On Hitting-Set Generators for Polynomials that Vanish Rarely

Dean Doron
Department of Computer Science, Stanford University, USA
ddoron@stanford.edu

Amnon Ta-Shma
The Blavatnik School of Computer Science, Tel-Aviv University, Israel
amnon@tau.ac.il

Roei Tell
Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Israel
roei.tell@weizmann.ac.il

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 $p\colon \mathbb{F}^n \to \mathbb{F}$ 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.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Hitting-set generators, Polynomials over finite fields, Quantified derandomization

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.7

Category RANDOM

Related Version A full version of the paper is available at `https://eccc.weizmann.ac.il/report/2019/119/`.

Funding *Dean Doron*: Supported by a Motwani Postdoctoral Fellowship and by NSF award CCF-1763311.
Amnon Ta-Shma: Supported by ISF grant 18/952 and by Len Blavatnik and the Blavatnik Family foundation.

Roei Tell: Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

Acknowledgements We are grateful to an exceptionally helpful anonymous reviewer, who pointed us to the work of Nie and Wang [33] and to follow-up works, suggested an alternative proof strategy for the lower bound in the special case of prime fields (the alternative proof for this special case appears in the full version), and helped improve our initial results.

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 Fact 14 and Fact 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., [32, 29, 26, 25, 9, 10, 8, 27, 43, 28, 12, 35], and the related survey by Viola [42].

Several years ago, Goldreich and Wigderson [18, 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 [18] (see also, e.g., [40, 11, 14]). And thirdly, as polynomial-based

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

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 arbitrary 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 [45], 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.

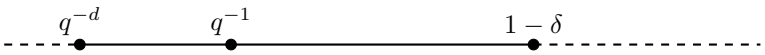

Figure 1 The two extremal values of ε (i.e., $\varepsilon = q^{-d}$ and $\varepsilon = 1-\delta$) and the expected $\varepsilon = q^{-1}$ for a random polynomial. (The parameter δ denotes the relative distance of the corresponding q-ary Reed-Muller code $RM(n,d)$.)

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., [4, 22]), but is currently settled only for $d = 2$ (see [37, 31]). 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 [24, 1]). 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., [30, 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 $p\colon \mathbb{F}^n \to \mathbb{F}$ 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 [19, 23, 21, 5, 7, 6]). 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 [18, 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 [40], 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 [40] 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 – 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

[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 $\Pr_{s \in S}[p(s) \neq 0] \geq \varepsilon$.

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 Section 5.[4]

Now, still referring to the combinatorial question, we observe that a result of Kaufmann, Lovett, and Porat [24], 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 [24]). *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 Theorem 1 and Theorem 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 [18] 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 $t < d - O(1)$:

▶ **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 [18] 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.

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 [33].

[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 $\log(n^{.99} \cdot (t/d))$.

	Seed length	Field Size	ε
Lower bounds			
[40]	$\Omega(d^{\Omega(1)} \cdot \log(n/d^{\Omega(1)}))$	$q = \text{poly}(n)$	$q^{-O(1)}$
Theorem 1	$\Omega((d/t) \cdot \log n)$ $(d \leq n^{.49})$	$2 \leq q \leq \text{poly}(n)$	q^{-t}
Theorem 23	$\Omega((d/t) \cdot \log(n^{.99} \cdot t/d))$ $(d/t \lessapprox q \cdot n^{.01})$	$2 \leq q \leq \text{poly}(n)$	q^{-t}
Upper bounds			
[18]	$O(\log n)$ (explicit)	$q = 2$	$2^{-d+O(1)}$
Theorem 2	$O((d-t)\log(\frac{n}{d-t})$ (non-explicit)	$q = 2$	2^{-t}
Theorem 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.

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:

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

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 [16, 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., $\texttt{Cl}^{(\texttt{d})}(S) = \mathbb{F}^n$) 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 $\left|\texttt{Cl}^{(\texttt{d})}(S)\right| = q^n$.

