# On the Probabilistic Degrees of Symmetric Boolean functions[*]

Srikanth Srinivasan[†]        Utkarsh Tripathi[‡]        S. Venkitesh[§]

### Abstract

The probabilistic degree of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined to be the smallest $d$ such that there is a random polynomial $\mathbf{P}$ of degree at most $d$ that agrees with $f$ at each point with high probability. Introduced by Razborov (1987), upper and lower bounds on probabilistic degrees of Boolean functions — specifically symmetric Boolean functions — have been used to prove explicit lower bounds, design pseudorandom generators, and devise algorithms for combinatorial problems.

In this paper, we characterize the probabilistic degrees of all symmetric Boolean functions up to polylogarithmic factors over all fields of fixed characteristic (positive or zero).

## 1   Introduction

Studying the combinatorial and computational properties of Boolean functions by representing them using multivariate polynomials (over some field $\mathbb{F}$) is an oft-used technique in Theoretical Computer Science. Such investigations into the complexity of Boolean functions have led to many important advances in the area (see, e.g. [Bei93, O'D14, Wil14] for a large list of such results).

An "obvious" way of representing a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is via a multilinear polynomial $P \in \mathbb{F}[x_1,\dots,x_n]$ such that $P(a) = f(a)$ for all $a \in \{0,1\}^n$. While such a representation has the advantage of being unique, understanding the computational complexity of $f$ sometimes requires us to understand polynomial representations where we allow some notion of error in the representation. Many such representations have been studied, but we concentrate here on the notion of *Probabilistic degree* of a Boolean function, introduced implicitly in a paper of Razborov [Raz87]. It is defined as follows.

**Definition 1** (Probabilistic polynomial and Probabilistic degree). *Given a Boolean function $f :$ $\{0,1\}^n \to \{0,1\}$ and an $\varepsilon > 0$, an $\varepsilon$-error probabilistic polynomial for $f$ is a random polynomial*

**P** *(with some distribution having finite support) over* $\mathbb{F}[x_1, \ldots, x_n]$ *such that for each* $a \in \{0,1\}^n$,

$$\Pr_{\mathbf{P}}[\mathbf{P}(a) \neq f(a)] \leq \varepsilon.$$

*We say that the degree of* **P***, denoted* $\deg(\mathbf{P})$*, is at most* $d$ *if the probability distribution defining* **P** *is supported on polynomials of degree at most* $d$*. Finally, we define the* $\varepsilon$*-error probabilistic degree of* $f$*, denoted* $\mathrm{pdeg}_{\varepsilon}^{\mathbb{F}}(f)$*, to be the least* $d$ *such that* $f$ *has an* $\varepsilon$*-error probabilistic polynomial of degree at most* $d$*.*

*When the field* $\mathbb{F}$ *is clear from context, we use* $\mathrm{pdeg}_{\varepsilon}(f)$ *instead of* $\mathrm{pdeg}_{\varepsilon}^{\mathbb{F}}(f)$*.*

Intuitively, if we think of multivariate polynomials as algorithms and degree as a notion of efficiency, then a low-degree probabilistic polynomial for a Boolean function $f$ is an efficient *randomized* algorithm for $f$.

The study of the probabilistic degree itself is by now a classical topic, and has had important repercussions for other problems. We list three such examples below, referring the reader to the papers for definitions and exact statements of the results.

- Razborov [Raz87] showed strong upper bounds on the probabilistic degree of the OR function over fields of (fixed) positive characteristic. Along with lower bounds on the probabilistic degree of some symmetric Boolean functions,[1] this led to the first lower bounds for the Boolean circuit class $\mathrm{AC}^0[p]$, for prime $p$ [Raz87, Smo87a, Smo93a].

- Tarui [Tar93] and Beigel, Reingold and Spielman [BRS91] showed upper bounds on the probabilistic degree of the OR function over any characteristic (and in particular over the reals). This leads to probabilistic degree upper bounds for the circuit class $\mathrm{AC}^0$, which was used by Braverman [Bra10] to resolve a long-standing open problem of Linial and Nisan [LN90] regarding pseudorandom generators for $\mathrm{AC}^0$.

- Alman and Williams [AW15] showed that for constant error, the probabilistic degree of any symmetric Boolean function is at most $O(\sqrt{n})$, and used this to obtain the first subquadratic algorithm for an offline version of the Nearest Neighbour problem in the Hamming metric.

In all the above results, it was important to understand the probabilistic degree of a certain class of symmetric Boolean functions. However, the problem of *characterizing* the probabilistic degree of symmetric Boolean functions in general does not seem to have been considered. This is somewhat surprising, since this problem has been considered in a variety of other computational models, such as $\mathrm{AC}^0$ circuits of polynomial size [FKPS85, BW87], $\mathrm{AC}^0[p]$ circuits of quasipolynomial size [Lu01], Approximate degree[2] [Pat92] and constant-depth Perceptrons[3] of quasipolynomial size [ZBT93].

---

[1]Recall that a *symmetric* Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a function such that $f(x)$ depends only on the Hamming weight of $x$. Examples include the threshold functions, Parity (counting modulo 2), etc..

[2]A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is said to have approximate degree at most $d$ if there is a degree $d$ polynomial $P \in \mathbb{R}[x_1, \ldots, x_n]$ such that at each $a \in \{0,1\}^n$, $|f(a) - P(a)| \leq 1/4$.

[3]These are constant-depth circuits that have an output Majority gate with $\mathrm{AC}^0$ circuits feeding into it.

**Our result.** In this paper, we give an almost-complete understanding of the probabilistic degrees of all symmetric Boolean functions over all fields of fixed positive characteristic and characteristic 0. For each Boolean function $f$ on $n$ variables, our upper bounds and lower bounds on $\operatorname{pdeg}(f)$ are separated only by polylogarithmic factors in $n$.

We now introduce some notation and give a formal statement of our result. We shall use the notation $[a, b]$ to denote an interval in $\mathbb{R}$ as well as an interval in $\mathbb{Z}$; the distinction will be clear from the context. Throughout, fix some field $\mathbb{F}$ of characteristic $p$ which is either a fixed positive constant or 0. Let $n$ be a growing integer parameter which will always be the number of input variables. We use $s\mathcal{B}_n$ to denote the set of all symmetric Boolean functions on $n$ variables. Note that each symmetric Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ is uniquely specified by a string $\operatorname{Spec} f : [0, n] \to \{0, 1\}$, which we call the *Spectrum* of $f$, in the sense that for any $a \in \{0, 1\}^n$, we have

$$f(a) = \operatorname{Spec} f(|a|).$$

Given a $f \in s\mathcal{B}_n$, we define the *period of $f$*, denoted $\operatorname{per}(f)$, to be the smallest positive integer $b$ such that $\operatorname{Spec} f(i) = \operatorname{Spec} f(i + b)$ for all $i \in [0, n - b]$. We say $f$ is *$k$-bounded* if $\operatorname{Spec} f$ is constant on the interval $[k, n - k]$; let $B(f)$ denote the smallest $k$ such that $f$ is $k$-bounded.

**Standard decomposition of a symmetric Boolean function [Lu01].** Fix any $f \in s\mathcal{B}_n$. Among all symmetric Boolean functions $f' \in s\mathcal{B}_n$ such that $\operatorname{Spec} f'(i) = \operatorname{Spec} f(i)$ for all $i \in [\lceil n/3 \rceil, \lfloor 2n/3 \rfloor]$, we choose a function $g$ such that $\operatorname{per}(g)$ is as small as possible. We call $g$ the *periodic part* of $f$. Define $h \in s\mathcal{B}_n$ by $h = f \oplus g$. We call $h$ the *bounded part* of $f$.

We will refer to the pair $(g, h)$ as a *standard decomposition* of the function $f$. Note that we have $f = g \oplus h$.

**Observation 2.** *Let $f \in s\mathcal{B}_n$ and let $(g, h)$ be a standard decomposition of $f$. Then, $\operatorname{per}(g) \leq \lfloor n/3 \rfloor$ and $B(h) \leq \lceil n/3 \rceil$.*

In this paper, we prove the following upper and lower bounds for the probabilistic degrees of symmetric Boolean functions.

**Theorem 3** (Upper bounds on probabilistic degree). *Let $\mathbb{F}$ be a field of constant characteristic $p$ (possibly 0) and $n \in \mathbb{N}$ be a growing parameter. Let $f \in s\mathcal{B}_n$ be arbitrary and let $(g, h)$ be a standard decomposition of $f$. Then we have the following for any $\varepsilon > 0$.*

1. *If $\operatorname{per}(g) = 1$, then $\operatorname{pdeg}_\varepsilon^{\mathbb{F}}(g) = 0$.*

   *If $\operatorname{per}(g)$ is a power of $p$, then $\operatorname{pdeg}_\varepsilon^{\mathbb{F}}(g) \leq \operatorname{per}(g)$, [Lu01]*

   *(Note that $\operatorname{per}(g)$ cannot be a power of $p$ if $p = 0$.)*

2. *$\operatorname{pdeg}_\varepsilon^{\mathbb{F}}(h) = \widetilde{O}(\sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon))$ if $B(h) > 1$ and $0$ otherwise, and*

3. *$\operatorname{pdeg}_\varepsilon^{\mathbb{F}}(f) = \begin{cases} O(\sqrt{n \log(1/\varepsilon)}) & \text{if } \operatorname{per}(g) > 1 \text{ and not a power of } p, \text{ [AW15]} \\ O(\min\{\sqrt{n \log(1/\varepsilon)}, \operatorname{per}(g)\}) & \text{if } \operatorname{per}(g) \text{ a power of } p \text{ and } B(h) = 0, \\ \widetilde{O}(\min\{\sqrt{n \log(1/\varepsilon)}, \operatorname{per}(g) + & \text{otherwise.} \\ \quad \sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon)\}) & \end{cases}$*

3

where the $\widetilde{O}(\cdot)$ hides polylogarithmic factors in $n$ (and are independent of $\varepsilon$). When $p$ is positive, we can replaced the $\widetilde{O}(\cdot)$ with $O(\cdot)$ in all the above bounds.

We obtain almost (up to polylogarithmic factors) matching lower bounds for all symmetric Boolean functions over all fields and all errors.

