muller code) *There exists a universal constant $\alpha > 0$ such that the following holds. For any sufficiently large $n \in \mathbb{N}$, $d \leq n^{.99}$, and $t \in \mathbb{N}$ such that $t < \alpha \cdot d$, the seed length of any hitting-set generator for $\mathcal{P}_{n,2,d,2^{-t}}$ is at least $\Omega((d/t) \cdot \log(n))$.*

Proof: Let $\delta > 0$ be a sufficiently small constant, let $n \in \mathbb{N}$ be sufficiently large, let $d \leq (1-\varepsilon) \cdot n$, and let $t < d$. Assume towards a contradiction that there exists a hitting-set $S \subseteq \mathbb{F}_2^n$ for $\mathcal{P}_{n,2,d,2^{-t}}$ of size $\delta \cdot \binom{n+d/t}{d/t}$.

Now, let $d_0 = \lfloor d/4t \rfloor$, let $t' = 2t$, and let $m = \lfloor n/t' \rfloor$. Let $\mathcal{C}' \subseteq \mathbb{F}_2^m$ be a linear subcode of the Reed-Muller code $\mathbb{F}_2^m \to \mathbb{F}_2$ of degree d_0 that has dimension $\gamma \cdot \binom{m+d_0}{d_0}$ and relative distance at least .49, whose existence is guaranteed by Corollary 43.

We construct a hitting-set for $\mathcal{C}'$ as follows. For every polynomial $p \in \mathcal{C}'$, consider the polynomial $p' \colon \mathbb{F}_2^{m \cdot t'} \to \mathbb{F}_2$ such that $p'(z) = \texttt{mvOR}(p(z^{(1)}), ..., p(z^{(t')}))$, where we think of $z = z^{(1)}, ..., z^{(t')}$ such that for each i it holds that $z^{(i)}$ is an m-bit string. Note that the degree of p' is less than $4d_0 \cdot t \leq d$, and that $\Pr_{z \in \mathbb{F}_2^{t' \cdot m}}[p'(z) = 0] = \Pr_{x \in \mathbb{F}_2^m}[p(x) = 0]^{t'} \leq 2^{-t}$. By our assumption that S is a hitting-set for $\mathcal{P}_{n,2,d,2^{-t}}$, there exists $z \in S$ such that $p'(z) \neq 0$, which implies that for some $i \in [t']$ it holds that $p(z^{(i)}) \neq 0$.

Thus, the set $S_0 = \{z^{(i)} : z \in S, i \in [t']\}$ is a hitting-set for $\mathcal{C}'$ of size at most $2t \cdot |S|$. Now, relying on Fact 14, any hitting-set for $\mathcal{C}'$ is of size at least $\dim(\mathcal{C}') = \gamma \cdot \binom{m+d_0}{d_0}$, and hence $|S| \geq \gamma \cdot \binom{m+d_0}{d_0}/t$. The seed length for sampling from S is thus at least

$$\Omega\left(d_0 \cdot \log((m+d_0)/d_0) - \log(t)\right) = \Omega\left((d/t) \cdot \log(n/d) - \log(t)\right) ,$$

which simplifies to $\Omega((d/t) \cdot \log(n))$ relying on the hypotheses that $d \leq n^{.99}$ and $t \leq \alpha \cdot d$, for a sufficiently small universal constant $\alpha > 0$. ■

To generalize the foregoing argument to fields other than $\mathbb{F}_2$, note that the only place where we used the fact that the field is $\mathbb{F}_2$ is when deducing the existence of a large subcode of the Reed-Muller code (i.e., in Corollary 43, which relied on Lemma 42). The argument can be generalized to any prime field, relying on a generalization of Lemma 42 to arbitrary prime fields that was recently proved by Beame, Gharan, and Yang [BGY18]. However, we are not aware of an analogous result for non-prime fields.