Loosely speaking, the main result of Nie and Wang [33] extends Observation 4 by showing that that for any $S \subseteq \mathbb{F}^n$ it holds that $\left|\texttt{Cl}^{(\texttt{d})}(S)\right| \leq \frac{|S|}{\binom{n+d}{d}} \cdot |\mathbb{F}|^n$. 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]

[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 [3], 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 $S \subseteq \mathbb{F}^n$).

We take another approach to extending Observation 4, 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 $\left|\mathtt{Cl}^{(\mathtt{d}/2(\mathtt{t}+1))}(S)\right| > \frac{1}{2} \cdot q^{n-t}$.*

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. 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 Appendix C 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 [33]. For more details see Appendix C.

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:

[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 $|T|/|F|^m \geq \delta$, for all but at most 2^k of the inputs $z \in \mathbb{F}^n$ there exists $i \in \{0,1\}^\ell$ such that $\mathsf{Disp}(z,i) \in T$.

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 $\Pr_{z\in\mathbb{F}^n}[\forall i\ \mathsf{Disp}(z,i) \notin T] \leq \varepsilon = 2^k/q^n$. 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 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.* [2].[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 $q = \mathrm{poly}(n)$ 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 $k = (n-t)\cdot\log(q)$.

[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 $1 - 1/q$.

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 Definition 16 and Definition 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 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 5).

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]

In the special case when $\mathbb{F}$ is a *prime field*, 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

[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 $\delta = \Omega(1)$.

dimension linear in $\binom{m+d_0}{d_0}$) that still has distance $\Omega(1)$; see [15, Appendix B] for details. 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 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 $\mathsf{randomized\ test}$ $\mathbf{p}$ 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 Appendix A)**.** *Assume that 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 20). *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 $\mathsf{Disp}\colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$. Then, for $d = 4d_0 \cdot t$, if $\ell \leq \frac{d}{8t}\cdot\log(mt/d)$, 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 21).

Our lower bounds, which include Theorem 1 and various extensions and are presented in Section 5, are proved by instantiating Lemma 7 with specific useful dispersers. In a gist, Theorem 1 and some extensions are proved using a linear disperser that we obtain by modifying the extractor by Shaltiel and Umans [36]; the original extractor works over the binary alphabet, and we modify it to a linear disperser over an arbitrary field $\mathbb{F}_q$ (see Appendix B for details). Another extension of Theorem 1, which applies only to fields of constant size, is proved using a linear disperser that is based on the recent construction of "linear 1-local expanders" by Goldreich [17], following Viola and Wigderson [44].

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 [18], by extending a proof approach from [40]. 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 [43] for low-degree polynomials.

In more detail, we say that a polynomial $p\colon \mathbb{F}_2^n \to \mathbb{F}_2$ 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 $\Pr_h[\mathbf{h}(x) = p(x)] \geq .99$. 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 $\vec{a} = a^{(1)},\ldots,a^{(k)} \in \mathbb{F}_2^n$, where $k = t - O(1)$, and outputting the polynomial $h_{\vec{a}} = \Delta_{a^{(k)}}\Delta_{a^{(k-1)}}\ldots\Delta_{a^{(1)}}(p) + 1$; 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 $\Pr[\mathbf{h}(x) = p(x)] \geq .99$.

[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.

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. 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 [43] for such polynomials, which uses a seed of length $O((d-t)\cdot(2^{d-t}+\log(n)))$, is also a HSG for $\mathcal{P}_{n,2,d,2^{-t}}$.

The proofs of Theorem 2 and Theorem 3 appear in the full version (see [15, Section 5]).