**Theorem 4** (Lower bounds on probabilistic degree). *Let $\mathbb{F}$ be a field of constant characteristic $p$ (possibly $0$) and $n \in \mathbb{N}$ be a growing parameter. Let $f \in s\mathcal{B}_n$ be arbitrary and let $(g, h)$ be a standard decomposition of $f$. Then for any $\varepsilon \in [1/2^n, 1/3]$, we have*

1. $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(g) = \widetilde{\Omega}(\sqrt{n \log(1/\varepsilon)})$ *if* $\mathrm{per}(g) > 1$ *and is not a power of $p$ and* $\widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g)\})$ *otherwise.*

2. $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(h) = \widetilde{\Omega}(\sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon))$ *if* $B(h) \geq 1$, *and*

3. $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(f) = \begin{cases} \widetilde{\Omega}(\sqrt{n \log(1/\varepsilon)}) & \text{if } \mathrm{per}(g) > 1 \text{ and not a power of } p, \\ \widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g)\}) & \text{if } \mathrm{per}(g) \text{ a power of } p \text{ and } B(h) = 0, \\ \widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g) & \text{otherwise.} \\ \quad + \sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon)\}) \end{cases}$

*where the $\widetilde{\Omega}(\cdot)$ hides $\mathrm{poly}(\log n)$ factors (independent of $\varepsilon$).*

**Remark 5.** *A natural open question following our results is to remove the polylogarithmic factors separating our upper and lower bounds. We remark that in characteristic $0$, such gaps exist even for the very simple OR function despite much effort [MNV16, HS16, BHMS18]. Over positive characteristic, there is no obvious barrier, but our techniques fall short of proving tight lower bounds for natural families of functions such as the Exact Threshold functions (defined in Section 2).*

## 1.1 Proof Outline

For the outline below, we assume that the field is of fixed positive characteristic $p$.

**Upper bounds.** Given a symmetric Boolean function $f$ on $n$ variables with standard decomposition $(g, h)$, it is easy to check that $\mathrm{pdeg}_\varepsilon(f) = O(\mathrm{pdeg}_\varepsilon(g) + \mathrm{pdeg}_\varepsilon(h))$. So it suffices to upper bound the probabilistic degrees of periodic and bounded functions respectively.

For periodic functions $g$ with period a power of $p$, Lu [Lu01] showed that the *exact* degree of the Boolean functions is at most $\mathrm{per}(g)$. If the period is not a power of $p$, then we use the upper bound of Alman and Williams [AW15] that holds for all symmetric Boolean functions (as we show below, this is nearly the best that is possible).

For a $t$-constant function $h$ (defined in Section 3), we use the observation that any $t$-constant function is essentially a linear combination of the threshold functions $\mathrm{Thr}_n^0, \ldots, \mathrm{Thr}_n^t$ (defined in Section 2) and so it suffices to construct probabilistic polynomials for $\mathrm{Thr}_n^i$ for $i \in [0, t]$.[4]

---

[4]We actually need to construct probabilistic polynomials for all the threshold functions simultaneously. We ignore this point in this high-level outline.

Our main technical upper bound is a new probabilistic degree upper bound of $O(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon))$ for any threshold function $\text{Thr}_n^t$. This upper bound interpolates smoothly between a classical upper bound of $O(\log(1/\varepsilon))$ due to Razborov [Raz87] for $t = 1$ and a recent result of Alman and Williams [AW15] that yields $O(\sqrt{n \log(1/\varepsilon)})$ for $t = \Omega(n)$.

The proof of our upper bound is based on the beautiful inductive construction of Alman and Williams [AW15] which gives their above-mentioned result. The key difference between our proof and the proof of [AW15] is that we need to handle separately the case when the error $\varepsilon \leq 2^{-\Omega(t)}$.[5] In [AW15], this is a trivial case since any function on $n$ Boolean variables has an exact polynomial of degree $n$ which is at most $O(\sqrt{n \log(1/\varepsilon)})$ when $\varepsilon \leq 2^{-\Omega(n)}$. In our setting, the correct bound in this case is $O(\log(1/\varepsilon))$, which is non-obvious. We obtain this bound by a suitable modification of Razborov's technique (for $t = 1$) to handle larger thresholds.

**Lower bounds.** Here, our proof follows a result of Lu [Lu01], who gave a characterization of symmetric Boolean functions that have quasipolynomial-sized $\text{AC}^0[p]$ circuits.[6] To show circuit lower bounds for a symmetric Boolean function $h$, Lu showed how to convert a circuit $C$ computing $h$ to a circuit $C'$ computing either the Majority or a $\text{MOD}_q$ function (where $q$ and $p$ are relatively prime). Since both of these are known to be hard for $\text{AC}^0[p]$ [Raz87, Smo87a], we get the lower bound.

Lu's basic idea was to use a few restrictions[7] of $h$ along with some additional circuitry to compute either Majority or $\text{MOD}_q$. These functions are also known to have large probabilistic degree (in fact, this is the source of the $\text{AC}^0[p]$ lower bound), and so this high-level idea seems applicable to our setting as well. Indeed we do use this strategy, but our proofs are different when it comes down to the details. As Lu's aim was to derive optimal circuit lower bounds for $h$, his reductions were tailored towards using as small an amount of additional circuitry as possible. Our focus, however, is to prove the best possible probabilistic degree lower bound, so we would like our reductions to be computable by polynomials of small degree. This makes the actual reductions quite different.[8]

## 2  Preliminaries

**Some Boolean functions.** Fix some positive $n \in \mathbb{N}$. The *Majority* function $\text{Maj}_n$ on $n$ Boolean variables accepts exactly the inputs of Hamming weight greater than $n/2$. For $t \in [0, n]$, the *Threshold* function $\text{Thr}_n^t$ accepts exactly the inputs of Hamming weight at least $t$; and similarly, the *Exact Threshold* function $\text{EThr}_n^t$ accepts exactly the inputs of Hamming weight exactly $t$. Finally, for $b \in [2, n]$ and $i \in [0, b - 1]$, the function $\text{MOD}_n^{b,i}$ accepts exactly those inputs $a$ such that $|a| \equiv i \pmod{b}$. In the special case that $i = 0$, we also use $\text{MOD}_n^b$.

---

[5]This case comes up naturally in the inductive construction, even if one is ultimately only interested in the case when $\varepsilon$ is a constant.

[6]Recall that an $\text{AC}^0[p]$ circuit is a constant-depth circuit made up of gates that can compute the Boolean functions AND, OR, NOT and $\text{MOD}_p$ (defined below).

[7]A restriction of a Boolean function is obtained by setting some of its input variables to constants in $\{0, 1\}$.

[8]In an earlier version of this paper, we actually used Lu's reductions (and variants thereof) directly in the setting of probabilistic polynomials. This still works in certain parameter regimes because the additional circuitry itself has low *probabilistic* degree. However, in the setting of small error, this strategy seems to yield suboptimal results.

**Fact 6.** *We have the following simple facts about probabilistic degrees. Let $\mathbb{F}$ be any field.*

1. *(Error reduction [HS16]) For any $\delta < \varepsilon \leq 1/3$ and any Boolean function $f$, if $\mathbf{P}$ is an $\varepsilon$-error probabilistic polynomial for $f$, then $\mathbf{Q} = M(\mathbf{P}_1, \ldots, \mathbf{P}_\ell)$ is a $\delta$-error probabilistic polynomial for $f$ where $M$ is the exact multilinear polynomial for $\mathrm{Maj}_\ell$ and $\mathbf{P}_1, \ldots, \mathbf{P}_\ell$ are independent copies of $\mathbf{P}$. In particular, we have $\mathrm{pdeg}_\delta^{\mathbb{F}}(f) \leq \mathrm{pdeg}_\varepsilon^{\mathbb{F}}(f) \cdot O(\log(1/\delta)/\log(1/\varepsilon))$.*

2. *(Composition) For any Boolean function $f$ on $k$ variables and any Boolean functions $g_1, \ldots, g_k$ on a common set of $m$ variables, let $h$ denote the natural composed function $f(g_1, \ldots, g_k)$ on $m$ variables. Then, for any $\varepsilon, \delta > 0$, we have $\mathrm{pdeg}_{\varepsilon+k\delta}^{\mathbb{F}}(h) \leq \mathrm{pdeg}_\varepsilon^{\mathbb{F}}(f) \cdot \max_{i \in [k]} \mathrm{pdeg}_\delta^{\mathbb{F}}(g_i)$.*

3. *(Sum) Assume that $f, g_1, \ldots, g_k$ are all Boolean functions on a common set of $m$ variables such that $f = \sum_{i \in [k]} g_i$. Then, for any $\delta > 0$, we have $\mathrm{pdeg}_{k\delta}^{\mathbb{F}}(f) \leq \max_{i \in [k]} \mathrm{pdeg}_\delta^{\mathbb{F}}(g_i)$.*

## 2.1 Some previous results on probabilistic degree

The following upper bounds on probabilistic degrees of OR and AND functions were proved by Razborov [Raz87] and Smolensky [Smo87a] in the case of positive characteristic and Tarui [Tar93] and Beigel, Reingold and Spielman [BRS91] in the general case. For the latter, we state a slightly tighter result that follows from [Bra10, Lemma 8].

**Lemma 7** (Razborov's upper bound on probabilistic degrees of OR and AND). *Let $\mathbb{F}$ be a field of characteristic $p$. For $p > 0$, we have*

$$\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{OR}_n) = \mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{AND}_n) \leq p\lceil \log(1/\varepsilon))\rceil. \tag{1}$$

*For any $p$, we have*

$$\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{OR}_n) = \mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{AND}_n) \leq 4\lceil \log n \rceil \cdot \lceil \log(1/\varepsilon) \rceil. \tag{2}$$

*Further, the probabilistic polynomials have one-sided error in the sense that on the all $0$ input, they output $0$ with probability $1$.*

We now recall two probabilistic degree lower bounds due to Smolensky [Smo87b, Smo93b], building on the work of Razborov [Raz87].