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., [20]). *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 $\mathtt{OR}\colon\mathbb{F}^k\to\mathbb{F}$ 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 $\mathtt{mvOR}\colon\mathbb{F}^k\to\mathbb{F}$ is a* multivalued OR function *if $\mathtt{mvOR}(0^k)=0$, but $\mathtt{mvOR}(x)\neq 0$ for every $x\neq 0^k$.*

For a fixed field $\mathbb{F}$ there are many different k-variate multivalued OR functions. Indeed, the standard $\mathtt{OR}$ 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 [40, Proposition 7.3] for the construction, and see e.g. [13, 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 $H = \{H_n\colon \{0,1\}^{\ell(n)} \to \mathbb{F}_{q(n)}^n\}$ 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 [27, Lemma 2.4]).

Lastly, we recall the standard lower bound on the size of hitting-sets for polynomials of degree d and state the complementary upper-bound that is obtained by a standard probabilistic argument. (For proofs see [15, Section 3].)

▶ **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 $d \cdot \log(n/d)$.*

▶ **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 $O(d \cdot \log(n/d) + \log\log(q))$.*

3.3 Dispersers and Extractors

We recall the definition of dispersers $\mathsf{Disp}\colon [N] \times \{0,1\}^\ell \to [M]$, 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.*[11]

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 $\mathsf{Disp}\colon \mathbb{F}^n \times \{0,1\}^\ell \to \mathbb{F}^m$ 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 5 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[12] 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.*

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 Lower Bounds from Low-Degree Dispersers

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.

[11] In this work we take the *hitter* view of a disperser, which is equivalent to the following standard definition of dispersers: For every random variable $\mathbf{x} \sim \mathbb{F}^n$ with min-entropy k, $\mathsf{Disp}(\mathbf{x}, \mathbf{u}_\ell)$ has support size at least $(1-\delta)q^m$.

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

▶ **Proposition 19** (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 22. 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 22 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 19, 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 20** (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 19 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 21** (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* 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 20, 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. ■

5 Lower Bounds Over General Finite Fields

In this section we describe our lower bounds on the seed length of HSGs for polynomials that vanish rarely, which are proved by instantiating the approach from Section 4 with specific dispersers that are suitable for the corresponding parameter settings.

We prove three incomparable lower bounds. Our first and main lower bound 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 $\Omega(d)$. For details see Theorem 23 in Appendix B.

Our two additional lower bounds, which are detailed and proved in the full version (see [15, Section 6.3]), hold in more specific settings than the foregoing lower bound, but have advantages over this bound. The first of the two lower bounds 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 23, but holds for higher degrees up to $d \leq n^{1-\Omega(1)}$ without complicated conditions on d and t. 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 $t \lessapprox \sqrt{d}$.

[13] Recall, from Corollary 21, that this is the lower is that would be obtained if there exists a linear disperser with optimal parameters.

References

1. Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed-Muller codes for random erasures and errors. *IEEE Transactions on Information Theory*, 61(10):5229–5252, 2015.
2. 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. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.
3. Peter Beelen and Mrinmoy Datta. Generalized Hamming weights of affine Cartesian codes. *Finite Fields and their Applications*, 51:130–145, 2018.
4. E. R. Berlekamp and N. J. A. Sloane. Restrictions on weight distribution of Reed-Muller codes. *Information and Control*, 14:442–456, 1969.
5. Arnab Bhattacharyya. Polynomial decompositions in polynomial time. In *Proc. 22nd European Symposia on Algorithms*, pages 125–136. 2014.
6. 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)*, pages Art. No. 23, 29. 2016.
7. Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proc. 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1870–1889, 2015.
8. 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), pages 345–356, 2008.
9. Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *Proc. 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 21–30. 2005.
10. Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM Journal of Computing*, 39(6):2464–2486, 2010.
11. 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)*, pages 34–41, 2019.
12. Gil Cohen and Amnon Ta-Shma. Pseudorandom generators for low degree polynomials from algebraic geometry codes. *Electronic Colloquium on Computational Complexity: ECCC*, 20:155, 2013.
13. David A. Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, Cham, fourth edition, 2015.
14. Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly optimal pseudorandomness from hardness. *Electronic Colloquium on Computational Complexity: ECCC*, 26:99, 2019.
15. Dean Doron, Amnon Ta-Shma, and Roei Tell. On hitting-set generators for polynomials that vanish rarely (rev. 1). *Electronic Colloquium on Computational Complexity: ECCC*, 26:119, 2019.
16. Zeev Dvir. On the size of Kakeya sets in finite fields. *Journal of the American Mathematical Society*, 22(4):1093–1097, 2009.
17. Oded Goldreich. Deconstructing 1-local expanders. *Electronic Colloquium on Computational Complexity: ECCC*, 23:152, 2016.
18. Oded Goldreich and Avi Widgerson. On derandomizing algorithms that err extremely rarely. In *Proc. 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 109–118. 2014. Full version available online at *Electronic Colloquium on Computational Complexity: ECCC*, 20:152 (Rev. 2), 2013.
19. Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contributions to Discrete Mathematics*, 4(2):1–36, 2009.

694 **20** Venkatesan Guruswami, Atri Rudra1, and Madhu Sudan. Essential coding theory, 2019. Accessed at `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`.

697 **21** Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proc. 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 331–340. 2010.

699 **22** Tadao Kasami and Nobuki Tokura. On the weight structure of Reed-Muller codes. *IEEE Transactions on Information Theory*, IT-16:752–759, 1970.

701 **23** 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)*, pages 166–175, 2008.

704 **24** Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of Reed-Muller codes. *IEEE Transactions on Information Theory*, 58(5):2689–2696, 2012.

706 **25** Adam R. Klivans and Daniel Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 216–223, 2001.

709 **26** 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)*, pages 438–447. 1998.

712 **27** Shachar Lovett. Unconditional pseudorandom generators for low-degree polynomials. *Theory of Computing*, 5:69–82, 2009.

714 **28** Chi-Jen Lu. Hitting set generators for sparse polynomials over any finite fields. In *Proc. 27th Annual IEEE Conference on Computational Complexity (CCC)*, pages 280–286. 2012.

716 **29** M. Luby, B. Velickovic, and A. Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proc. 2nd Israel Symposium on Theory and Computing Systems*, pages 18–24, 1993.

719 **30** F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

721 **31** R. J. McEliece. Quadratic forms over finite fields and second-order Reed-Muller codes. *Space Program Summary*, 3(37–58):28–33, 1969.

723 **32** Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal of Computing*, 22(4):838–856, 1993.

725 **33** Zipei Nie and Anthony Y. Wang. Hilbert functions and the finite degree Zariski closure in finite field combinatorial geometry. *Journal of Combinatorial Theory. Series A*, 134:196–220, 2015.

728 **34** Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

730 **35** 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)*, volume 116, pages Art. No. 56, 20. 2018.

734 **36** Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *Journal of the ACM*, 52(2):172–216, 2005.

736 **37** Neil J. A. Sloane and Elwyn R. Berlekamp. Weight enumerator for second-order Reed-Muller codes. *IEEE Transactions on Information Theory*, IT-16:745–751, 1970.

738 **38** Amnon Ta-Shma and David Zuckerman. Extractor codes. *IEEE Transactions on Information Theory*, 50(12):3015–3025, 2004.

740 **39** Amnon Ta-Shma, David Zuckerman, and Shmuel Safra. Extractors from Reed-Muller codes. *Journal of Computer and System Sciences*, 72(5):786–812, 2006.

742 **40** Roei Tell. Improved bounds for quantified derandomization of constant-depth circuits and polynomials. In *Computational Complexity*, 2019.

744 **41** Salil P. Vadhan. *Pseudorandomness.* Foundations and Trends in Theoretical Computer Science. Now Publishers, 2012.