**Lemma 8** (Smolensky's lower bound for close-to-Majority functions). *For any field $\mathbb{F}$, any $\varepsilon \in (1/2^n, 1/5)$, and any Boolean function $g$ on $n$ variables that agrees with $\mathrm{Maj}_n$ on a $1 - \varepsilon$ fraction of its inputs, we have*

$$\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(g) = \Omega(\sqrt{n \log(1/\varepsilon)}).$$

**Lemma 9** (Smolensky's lower bound for MOD functions). *For $2 \leq b \leq n/2$, any $\mathbb{F}$ such that $char(\mathbb{F}) = p$ is coprime to $b$, any $\varepsilon \in (1/2^n, 1/(3b))$, there exists an $i \in [0, b-1]$ such that*

$$\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{MOD}_n^{b,i}) = \Omega(\sqrt{n \log(1/b\varepsilon)}).$$

**Remark 10.** *From the above lemma, it also easily follows that if $b \leq n/4$, then for every $i \in [0, b-1]$, we have $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(\mathrm{MOD}_n^{b,i}) = \Omega(\sqrt{n \log(1/b\varepsilon)})$. This is the usual form in which Smolensky's lower bound is stated. The above form is slightly more useful to us because it holds for $b$ up to $n/2$.*

We will also need the following result of Alman and Williams [AW15].

**Lemma 11.** *Let $\mathbb{F}$ be any field. For any $n \geq 1, \varepsilon > 0$ and $f \in s\mathcal{B}_n$, $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(f) = O(\sqrt{n \log(1/\varepsilon)})$.*

## 2.2 A string lemma

Given a function $w : I \to \{0, 1\}$ where $I \subseteq \mathbb{N}$ is an interval, we think of $w$ as a string from the set $\{0, 1\}^{|I|}$ in the natural way. For an interval $J \subseteq I$, we denote by $w|_J$ the substring of $w$ obtained by restriction to $J$.

The following simple lemma can be found, e.g. as a consequence of [JJJ80, Chapter I, Section 2, Theorem 1].

**Lemma 12.** *Let $w \in \{0, 1\}^+$ be any non-empty string and $u, v \in \{0, 1\}^+$ such that $w = uv = vu$. Then there exists a string $z \in \{0, 1\}^+$ such that $w$ is a power of $z$ (i.e. $w = z^k$ for some $k \geq 2$).*

**Corollary 13.** *Let $g \in s\mathcal{B}_n$ be arbitrary with $\mathrm{per}(g) = b > 1$. Then for all $i, j \in [0, n - b + 1]$ such that $i \not\equiv j \pmod{b}$, we have $\mathrm{Spec}\, g|_{[i,i+b-1]} \neq \mathrm{Spec}\, g|_{[j,j+b-1]}$.*

*Proof.* Suppose $\mathrm{Spec}\, g|_{[i,i+b-1]} = \mathrm{Spec}\, g|_{[j,j+b-1]}$ for some $i \not\equiv j \pmod{b}$. Assume without loss of generality that $i < j < i + b$. Let $u = \mathrm{Spec}\, g|_{[i,j-1]}, v = \mathrm{Spec}\, g|_{[j,i+b-1]}, w = \mathrm{Spec}\, g|_{[i+b,j+b-1]}$. Then $u = w$ and the assumption $uv = vw$ implies $uv = vu$. By Lemma 12, there exists a string $z$ such that $uv = z^k$ for $k \geq 2$ and therefore $\mathrm{per}(g) < b$. This contradicts our assumption on $b$. $\square$

# 3 Upper bounds

In this section, we will first prove upper bounds on the probabilistic degree of a special class of symmetric Boolean functions that we call *t-constant functions*, and then use it to prove Theorem 3.

## 3.1 Upper bound on probabilistic degree of $t$-constant functions

**Definition 14** ($t$-constant function). *For any positive $n \in \mathbb{N}$ and $t \in [0, n]$, a Boolean function $f \in s\mathcal{B}_n$ is said to be $t$-constant if $f|_{\{x:|x|\geq t\}}$ is a constant, that is, $\mathrm{Spec}\, f|_{[t,n]}$ is a constant.*

The following observation is immediate.

**Observation 15.** *A Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ is $t$-constant if and only if $f = \sum_{j=0}^{t} a_j \mathrm{Thr}_n^j$, for some $a_0, \ldots, a_t \in \{-1, 0, 1\}$. In other words, $f$ is $t$-constant if and only if there exists a linear polynomial $g(Y_0, \ldots, Y_t) = a_0 Y_0 + \cdots + a_t Y_t \in \mathbb{F}[Y_0, \ldots, Y_t]$ with $a_j \in \{-1, 0, 1\}$, $j \in [0, t]$ such that $f = g(\mathrm{Thr}_n^0, \ldots, \mathrm{Thr}_n^t)$.*

We will prove an upper bound on the probabilistic degree of $t$-constant Boolean functions. For this, we first generalize the notion of probabilistic polynomial and probabilistic degree to a *tuple* of Boolean functions. This generalization was implicit in [AW15].

**Definition 16** (Probabilistic poly-tuple and probabilistic degree). *Let $f = (f_1, \ldots, f_m) : \{0, 1\}^n \to \{0, 1\}^m$ be an $m$-tuple of Boolean functions and $\varepsilon \in (0, 1)$. An $\varepsilon$-error probabilistic poly-tuple for $f$ is a random $m$-tuple of polynomials $\mathbf{P}$ (with some distribution having finite support) from $\mathbb{F}[X_1, \ldots, X_n]^m$ such that*

$$\Pr_{\mathbf{P}}[\mathbf{P}(x) \neq f(x)] \leq \varepsilon, \quad \text{for all } x \in \{0, 1\}^n.$$

We say that the degree of $\mathbf{P}$ is at most $d$ if $\mathbf{P}$ is supported on $m$-tuples of polynomials $P = (P_1, \ldots, P_m)$ where each $P_i$ has degree at most $d$. Finally we define the $\varepsilon$-error probabilistic degree of $f$, denoted by $\mathrm{pdeg}_\varepsilon^{\mathbb{F}}(f)$, to be the least $d$ such that $f$ has an $\varepsilon$-error probabilistic poly-tuple of degree at most $d$.

We make a definition for convenience.

**Definition 17** (Threshold tuple). *For positive $n \in \mathbb{N}$, $t \in [0, n]$, an $(n, t)$-threshold tuple is any tuple of Boolean functions $(\mathrm{Thr}_n^{t_1}, \ldots, \mathrm{Thr}_n^{t_m})$, with $t_1, \ldots, t_m \in [0, t]$ and $\max\{t_1, \ldots, t_m\} \leq t$.*

The main theorem of this subsection is the following.

**Theorem 18.** *For any positive $n \in \mathbb{N}$, $t \in [0, n]$, if $T$ is an $(n, t)$-threshold tuple and $\varepsilon \in (0, 1/3)$, then*

$$\mathrm{pdeg}_\varepsilon(T) = \begin{cases} \widetilde{O}(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = 0, \\ O(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = p > 0. \end{cases}$$

As a corollary to the above theorem, we get an upper bound for the probabilistic degree of $t$-constant functions.

**Corollary 19.** *For any $t$-constant Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ and $\varepsilon \in (0, 1/3)$,*

$$\mathrm{pdeg}_\varepsilon(f) = \begin{cases} \widetilde{O}(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = 0, \\ O(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = p > 0. \end{cases}$$

*Proof.* By Observation 15, there exists $g(Y_0, \ldots, Y_t) = a_0 Y_0 + \cdots + a_t Y_t \in \mathbb{F}[Y_0, \ldots, Y_t]$ with $a_j \in \{-1, 0, 1\}$, $j \in [0, t]$ such that $f = g(\mathrm{Thr}_n^0, \ldots, \mathrm{Thr}_n^t)$. We note that $\deg g = 1$. So by Theorem 18, we get

$$\mathrm{pdeg}_\varepsilon(f) = \deg g \cdot \mathrm{pdeg}_\varepsilon(\mathrm{Thr}_n^0, \ldots, \mathrm{Thr}_n^t) = \begin{cases} \widetilde{O}(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = 0, \\ O(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)), & char(\mathbb{F}) = p > 0. \end{cases} \quad \square$$

**High-level outline of the proof.** The basic strategy behind the inductive construction of probabilistic poly-tuples is due to Alman and Williams [AW15]. We describe the construction of an $\varepsilon$-error probabilistic polynomial for a single threshold $\mathrm{Thr}_n^t$ (the construction for a tuple is similar). Assume that, by induction, we already have probabilistic polynomials $\mathbf{T}_{m,t,\varepsilon}$ for $\mathrm{Thr}_m^t$ where $m < n$. The idea is to try to use $\mathbf{T}_{m,t,\varepsilon}$ for $m < n$ to compute $\mathrm{Thr}_n^t(x)$. We do this by sampling: we sample a random subvector $\hat{\mathbf{x}}$ of length $n/10$ of $x$ by sampling uniform random entries of $x$ with replacement. If the Hamming weight of $x$ is "sufficiently far" from the threshold $t$, then the weight of $\hat{\mathbf{x}}$ is on the "same side" of $t/10$ as $x$ is of $t$ w.h.p. (say at least $1 - \varepsilon/4$); in particular, in this case $\mathbf{P}_{n/10, t/10, \varepsilon/4}$ gives the right answer with probability $1 - \varepsilon/4$ and we are done. However, if $|x|$ is "not sufficiently far" from $t$, then we need to do something else: here, we simply interpolate a polynomial that outputs the right answer on these values (see Theorem 21 below). Finally, to check which of "far" or the "not far" cases we are in, we again use the inductive hypothesis on the subvector $\hat{\mathbf{x}}$, which again gives the right answer with probability $1 - \varepsilon/4$. Putting these things together yields the $\varepsilon$-error probabilistic polynomial.

In the analysis of the construction above, the distance parameter (say $\theta$) that determines "far" vs. "not far" comes from the concentration properties of Bernoulli random variables (see Lemma 20 below). in our setting, $\theta$ is roughly $\sqrt{t \log(1/\varepsilon)}$. In particular, to check that $|x|$ is not much larger than $t$, we need to apply a probabilistic polynomial for the threshold function $\mathrm{Thr}_{n/10}^{t/10+\theta}$ to the random vector $\hat{\mathbf{x}}$. Here, to keep the threshold parameter bounded by $t$, we need that $\log(1/\varepsilon)$ is not much larger than $t$, or equivalently that $\varepsilon$ is not much smaller than $2^{-t}$.