42 Emanuele Viola. Guest column: correlation bounds for polynomials over $\{0,1\}$. *SIGACT News*, 40:27–44, 02 2009.

43 Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

44 Emanuele Viola and Avi Wigderson. Local expanders. *Computational Complexity*, 2017.

45 Ewald Warning. Bemerkung zur vorstehenden arbeit von herrn chevalley. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 1935.

A Randomized Tests

The proofs of both our upper bounds and of our lower bounds relies on a general observation that we explain here. The observation is essentially from [40, Sections 2.1 & 4], following a proof idea from [10].

Assume that we want to deterministically *find* an element in a set $\mathrm{G} \subseteq \mathbb{F}^n$. 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 [40, 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 [40] for details), and we will apply it in two specific settings. In the first setting, which is useful for our lower bound proofs, the set G is dense (i.e., $\Pr_{x\in\mathbb{F}^n}[x \in \mathrm{G}] \geq .99$), and can be decided by a distribution $\mathbf{p}$ over polynomials with small "one-sided" error (i.e., every $x \in \mathrm{G}$ is accepted with high probability, and every $x \notin \mathrm{G}$ 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 22** (randomized tests). *Let $\varepsilon, \rho > 0$ such that $\varepsilon + \rho < 1$, and let $\mathrm{G} \subseteq \mathbb{F}^n$ be such that $\Pr_{x\in\mathbb{F}^n}[x \in \mathrm{G}] \geq 1 - \varepsilon$. Assume that there exists a distribution $\mathbf{p}$ over polynomials $p\colon \mathbb{F}^n \to \mathbb{F}$ such that:*

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

We give the proof of Lemma 22 in the full version of the paper (see [15, Section 4]).

In the second setting, which is useful for our upper-bound proof (see [15, Section 5]), 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. We defer the details of the second setting to the full version (see [15, Section 4]).

B 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 23** (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 $q \leq n^{1/\gamma'}$ is a prime power, $d \leq n/4$, and:*

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

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 23, note that if we are willing to assume that $d \leq n^{.49}$, then we can choose $\gamma = .499$ and $\gamma' > 0$ that is sufficiently small, and the three conditions in Theorem 23 hold for every $q \leq n^{1/\gamma'}$ and $t \leq \gamma'' \cdot d$.

To prove Theorem 23 we will instantiate Corollary 20 with a linear disperser that we will construct relying on the extractor of Shaltiel and Umans [36]. Recall that [36] constructed an extractor $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^m$ 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 [39]).

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., $q \leq n^{\gamma'}$ for some small constant $\gamma' > 0$). To do so, we take as a starting-point their construction of a q_0-ary extractor from [36], 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 [36].

▶ **Definition 24** (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 $\mathsf{Ext}_0\colon \mathbb{F}_{q_0}^n \times \{0,1\}^\ell \to \mathbb{F}_{q_0}^m$ 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 $P\colon \mathbb{F}_{q_0}^{i-1} \to \mathbb{F}_{q_0}^{\rho^{-2}}$, it holds that $\Pr_{x\sim\mathbf{x}, u\sim\mathbf{u}_\ell}[P(\mathsf{Ext}_0(x,u)_1, ..., \mathsf{Ext}_0(x,u)_{i-1}) \ni \mathsf{Ext}_0(x,u)_i] \leq \rho$.*

▶ **Theorem 25** ([36, 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 $\mathsf{Ext}_0\colon \mathbb{F}_{q_0}^{n_0} \times \{0,1\}^{r\cdot\log(q_0)} \to \mathbb{F}_{q_0}^m$ 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 $p_x\colon \mathbb{F}_{q_0}^r \to \mathbb{F}_{q_0}$ 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 [36] the input of the extractor is represented in binary and interpreted as n_0 elements in $\mathbb{F}_q$, whereas in Theorem 25 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 [36, 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, given in the full version, generalizes an idea from [39]. 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 unpredictable", and hence the distribution of elements from $\mathbb{F}_q$ is a *block source* (see, e.g., [41, Section 6.3.1]). Following Nisan and Zuckerman [34], 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., [38, Claim 4.1]), relying on well-known constructions of such codes.[14]

▶ **Proposition 26** (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 $\mathfrak{C}\colon \mathbb{F}_q^\Delta \to \mathbb{F}_q^\Delta$ be a $(1 - 1/q - \rho, \rho^{-2})$-list-decodable code. Assume that $\mathsf{Ext}_0\colon \mathbb{F}_{q_0}^{n_0} \times \{0,1\}^{\ell_0} \to \mathbb{F}_{q_0}^m$ is a (k,ρ) q_0-ary extractor. Let $\mathsf{Ext}\colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$, where $n = n_0 \cdot \Delta$ and $\ell = \ell_0 + \log(\bar{\Delta})$, be defined by*

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

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.

We now combine Theorem 25 and Proposition 26 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 $\ell = O(\log(n/\delta))$. We provide the proof in the full version of the paper.

▶ **Theorem 27** (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 $\mathsf{Disp}\colon \mathbb{F}_q^n \times \{0,1\}^\ell \to \mathbb{F}_q^m$, where $\ell = O_{\gamma'}(\log(n/\delta))$ and $m = \Omega_{\gamma'}\left(k/n^{\gamma+\gamma'}\right)$.*

Finally, we deduce our lower bound from Theorem 23 using Corollary 20 with the linear disperser from Theorem 27.

Proof of Theorem 23. 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 27 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 27 hold due to our hypotheses that $d/t \leq \gamma'' \cdot \frac{q-1}{\log(q)} \cdot n^\gamma$ (which implies that

[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.

$\delta \geq 2^{-n^{\gamma}+\log(2qn)}$) 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 $m = \Omega(n^{1-(\gamma+\gamma')})$.

Relying on Corollary 20, 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 $\frac{q-1}{\log(q)} \cdot \log(nt/d) \geq 1/\gamma''$. ∎

C 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 [33]. To do so let us first define the degree-d closure of a set $S \subseteq \mathbb{F}^n$:

▶ **Definition 28** (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 $\mathtt{Cl}^{(\mathtt{d})}(S)$, by $\mathtt{Cl}^{(\mathtt{d})} = \{x \in \mathbb{F}^n : \forall p \in \mathcal{I}(S), p(x) = 0\}$, where $\mathcal{I}(S) = \{p : \mathbb{F}^n \to \mathbb{F} : \deg(p) = d \wedge \forall s \in S, p(s) = 0\}$.*

We now restate and 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 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 $\left|\mathtt{Cl}^{(\mathtt{d}/2(\mathtt{t}+1))}(S)\right| > \frac{1}{2} \cdot q^{n-t}$.*

Proof. For the first statement, let $S \subseteq \mathbb{F}^n$ be such that $\left|\mathtt{Cl}^{(\mathtt{d})}(S)\right| > q^{n-t}$. 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 $\left|\mathtt{Cl}^{(\mathtt{d}')}(S)\right| \leq \frac{1}{2} \cdot q^{n-t}$, 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 $T_1 = \mathbb{F}^n \setminus \mathtt{Cl}^{(\mathtt{d}')}(S)$. 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

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

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 $\left|\{x \notin \mathtt{Cl}^{(\mathtt{d})}(S) : \forall i \in [t], p_i(x) = 0\}\right| \leq |T_1|/q^{t+1} \leq \frac{1}{2} \cdot q^{-t}$. Let $p \colon \mathbb{F}^n \to \mathbb{F}$ be the multivalued OR of $p_1, ..., p_{t+1}$, defined by $p(x) = \mathtt{mvOR}(p_1(x), ..., p_t(x))$. 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 23 and the first item of Theorem 5. Specifically, we can deduce that for every $2 \leq q \leq \mathrm{poly}(n)$ 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 [33], 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 5), 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.