When $\varepsilon$ does fall below $2^{-t}$,[9] we need to do something different, as the above inductive strategy fails. This case does not occur in [AW15] since there $t = \Omega(n)$ and when $\varepsilon \leq 2^{-n}$, we can always use an exact polynomial representation of the threshold function (which has degree $n = O(\sqrt{n \log(1/\varepsilon)})$). In our setting, though, we aim for a bound of $\tilde{O}(\log(1/\varepsilon))$ in this case, which is non-trivial. To handle this case, we use a different construction, which is a modification of Razborov's probabilistic polynomial construction for the OR function (Lemma 7 above). This changes the base case of the induction and certain elements of the inductive analysis. Overall, though, we are able to use these ideas to obtain a probabilistic polynomial of degree $\tilde{O}(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon))$ (and only a constant-factor loss when the characteristic $p > 0$).

### 3.1.1 Proof of Theorem 18

Before we prove Theorem 18, we will gather a few results that we require. The following lemma is a particular case of Bernstein's inequality (Theorem 1.4, [DP09]).

**Lemma 20.** *Let $X_1, \ldots, X_m$ be independent and identically distributed Bernoulli random variables with mean $q$. Let $X = \sum_{i=1}^{m} X_i$. Then for any $\theta > 0$,*

$$\Pr\left[|X - mq| > \theta\right] \leq 2 \exp\left(-\frac{\theta^2}{2mq(1-q) + 2\theta/3}\right).$$

We will also need the following polynomial construction.

**Theorem 21** (Lemma 3.1, [AW15])**.** *For any symmetric Boolean function $f : \{0,1\}^n \to \{0,1\}$ and integer interval $[a,b] \subseteq [0,n]$, there exists a symmetric multilinear polynomial $EX_{[a,b]}f \in \mathbb{Z}[X_1, \ldots, X_n]$ such that $\deg(EX_{[a,b]}f) \leq b - a$ and $\mathrm{Spec}\left(EX_{[a,b]}f\right)|_{[a,b]} = \mathrm{Spec}\, f|_{[a,b]}$.*

**Remark 22.** *In particular, the polynomial $EX_{[a,b]}f$ may be interpreted as a polynomial over any field $\mathbb{F}$ satisfying the above property.*

We will now prove Theorem 18.

**Proof of Theorem 18**. For any $a = (a_1, \ldots, a_k), b = (b_1, \ldots, b_k) \in \mathbb{F}^k$, fix the notation $a * b = (a_1 b_1, \ldots, a_k b_k)$. Throughout, the notation $\mathbf{1}$ will denote the constant-1 vector of appropriate length.

For positive characteristic $p$, we prove that for any positive $n \in \mathbb{N}$, $t \in [0,n]$ and $\varepsilon \in (0, 2^{-100})$, any $(n,t)$-threshold tuple $T$ has an $\varepsilon$-error probabilistic poly-tuple $\mathbf{T}$ of degree at most $A_p \sqrt{t \log(1/\varepsilon)} + B_p \log(1/\varepsilon)$, for constants $A_p = B_p = 6,400,000p$ (we make no effort to optimize

---

[9]Note that this can occur even if we are only interested in the case of (say) constant error. Since the inductive strategy causes the error to drop at each stage, even if we start with constant $\varepsilon$, after a few stages we end up in the setting where $\varepsilon < 2^{-t}$.

the constants). For $p = 0$, we prove a similar result with a degree bound of $A_0 \log n \cdot \sqrt{t \log(1/\varepsilon)} + B_0 \log n \cdot \log(1/\varepsilon)$, for $A_0 = B_0 = 64,000,000$. This will prove the theorem for $\varepsilon \le 2^{-100}$. To prove the theorem for all $\varepsilon \le 1/3$, we use error reduction (Fact 6) and reduce the error to $2^{-100}$ and then apply the result for small error.

The proof is by induction on the parameters $n, t$ and $\varepsilon$. At any stage of the induction, given an $(n, t)$-threshold tuple with error parameter $\varepsilon$, we construct the required probabilistic poly-tuple by using the probabilistic poly-tuples (guaranteed by inductive hypothesis) for suitable threshold poly-tuples with $n/10$ inputs and error parameter $\varepsilon/4$. Thus the base cases of the induction are as follows.

**Base Case:** Suppose $n \le 10$. Let $T = (T_1, \ldots, T_m)$ be an $(n, t)$-threshold tuple. Let $Q_1, \ldots, Q_m$ be the unique multilinear polynomial representations of $T_1, \ldots, T_m$ respectively. Then $Q = (Q_1, \ldots, Q_m)$ is an $\varepsilon$-error probabilistic poly-tuple for $T$, for all $\varepsilon \in (0, 2^{-100})$, with $\deg Q \le n = 10 \le B_p \log(1/\varepsilon)$ for all $p$ (including 0). Hence the claim is proved in this case.

**Base Case:** Suppose $\varepsilon \le 2^{-t/160000}$. Let $T = (T_1, \ldots, T_m) = (\mathrm{Thr}_n^{t_1}, \ldots, \mathrm{Thr}_n^{t_m})$ be any $(n, t)$-threshold tuple and let $r = 6400000 \log(1/\varepsilon)$.

Suppose $n \le r$. Let $Q_1, \ldots, Q_m$ be the unique multilinear representations of $T_1, \ldots, T_m$ respectively. Then $Q = (Q_1, \ldots, Q_m)$ is an $\varepsilon$-error probabilistic polynomial with $\deg Q \le n \le r$. This proves the claim in this case.

Now suppose $n > r$. We first describe how to construct the probabilistic poly-tuple $\mathbf{P}$ in this case. Assume for now that $\mathrm{char}(\mathbb{F}) = p > 0$.

- Let $P_1 = (\mathrm{EX}_{[0,r]} T_1, \ldots, \mathrm{EX}_{[0,r]} T_m)$. Then $\deg P_1 \le r$.

- Choose a uniformly random hash function $\mathbf{H} : [n] \to [r]$ and let $\mathbf{S}_j = \mathbf{H}^{-1}(j)$, $j \in [r]$. Choose $\alpha_i \sim \mathbb{F}_p$, $i \in [n]$ independently and uniformly at random and define $\mathbf{L}_j(x) = \sum_{i \in \mathbf{S}_j} \alpha_i x_i$, $x \in \{0, 1\}^n$, $j \in [r]$. For $i \in [m]$, let $\mathbf{P}_2^{(i)} = Q_r^{(i)}(\mathbf{L}_1^{p-1}, \ldots, \mathbf{L}_r^{p-1})$, where $Q_r^{(i)}$ is the unique multilinear polynomial representation of $\mathrm{Thr}_r^{t_i}$. Let $\mathbf{P}_2 = (\mathbf{P}_2^{(1)}, \ldots, \mathbf{P}_2^{(m)})$. Note that $\deg(\mathbf{P}_2) \le (p-1) \cdot \left( \max_i \deg(Q_r^{(i)}) \right) \le (p-1) \cdot r$.

- Define $\mathbf{P} = \mathbf{1} - (\mathbf{1} - P_1) * (\mathbf{1} - \mathbf{P}_2)$, that is, $\mathbf{P} = (\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(m)})$, where $\mathbf{P}^{(i)} = \mathrm{OR}_2(P_1^{(i)}, \mathbf{P}_2^{(i)})$, for all $i \in [m]$. We have $\deg(\mathbf{P}) \le \deg(P_1) + \deg(\mathbf{P}_2) \le p \cdot r \le B_p \log(1/\varepsilon)$.

We now show that $\mathbf{P}$ is indeed an $\varepsilon$-error probabilistic poly-tuple for $T$. Note that since $\varepsilon \le 2^{-t/160000}$, we have $r = 6400000 \log(1/\varepsilon) \ge 40t > t$. Thus $t_i \le t \le r$, for all $i \in [m]$. Now fix any $a \in \{0, 1\}^n$. Let $Z_a = \{i \in [m] : \mathrm{Thr}_n^{t_i}(a) = 0\}$ and $N_a = \{i \in [m] : \mathrm{Thr}_n^{t_i}(a) = 1\}$. So we have $|a| < t_i \le t \le r$ and hence $\mathrm{EX}_{[0,r]} T_i(a) = 0$, for all $i \in Z_a$. Also $|(\mathbf{L}_1^{p-1}(a), \ldots, \mathbf{L}_r^{p-1}(a))| \le |a| < t_i$ w.p.1, and so $\mathbf{P}_2^{(i)}(a) = Q_r^{(i)}((\mathbf{L}_1^{p-1}(a), \ldots, \mathbf{L}_r^{p-1}(a))) = 0$ w.p.1, for all $i \in Z_a$ simultaneously. Thus $\mathbf{P}^{(i)}(a) = 0$ w.p.1, for all $i \in Z_a$ simultaneously.

Further we have $|a| \ge t_i$, for all $i \in N_a$. We will now show that $\mathbf{P}^{(i)}(a) = 1$ w.p. at least $1 - \varepsilon$, for all $i \in N_a$ simultaneously. If $|a| \le r$, then again $P_1^{(i)}(a) = 1$, for all $i \in N_a$ and so $\mathbf{P}^{(i)}(a) = 1$ w.p.1. Now suppose $|a| \ge r$ (note that in this case, $N_a = [m]$). Without loss of generality, assume $t_1 \le \cdots \le t_m = t$. Then we have $\mathbf{P}_2^{(1)}(a) \ge \cdots \ge \mathbf{P}_2^{(m)}(a)$ w.p. 1, under the order $1 > 0$. So it is enough to show that $\mathbf{P}_2^{(m)}(a) = 1$ w.p. at least $1 - \varepsilon$.

Define $I(\mathbf{H}) = \{j \in [r] : \operatorname{supp}(a) \cap \mathbf{S}_j \neq \emptyset\}$. We get

$$\Pr\left[\mathbf{P}_2^{(m)}(a) = 0\right] = \Pr\left[\mathbf{P}_2^{(m)}(a) = 0 \;\middle|\; |I(\mathbf{H})| < r/10\right] \cdot \Pr\left[|I(\mathbf{H})| < r/10\right]$$

$$+ \Pr\left[\mathbf{P}_2^{(m)}(a) = 0 \;\middle|\; |I(\mathbf{H})| \geq r/10\right] \cdot \Pr\left[|I(\mathbf{H})| \geq r/10\right]$$

$$\leq \Pr\left[|I(\mathbf{H})| < r/10\right] + \max_{\mathbf{H}:|I(\mathbf{H})|\geq r/10} \Pr\left[\mathbf{P}_2^{(m)}(a) = 0 \;\middle|\; \mathbf{H}\right].$$

Since the function $\mathbf{H} : [n] \to [r]$ is chosen uniformly at random, the probability that $I(\mathbf{H}) \subseteq I$ for any set $I \subseteq [r]$ is $(|I|/r)^{|a|}$. Using the fact that $|a| > r$ and the union bound, we get

$$\Pr\left[|I(\mathbf{H})| < r/10\right] \leq \sum_{I \subseteq [r], |I|=r/10} \Pr\left[I(\mathbf{H}) \subset I\right] \leq \binom{r}{r/10} \frac{1}{10^r} \leq \frac{1}{4^r} \leq \frac{\varepsilon}{4}.$$

Now fix any $\mathbf{H}$ such that $|I(\mathbf{H})| \geq r/10$, and let $\ell = |I(\mathbf{H})|$. Note that $\mathbf{P}_2^{(m)}(a)$ is $0$ if and only if at most $t - 1$ many $\mathbf{L}_j(a)$ are non-zero. We consider only $j \in I(\mathbf{H})$. For each $j \in I(\mathbf{H})$, the probability that $\mathbf{L}_j(a)$ is non-zero is $1 - 1/p \geq 1/2$. Let $\mathbf{Z}$ denote the number of such $\mathbf{L}_j$ $(j \in I(\mathbf{H}))$. Thus, the expected value of $\mathbf{Z}$ is at least $\ell/2 \geq r/20 \geq 2t$. Thus, by Lemma 20,

$$\Pr\left[\mathbf{P}_2^{(m)}(a) = 0 \;\middle|\; \mathbf{H}\right] = \Pr\left[|I(\mathbf{H}) \cap \{j : \mathbf{L}_j(a) = 1\}| \leq t - 1 \mid \mathbf{H}\right]$$

$$\leq \Pr\left[|\mathbf{Z} - \mathbf{E}[\mathbf{Z}]| \geq \ell/4 \mid \mathbf{H}\right]$$

$$\leq 2\exp\left(-\frac{\ell^2/16}{2 \cdot \ell \cdot (1/4) + (2/3) \cdot (\ell/4)}\right) < \frac{\varepsilon}{2}.$$

where for the final inequality we have used the fact that $\ell \geq r/10 \geq 640000 \log(1/\varepsilon)$. Thus $\Pr\left[\mathbf{P}_2^{(m)}(a) = 0\right] \leq \varepsilon$, proving the claim when $\operatorname{char}(\mathbb{F}) = p > 0$.

Now suppose $\operatorname{char}(\mathbb{F}) = 0$. Then we use the same construction as above except for one change: for $i \in [m]$ we let $\mathbf{P}_2^{(i)} = Q_r^{(i)}(\mathbf{O}_1, \ldots, \mathbf{O}_r)$, where $Q_r^{(i)}$ is the unique multilinear polynomial representation of $\operatorname{Thr}_r^{t_i}$, and for $j \in [r]$, $\mathbf{O}_j$ is the $1/4$-error probabilistic polynomial for $\operatorname{OR}_{\mathbf{S}_j}$, the OR function on variables $(X_k : k \in \mathbf{S}_j)$, given to us by Lemma 7. One can verify that the degree in this case is bounded as above by $10r \log n \leq B_0 \log n \cdot \log(1/\varepsilon)$. The rest of the analysis follows similarly, proving the base case when $\operatorname{char}(\mathbb{F}) = 0$.

**Inductive Construction.** For any positive characteristic $p$, any $n' < n$, $t' \in [0, n']$ and $\varepsilon' \in (0, 2^{-100})$, assume the existence of an $\varepsilon'$-error probabilistic poly-tuple for any $(n', t')$-threshold tuple, with degree at most $A_p\sqrt{t' \log(1/\varepsilon')} + B_p \log(1/\varepsilon')$; similarly, for characteristic zero, assume we have a probabilistic poly-tuple of degree $A_0 \log n \cdot \sqrt{t' \log(1/\varepsilon')} + B_0 \log n \cdot \log(1/\varepsilon')$.

We now consider an $(n, t)$-threshold tuple $T = (T_1, \ldots, T_m) = (\operatorname{Thr}_n^{t_1}, \ldots, \operatorname{Thr}_n^{t_m})$. Assume that the parameter $\varepsilon > 2^{-t/160000}$ since otherwise we can use the construction from the base case. Define

$$T' = (T_1', \ldots, T_m') = \left(\operatorname{Thr}_{n/10}^{t_1/10}, \ldots, \operatorname{Thr}_{n/10}^{t_m/10}\right),$$

$$T_+'' = (T_{1,+}'', \ldots, T_{m,+}'') = \left(\operatorname{Thr}_{n/10}^{t_1/10+20\sqrt{t \log(1/\varepsilon)}}, \ldots, \operatorname{Thr}_{n/10}^{t_m/10+20\sqrt{t \log(1/\varepsilon)}}\right),$$

$$T_-'' = (T_{1,-}'', \ldots, T_{m,-}'') = \left(\operatorname{Thr}_{n/10}^{t_1/10-20\sqrt{t \log(1/\varepsilon)}}, \ldots, \operatorname{Thr}_{n/10}^{t_m/10-20\sqrt{t \log(1/\varepsilon)}}\right).$$

By induction hypothesis, let $\mathbf{T}', \mathbf{T}''_+, \mathbf{T}''_-$ be $\varepsilon/4$-error probabilistic poly-tuples for $T', T''_+, T''_-$ respectively. Let $\mathbf{N}'' = (\mathbf{1} - \mathbf{T}''_+) * \mathbf{T}''_-$. For any $x \in \{0,1\}^n$, choose a random subvector $\hat{\mathbf{x}} \in \{0,1\}^{n/10}$ with each coordinate of $\hat{\mathbf{x}}$ chosen independently and uniformly at random from among the $n$ coordinates of $x$, with replacement. Define

$$\mathbf{T}(x) = \mathbf{N}''(\hat{\mathbf{x}}) * E(x) + (\mathbf{1} - \mathbf{N}'')(\hat{\mathbf{x}}) * \mathbf{T}'(\hat{\mathbf{x}}),$$

where $E = (E_1, \ldots, E_m)$, with $E_i = \mathrm{EX}_{[t_i - 300\sqrt{t\log(1/\varepsilon)}, t_i + 300\sqrt{t\log(1/\varepsilon)}]} \mathrm{Thr}_n^{t_i}$, $i \in [m]$. We will now prove that $\mathbf{T}$ is an $\varepsilon$-error probabilistic poly-tuple for $T$.

**Correctness of Inductive Construction.** We now check that the construction above gives an $\varepsilon$-error probabilistic poly-tuple for $T$. Fix any $a \in \{0,1\}^n$. Let $\hat{a} \in \{0,1\}^{n/10}$ be chosen as given in the inductive construction.

Suppose $|a| \leq 2t$. Let $\theta = 10\sqrt{t\log(1/\varepsilon)}$. Applying Lemma 20, we get

$$\Pr\left[||\hat{a}| - |a|/10| > \theta\right] \leq 2\exp\left(-\frac{\theta^2}{2 \cdot |a| \cdot (|a|/n)(1 - |a|/n) + (2\theta/3)}\right)$$

$$\leq 2\exp\left(-\frac{100t\log(1/\varepsilon)}{2 \cdot (2t) \cdot (1/4) + 7\sqrt{t\log(1/\varepsilon)}}\right)$$

$$\leq 2\exp\left(-\frac{100t\log(1/\varepsilon)}{t + 7\sqrt{t\log(1/\varepsilon)}}\right)$$

$$\leq 2\exp\left(-\frac{100t\log(1/\varepsilon)}{2t}\right) \leq \varepsilon/4.$$

where for the third inequality, we have used the fact that $\log(1/\varepsilon) \leq t/160000$.

By induction hypothesis, the probability that $\mathbf{T}'(\hat{a})$ does not agree with $T'(\hat{a})$ is at most $\varepsilon/4$, and similarly for $\mathbf{T}''_+$ and $\mathbf{T}''_-$. Let $\mathcal{G}_a$ be the event that none of the above events occur and that $||\hat{a}| - (|a|/10)| \leq \theta$; by a union bound, the event $\mathcal{G}_a$ occurs with probability at least $1 - \varepsilon$. In this case, we show that $\mathbf{T}(a) = T(a)$, which will prove the correctness of the construction in the case that $|a| \leq 2t$.

To see this, observe the following for each $i \in [m]$.

- $\mathbf{T}'_i(\hat{a}) = T_i(a)$ if $||a| - t_i| > 10\theta$. This is because $\mathbf{T}'_i(\hat{a}) = T'_i(\hat{a})$ by our assumption that the event $\mathcal{G}_a$ has occurred. Further, we also have $T'_i(\hat{a}) = T_i(a)$ since $|\hat{a} - |a|/10| \leq \theta$ (by occurrence of $\mathcal{G}_a$) and hence $|a| \geq t_i$ if and only if $|\hat{a}| \geq t_i/10$.

- If $||a| - t_i| > 30\theta$, then $\mathbf{N}''_i(\hat{a}) = 0$. This is because $||\hat{a}| - |a|/10| \leq \theta$ and hence $||\hat{a}| - t_i/10| > 2\theta$. Hence, either $\mathbf{T}''_{i,+}(\hat{a}) = 1$ or $\mathbf{T}''_{i,-}(\hat{a}) = 0$ and therefore, $\mathbf{N}''_i(\hat{a}) = 0$.

  Thus, when $||a| - t_i| > 30\theta$, the definition of $\mathbf{T}$ yields $\mathbf{T}_i(a) = \mathbf{T}'_i(\hat{a})$ which equals $T_i(a)$ whenever $||a| - t_i| > 10\theta$ as argued above. We are therefore done in this case.

- If $||a| - t_i| < 10\theta$, then $\mathbf{N}''_i(\hat{a}) = 1$. This is similar to the analogous statement above.

  Therefore, when $||a| - t_i| < 10\theta$, we have $\mathbf{T}_i(a) = E_i(a) = T_i(a)$ as $|a| \in [t_i - 300\sqrt{t\log(1/\varepsilon)}, t_i + 300\sqrt{t\log(1/\varepsilon)}]$. Hence, we are done in this case also.

12

- If $10\theta \le ||a| - t_i| \le 30\theta$, then $E_i(a) = \mathbf{T}'(\hat{\mathbf{a}}) = T_i(a)$. Since $\mathbf{N}''_i(\hat{\mathbf{a}}) \in \{0, 1\}$ for each $i \in [m]$, we again obtain $\mathbf{T}_i(a) = T_i(a)$.

This shows that for any $a$ such that $|a| \le 2t$, whenever $\mathcal{G}_a$ does not occur, $\mathbf{T}(a) = T(a)$.

Now suppose $|a| > 2t$. Then by Bernstein's inequality (Lemma 20), we get

$$\Pr\left[|\hat{\mathbf{a}}| < 1.5t/10\right] \le \Pr\left[\left||\hat{\mathbf{a}}| - (|a|/10)\right| \ge (|a|/40)\right]$$

$$\le 2\exp\left(-\frac{(|a|/40)^2}{2 \cdot |a| \cdot (|a|/n)(1 - (|a|/n)) + (2/3) \cdot (|a|/40)}\right)$$

$$\le 2\exp\left(-\frac{(|a|/40)^2}{2 \cdot |a| \cdot (1/4) + (2/3) \cdot (|a|/40)}\right)$$

$$\le 2\exp\left(-\frac{(|a|/40)^2}{(3|a|/5)}\right) \le 2\exp\left(-\frac{|a|}{960}\right)$$

$$\le 2\exp(-t/480) < \varepsilon/2.$$

Also, by the induction hypothesis, the probability that $\mathbf{T}'(\hat{\mathbf{a}})$ does not agree with $T'(\hat{\mathbf{a}})$ is at most $\varepsilon/4$, and similarly for $\mathbf{T}''_+$ and $\mathbf{T}''_-$. Let $\mathcal{G}_a$ denote the event that none of the above events occur and that $|a| \ge 3t/20$; we have $\Pr\left[\mathcal{G}_a\right] \ge 1 - \varepsilon$. As above, we show that when $\mathcal{G}_a$ occurs, then $\mathbf{T}(a) = T(a)$.

To see this, we proceed as follows.

- Since $|a| \ge 2t$ and $|\hat{\mathbf{a}}| \ge 3t/20$, both $T(a)$ and $\mathbf{T}'_i(\hat{a})$ are both the constant-1 vector.

- Further, we note that we have $\mathbf{N}''_i(\hat{\mathbf{a}}) = 0$ for each $i \in [m]$. This is because $||\hat{\mathbf{a}}| - t_i/10| \ge (|\hat{\mathbf{a}}| - t/10) \ge t/20 > 20\sqrt{t\log(1/\varepsilon)}$, where the final inequality uses $t > 160000\log(1/\varepsilon)$.

  This implies that $\mathbf{T}_i(a) = \mathbf{T}'_i(\hat{\mathbf{a}}) = 1$ for each $i \in [m]$.

Hence, when $\mathcal{G}_a$ does not occur, we have $\mathbf{T}(a) = T(a)$, which proves the correctness of the construction.

**Correctness of Degree.** We need to argue that $\deg(\mathbf{T})$ satisfies the inductive claim. Suppose $\mathrm{char}(\mathbb{F}) = p > 0$. We have

$$\deg \mathbf{T} \le \deg \mathbf{N}'' + \max\{\deg E, \deg \mathbf{T}'\}$$
$$\le \deg \mathbf{T}''_+ + \deg \mathbf{T}''_- + \max\{\deg E, \deg \mathbf{T}'\}.$$

Recall that $A_p = B_p = 6400000p$. Now

$$
\deg \mathbf{T}''_+ + \deg \mathbf{T}''_- \leq A_p\left(\sqrt{\frac{t}{10} + 20\sqrt{t\log\left(\frac{1}{\varepsilon}\right)}} + \sqrt{\frac{t}{10} - 20\sqrt{t\log\left(\frac{1}{\varepsilon}\right)}}\right)\sqrt{\log\left(\frac{4}{\varepsilon}\right)}
$$

$$
+ 2B_p\log\left(\frac{4}{\varepsilon}\right)
$$

$$
\leq A_p\sqrt{\left(\frac{t}{5} + 2\sqrt{\frac{t^2}{100} - 400t\log\left(\frac{1}{\varepsilon}\right)}\right)\log\left(\frac{4}{\varepsilon}\right)} + 2B_p\log\left(\frac{4}{\varepsilon}\right)
$$

$$
\leq A_p\sqrt{\left(\frac{t}{5} + 2\sqrt{\frac{t^2}{100} - \frac{t^2}{400}}\right)\log\left(\frac{4}{\varepsilon}\right)} + 2B_p\log\left(\frac{4}{\varepsilon}\right)
$$

$$
\leq A_p\sqrt{\left(\frac{2+\sqrt{3}}{10}\right)t\log\left(\frac{4}{\varepsilon}\right)} + 2B_p\log\left(\frac{4}{\varepsilon}\right)
$$

$$
\leq A_p\sqrt{\frac{38t}{100}\log\left(\frac{4}{\varepsilon}\right)} + 2B_p\log\left(\frac{4}{\varepsilon}\right),
$$

and

$$
\max\{\deg E, \deg \mathbf{T}'\} \leq \max\left\{600\sqrt{t\log\left(\frac{1}{\varepsilon}\right)}, A_p\sqrt{\frac{t}{10}\log\left(\frac{4}{\varepsilon}\right)} + B_p\log\left(\frac{4}{\varepsilon}\right)\right\}
$$

$$
= A_p\sqrt{\frac{t}{10}\log\left(\frac{4}{\varepsilon}\right)} + B_p\log\left(\frac{4}{\varepsilon}\right).
$$

So we get

$$
\deg \mathbf{T} \leq \left(\sqrt{\frac{38}{100}} + \sqrt{\frac{1}{10}}\right)A_p\sqrt{t\log\left(\frac{4}{\varepsilon}\right)} + 3B_p\log\left(\frac{4}{\varepsilon}\right)
$$

$$
\leq \frac{94}{100}A_p\sqrt{\left(t\log\left(\frac{1}{\varepsilon}\right) + 2t\right)} + 3B_p\log\left(\frac{1}{\varepsilon}\right) + 6B_p
$$

$$
\leq \frac{95}{100}A_p\sqrt{t\log\left(\frac{1}{\varepsilon}\right)} + 4B_p\log\left(\frac{1}{\varepsilon}\right)
$$

$$
\leq \frac{95}{100}A_p\sqrt{t\log\left(\frac{1}{\varepsilon}\right)} + 3B_p\log\left(\frac{1}{\varepsilon}\right) + B_p\log\left(\frac{1}{\varepsilon}\right)
$$

$$
\leq A_p\sqrt{t\log\left(\frac{1}{\varepsilon}\right)} + B_p\log\left(\frac{1}{\varepsilon}\right).
$$

where the third inequality uses $\varepsilon \leq 2^{-100}$ and the final inequality uses $t > 160000\log(1/\varepsilon)$.

Now if $\mathrm{char}(\mathbb{F}) = 0$, then we get a similar degree bound with $A_0 = B_0 = 64000000$. This completes the argument for correctness of degree. $\qquad\square$

## 3.2  Upper bound on $\mathrm{pdeg}_\varepsilon(g)$

This result is due to Lu [Lu01] but a proof is sketched here for completeness.

Recall that $\mathrm{char}(\mathbb{F}) = p$.

When $\mathrm{per}(g) = 1$, $g$ is a constant function and hence the result is trivial. So assume that $\mathrm{per}(g) = p^t$ for $t \geq 1$. In this case, we show that $g$ can be represented *exactly* as a linear combination of elementary symmetric polynomials of degree at most $D = p^t - 1$, which clearly proves the upper bound stated in the theorem. To see that every such $g$ has such a representation, we proceed as follows.

Let $V$ be the vector space generated by all functions $f : \{0,1\}^n \to \mathbb{F}$ such that $f$ can be written as a linear combination of elementary symmetric polynomials of degree at most $D$. Clearly, since there is a 1-1 correspondence between multilinear polynomials and functions from $\{0,1\}^n$ to $\mathbb{F}$, the vector space $V$ has dimension exactly $D+1 = p^t$. Each function in $f$ is a symmetric (not necessarily Boolean) function on $\{0,1\}^n$. Further, a standard application of Lucas' theorem (see [Luc78]) shows that each $f \in V$ satisfies

$$\mathrm{Spec}\, f(i) = \mathrm{Spec}\, f(i + p^t) \tag{3}$$

for each $i \leq n - p^t$.

Now, consider the vector space $W$ of all functions $f' : \{0,1\}^n \to \mathbb{F}$ that satisfy property (3). Clearly, $W \subseteq V$. Furthermore, the functions $\mathrm{MOD}^i_{p^t}$ ($i \in [0, p^t - 1]$) is a set of $p^t$ many linearly independent functions in $W$. Hence, the dimension of $W$ is exactly $p^t$ and therefore, $W = V$.

Since $g \in W$, we immediately see that $g \in V$ and is hence a linear combination of elementary symmetric polynomials of degree at most $D$.

## 3.3  Upper bound for $\mathrm{pdeg}_\varepsilon(h)$.

Let $B(h) = k$. Thus we can write $h = h_1 + (1 - \widetilde{h_2})$, for $k$-constant symmetric Boolean functions $h_1, h_2$, where $\widetilde{h_2}(x_1, \ldots, x_n) = h_2(1 - x_1, \ldots, 1 - x_n)$. But then by Corollary 19, $\mathrm{pdeg}_\varepsilon(h_1) = \mathrm{pdeg}_\varepsilon(h_2) = O(\sqrt{k \log(1/\varepsilon)} + \log(1/\varepsilon))$ and so $\mathrm{pdeg}_\varepsilon(h) = O(\sqrt{k \log(1/\varepsilon)} + \log(1/\varepsilon))$ over positive characteristic $p$. For $p = 0$, we obtain the same upper bound up to log-factors.

## 3.4  Upper bound for $\mathrm{pdeg}_\varepsilon(f)$.

Let $(g, h)$ be the standard decomposition of $f$. So $f = g \oplus h = g + h - 2gh$. Further, we already have the Alman-Williams bound of $O(\sqrt{n \log(1/\varepsilon)})$ on $\mathrm{pdeg}_\varepsilon(f)$ (Lemma 11). So we get $\mathrm{pdeg}_\varepsilon(f) = O(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g) + \sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon)\})$ over positive characteristic and the same bound up to log-factors over characteristic 0. This concludes the proof of Theorem 3.

# 4  Lower Bounds

We now prove the lower bounds given in Theorem 4. Throughout this section, let $\mathbb{F}$ be any field. We use $\mathrm{pdeg}_\varepsilon(\cdot)$ instead of $\mathrm{pdeg}^{\mathbb{F}}_\varepsilon(\cdot)$.

**High-level outline of proof.** We use a similar proof strategy to Lu [Lu01], who gave a characterization of symmetric functions computable by quasipolynomial-sized $\mathrm{AC}^0[p]$ circuits. To prove a lower bound on the probabilistic degree of a symmetric function $f \in s\mathcal{B}_n$, we will use known lower bounds for Majority (Lemma 8) and $\mathrm{MOD}_q$ functions (Lemma 9), where $q$ is relatively prime to the characteristic $p$. The basic idea is to use a few "restrictions" of $f$ to compute either a Majority function or a $\mathrm{MOD}_q$ function.

Here, a restriction of $f$ is a function $h \in s\mathcal{B}_m$ obtained by setting a few inputs of $f$ to 0s and 1s. That is, we define $h(x) = f(x0^a1^{n-m-a})$ for some $a$. From the definition of such an $h$, it is clear that for any $\delta > 0$, $\mathrm{pdeg}_\delta(h) \leq \mathrm{pdeg}_\delta(f)$. We design a small number of such restrictions $h_1, \ldots, h_\ell$ and a "combining function" $P : \{0,1\}^\ell \to \{0,1\}$ such that either Majority of $\mathrm{MOD}_q$ can be written as $P(h_1, \ldots, h_\ell)$.

The main restriction we will have on $P$ is that it should be a low-degree polynomial.[10] Given this, using Fact 6, we can write

$$\mathrm{pdeg}_\varepsilon(P(h_1, \ldots, h_\ell)) \leq \deg(P) \cdot \max_i \mathrm{pdeg}_{\varepsilon/\ell}(h_i) \leq \deg(P) \cdot \mathrm{pdeg}_{\varepsilon/\ell}(f) \leq \mathrm{pdeg}_\varepsilon(h) \cdot (\log b \cdot \log \ell).$$

Using lower bounds on the probabilistic degree of Majority and $\mathrm{MOD}_q$, we then get lower bounds on the probabilistic degree of $f$.

The non-trivial part is to determine the hard function we use in the reduction and how to carry out the reduction with a polynomial $P$ of low degree. Both of these are dependent on the structure of $\mathrm{Spec}\, f$. We give the details below.

We start with a preliminary lemma.

**Lemma 23.** *Let $\mathcal{F}$ be a set of functions mapping $[0, n-1]$ to $\{0,1\}$ such that*

- *$1 - f \in \mathcal{F}$, for all $f \in \mathcal{F}$.*

- *For every $i, j \in [0, n-1]$, $i \neq j$, there exists $f \in \mathcal{F}$ such that $f(i) = 1$, $f(j) = 0$.*

*Then there exists $S \subseteq \mathcal{F}$ such that $|S| \leq \log n$ and $f_S = \prod_{f \in S} f$ has support size 1.[11]*

*Proof.* It is enough to prove that for every positive integer $k \leq \lfloor \log n \rfloor$, there exists $S_k \subseteq \mathcal{F}$ such that $|S| \leq k$ and $1 \leq |\mathrm{supp}(f_{S_k})| \leq n/2^k$. We do so by induction on $k$.

If $k = 1$, consider any $f \in \mathcal{F}$. If $|\mathrm{supp}(f)| \leq n/2$, then we choose $S = \{f\}$; otherwise we have $|\mathrm{supp}(1 - f)| \leq n/2$ and then we choose $S = \{1 - f\}$. Now consider any $k \in [\lfloor \log n \rfloor - 1]$ and assume the existence of $S_k$. If $|\mathrm{supp}(f_{S_k})| = 1$, then we choose $S_{k+1} = S_k$, which satisfies the required conditions. Now suppose $|\mathrm{supp}(f_{S_k})| > 1$. Choose any $i, j \in \mathrm{supp}(f_{S_k})$, $i \neq j$. By the given condition, there exists $f_{i,j} \in \mathcal{F}$ such that $f_{i,j}(i) = 1$ and $f_{i,j}(j) = 0$. If $|\mathrm{supp}(f_{i,j}f_{S_k})| \leq |\mathrm{supp}(f_{S_k})|/2$, then $|\mathrm{supp}(f_{i,j}f_{S_k})| \leq n/2^{k+1}$ and so we choose $S_{k+1} = S_k \cup \{f_{i,j}\}$; otherwise we have $|\mathrm{supp}((1 - f_{i,j})f_{S_k})| \leq |\mathrm{supp}(f_{S_k})| \leq n/2^{k+1}$ and so we choose $S_{k+1} = S_k \cup \{1 - f_{i,j}\}$.

At the end of this process, we have a set $S$ such that $1 \leq |\mathrm{supp}(f_S)| \leq n/2^{\lfloor \log n \rfloor} < 2$ and hence $|\mathrm{supp}(f_S)| = 1$. $\qquad\square$

**Remark 24.** *The above lemma is also true if we take any interval $I = [a, a+m-1]$ in $\mathbb{Z}$ instead of $[0, m-1]$. This can be checked by further taking the bijection $\phi_a : [a, a+m-1] \to [0, m-1]$ defined as $\phi(x) = x - a$, $x \in [a, a+m-1]$.*

---

[10]This is the point of divergence from Lu's result. His constraint on the function $P$ was that it have a small $\mathrm{AC}^0[p]$ circuit. Our focus is different.

[11]Here, the product between functions mapping $[0, n-1]$ to $\mathbb{F}$ is defined pointwise.

## 4.1 Lower bound on $\mathrm{pdeg}_\varepsilon(g)$

We start with a simple consequence of Lemma 23.

Let $u : [0, m-1] \to \{0, 1\}$ be any function. Let $\tau_m : [0, m-1] \to [0, m-1]$ be the cyclic-shift operator: that is, $\tau_m(i) = i+1$ for $i < m-1$, and $\tau_m(m-1) = 0$. We say that $u$ is *periodic* if $u \circ \tau_m^j = u$ for some $j \in [m-1]$, and *aperiodic* otherwise.

**Lemma 25.** *Let $\mathbb{F}$ be any field. Fix any $u : [0, m-1] \to \{0, 1\}$ that is aperiodic. Let $u_j = f \circ \tau_m^j$ for $j \in [0, m-1]$. For any $g : [0, m-1] \to \mathbb{F}$, there is a $P \in \mathbb{F}[Y_0, \ldots, Y_m-1]$ with degree at most $\log m$ such that $g = P(u_0, \ldots, u_{m-1})$.*

*Proof.* For any $i \in [0, m-1]$, let $\delta_i : [0, m-1] \to \{0, 1\}$ be the delta function supported only at $i$. By linearity, it suffices to show the lemma for $g \in \{\delta_0, \ldots, \delta_{m-1}\}$. Further, by symmetry, it suffices to show that there is at least one $i$ such that $\delta_i$ is a polynomial of degree at most $\log m$ in $u_0, \ldots, u_{m-1}$.

To prove the latter, we use Lemma 23. Let $\mathcal{F} = \{u_j : j \in [0, m-1]\} \cup \{1 - u_j : j \in [0, m-1]\}$. We first note that $\mathcal{F}$ satisfies the hypotheses of Lemma 23. Clearly, $v \in \mathcal{F}$ implies $1 - v \in \mathcal{F}$. Now consider any $i, j \in [0, m-1]$, $i \neq j$. If $v(i) = v(j)$ for all $v \in \mathcal{F}$, then in particular, we have $u_k(i) = u_k(j)$, for all $k \in [0, m-1]$. This implies $u$ is periodic, which is a contradiction. Thus there exists $v \in \mathcal{F}$ such that $v(i) = 0$, $v(j) = 1$. So by Lemma 23, there exists $S \subseteq \mathcal{F}$, $|S| \leq \log m$ such that $\prod_{v \in S} v$ has support size 1. We have $S = \{\hat{u}_{i_1}, \ldots, \hat{u}_{i_t}\}$, where for every $k \in [t]$, $\hat{u}_{i_k}$ is either $u_{i_k}$ or $1 - u_{i_k}$. So now for $k \in [t]$, define

$$M_k(Y_0, \ldots, Y_{m-1}) = \begin{cases} Y_{i_k}, & \hat{u}_{i_k} = u_{i_k} \\ 1 - Y_{i_k}, & \hat{u}_{i_k} = 1 - u_{i_k} \end{cases}$$

Then we have the required polynomial as $P(Y_0, \ldots, Y_{m-1}) = \prod_{k \in [t]} M_k(Y_0, \ldots, Y_{m-1})$. Since $\prod_{v \in S} v$ has support size 1, we have $P(u_0, \ldots, u_{m-1}) = \delta_i$, for some $i \in [0, m-1]$. □

We now prove the lower bound for $\mathrm{pdeg}_\varepsilon(g)$.

**Lemma 26.** *For any $\varepsilon \in [2^{-n}, 1/3]$,*

$$\mathrm{pdeg}_\varepsilon(g) \geq \begin{cases} \Omega\left(\dfrac{\sqrt{n \log(1/\varepsilon)}}{\log^2 n}\right), & \mathrm{per}(g) \text{ is not a power of } p \\ \Omega\left(\dfrac{\min\{\mathrm{per}(g), \sqrt{n \log(1/\varepsilon)}\}}{\log^2 n}\right), & \mathrm{per}(g) \text{ is a power of } p. \end{cases}$$

*Proof.* By Fact 6 item 1, we know that $\mathrm{pdeg}_\varepsilon(g) = \Theta(\mathrm{pdeg}_\delta(g))$ as long as $\delta = \varepsilon^{\Theta(1)}$. In particular, we may assume without loss of generality that $\varepsilon \in [2^{-n/10}, 1/5]$.

Let $b$ denote $\mathrm{per}(g)$. We know (Observation 2) that $b := \mathrm{per}(g) \leq \lfloor n/3 \rfloor$.

We have two cases.

$b$ **is not a power of** $p$. Define the function $u : [0, b-1] \to \{0, 1\}$ by $u(i) = \mathrm{Spec}\, g(i)$. Note that for $u_j = u \circ \tau_b^j$ ($j \in [0, b-1]$), as defined in the statement of Lemma 25, we have $u_j = \mathrm{Spec}\, g(j + i)$ (the latter is well defined as $j + i < 2b < n$). Further, as $b$ is the period of $g$,

17

Corollary 13 implies that $u \neq u_j$ for all $j \in [b-1]$. This means that $u$ is aperiodic (as defined above).

Let $q$ be any prime divisor of $b$ distinct from $p$. For each $i \in [0, q-1]$, define a function $v_i : [0, b-1] \to \{0, 1\}$ by $v_i(j) = 1$ iff $j \equiv i \pmod{q}$. Lemma 25 implies that for each $i$, there is a $P_i(Y_0, \ldots, Y_{b-1})$ of degree at most $\log b$ such that $P_i(u_0, \ldots, u_{b-1}) = v_i$.

Fix any $i \in [0, b-1]$ and consider the function $G_i : \{0, 1\}^{n-b} \to \mathbb{F}$ defined by $G_i(x) = P_i(g(x0^b), g(x0^{b-1}1), \ldots, g(x01^{b-1}))$. Clearly, as all the inputs to $P_i$ are $b$-periodic symmetric functions, the same holds for the function $G_i$. Further, for any $j \in [0, b-1]$, we see that

$$
\begin{aligned}
\operatorname{Spec} G_i(j) &= P_i(\operatorname{Spec} g(j), \ldots, \operatorname{Spec} g(j+b-1)) \\
&= P_i(\operatorname{Spec} u_0(j), \ldots, \operatorname{Spec} u_{b-1}(j)) \\
&= \begin{cases} 1 & \text{if } j \equiv i \pmod{q}, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

This implies that $G_i$ is in fact the $\operatorname{MOD}_{n-b}^{q,i}$ function. Note also that $q \leq b \leq (n-b)/2$. This will be relevant below as we will apply Lemma 9 to one of the functions $G_0, \ldots, G_{q-1}$.

By Fact 6, for any $\delta > 0$, we have that

$$
\operatorname{pdeg}_\delta(G_i) \leq (\log b) \cdot \max_{i \in [0, b-1]} \operatorname{pdeg}_{\delta/b}(g(x0^{b-i}1^i))
$$

$$
\leq (\log b) \cdot \operatorname{pdeg}_{\delta/b}(g) \leq (\log b) \cdot \operatorname{pdeg}_\varepsilon(g) \cdot O\left(\frac{\log(b/\delta)}{\log(1/\varepsilon)}\right).
$$

In particular, setting $\delta = \varepsilon/q$, we have $\operatorname{pdeg}_{\varepsilon/q}(G_i) \leq \operatorname{pdeg}_\varepsilon(g) \cdot O(\log^2 b)$ for each $i \in [0, b-1]$. On the other hand, Lemma 9 implies that for some $i \in [0, q-1]$, $\operatorname{pdeg}_{\varepsilon/b}(G_i) = \Omega(\sqrt{n \log(1/\varepsilon)})$. Putting these together, we obtain the claimed lower bound on $\operatorname{pdeg}_\varepsilon(g)$.

$b$ **is a power of** $p$. In this case, we first choose parameters $m, \delta$ with the following properties.

(P1) $m \in [n-b]$ and $m \equiv (n-b) \pmod{2}$.

(P2) $(1/5) \geq \delta \geq \max\{\varepsilon, 1/2^m\}$.

(P3) $4\sqrt{m \log(1/\delta)} \leq b$.

(P4) $\sqrt{m \log(1/\delta)} = \Omega(\min\{b, \sqrt{n \log(1/\varepsilon)}\})$.

We will show later how to find $m, \delta$ satisfying these properties. Assuming this for now, we first prove the lower bound on $\operatorname{pdeg}_\varepsilon(g)$. First, we fix some $b$-periodic symmetric function $G : \{0, 1\}^{n-b} \to \{0, 1\}$ so that $G$ agrees with $\operatorname{Maj}_{n-b}$ on all inputs of weight $a \in ((n-b)/2 - 2\sqrt{m \log(1/\delta)}, (n-b)/2 + 2\sqrt{m \log(1/\delta)})$ (it is possible to define such a $b$-periodic $G$ because of property (P3) above).

As in the previous case, we can use Lemma 25 to find a polynomial $P \in \mathbb{F}[Y_0, \ldots, Y_{b-1}]$ of degree at most $\log b$ so that $\operatorname{Spec} G(j) = P(\operatorname{Spec} g(j), \operatorname{Spec} g(j+1), \ldots, \operatorname{Spec} g(j+(b-1)))$ for all $j \in [0, n-b]$. (We omit the proof as it is very similar.) This implies that for any $x \in \{0, 1\}^{n-b}$,

$$
G(x) = P(g(x0^b), g(x0^{b-1}1), \ldots, g(x01^{b-1})).
$$

In particular, $\text{pdeg}_\delta(G) \leq (\log b) \cdot \text{pdeg}_{\delta/b}(g) \leq O(\log^2 b) \cdot \text{pdeg}_\delta(g)$.

Now, consider $G' : \{0,1\}^m \to \{0,1\}$ defined by $G'(x) = G(x0^t1^t)$ where $t = (n-b-m)/2$ (note that (P1) implies that $(n-b-m)$ is even). Clearly, $G'$ agrees with $G$, and hence $\text{Maj}_m$, on inputs of Hamming weight $a \in (m/2 - 2\sqrt{m \log(1/\delta)}, m/2 + 2\sqrt{m \log(1/\delta)})$. By the Bernstein inequality (Lemma 20), it follows that $G'$ agrees with $\text{Maj}_m$ on at least a $(1-\delta)$ fraction of its inputs. Hence, by Lemma 8, it follows that $\text{pdeg}_\delta(G') = \Omega(\sqrt{m \log(1/\delta)})$.

However, we have $\text{pdeg}_\delta(G') \leq \text{pdeg}_\delta(G)$ which in turn is bounded by $O(\log^2 b) \cdot \text{pdeg}_\delta(g)$ as argued above. So we obtain

$$\text{pdeg}_\varepsilon(g) \geq \text{pdeg}_\delta(g) = \Omega\left(\frac{\sqrt{m \log(1/\delta)}}{\log^2 b}\right) = \Omega\left(\frac{\min\{b, \sqrt{n \log(1/\varepsilon)}\}}{\log^2 n}\right)$$

where the first inequality follows from the fact that $\delta \geq \varepsilon$ (by (P2)), and the second equality follows from property (P4) above and the fact that $b \leq n$.

It remains to show that we can choose $m, \delta$ satisfying (P1)-(P4) as above. This we do as follows.

1. If $b \leq 10\sqrt{n}$, we take $m$ to be the largest integer satisfying (P1) and such that $m \leq b^2/100$. The parameter $\delta$ is set to $1/5$.

2. If $b \geq n/10$, we take $m$ to be the largest integer satisfying (P1) and such that $m \leq n/100$, and set $\delta = \max\{\varepsilon, 2^{-m}\}$.

3. Finally, if $10\sqrt{n} < b < n/10$, then we take $m = n - b$ and $\delta = \max\{\varepsilon, 2^{-b^2/16m}\}$.

In each case, the verification of properties (P1)-(P4) is a routine computation. (We assume throughout that $b$ is greater than a suitably large constant, since otherwise the statement of the lemma is trivial.) This concludes the proof. $\qquad\square$

## 4.2 Lower bound on $\text{pdeg}_\varepsilon(h)$

We start with the special case of thresholds.

**Lemma 27.** *Assume $1 \leq t \leq n/2$. For any $\varepsilon \in [2^{-n}, 1/3]$,*

$$\text{pdeg}_\varepsilon(\text{Thr}_n^t) = \Omega(\sqrt{t \log(1/\varepsilon)} + \log(1/\varepsilon)).$$

*Proof.* By Fact 6 item 1, we can assume that $\varepsilon \in [2^{-n/2}, 1/5]$. The proof breaks into two cases depending on the relative magnitudes of $\varepsilon$ and $2^{-t}$.

Consider the case when $\varepsilon \geq 1/2^t$. In this case, it suffices to show that $\text{pdeg}_\varepsilon(\text{Thr}_n^t) = \Omega(\sqrt{t \log(1/\varepsilon)})$. Note that $\text{Maj}_{2t-1}(x) = \text{Thr}_{2t-1}^t(x) = \text{Thr}_n^t(x0^{n-2t+1})$ for any $x \in \{0,1\}^{2t-1}$. Therefore, we have by Lemma 8 that

$$\text{pdeg}_\varepsilon(\text{Thr}_n^t) \geq \text{pdeg}_\varepsilon(\text{Maj}_{2t-1}) = \Omega(\sqrt{t \log(1/\varepsilon)}).$$

Now, consider the case when $\varepsilon < 2^{-t}$. Now, it suffices to show that $\text{pdeg}_\varepsilon(\text{Thr}_n^t) = \Omega(\log(1/\varepsilon))$. Note that we have $\text{OR}_{\lceil n/2 \rceil}(x) = \text{Thr}_{\lceil n/2 \rceil}^1(x) = \text{Thr}_n^t(x1^{t-1}0^{\lfloor n/2 \rfloor - t + 1})$, and so the lemma is implied by the following statement. For any positive integer $m$ and $\varepsilon \in [2^{-m}, 1/3]$

$$\text{pdeg}_\varepsilon(\text{OR}_m) = \Omega(\log(1/\varepsilon)). \tag{4}$$

While the above is possibly folklore, we don't know of a reference with a proof, so we give one here. Assume $m, \varepsilon$ as above and let $D$ denote $\mathrm{pdeg}_\varepsilon(\mathrm{OR}_m)$. By setting some bits to 0, we also get $\mathrm{pdeg}_\varepsilon(\mathrm{OR}_{m_1}) \leq D$ where $m_1 = \lfloor \log(1/\varepsilon) - 1 \rfloor$. Fix such a probabilistic polynomial $\mathbf{P}$ of degree at most $D$ for $\mathrm{OR}_{m_1}$. We have for every $x \in \{0,1\}^{m_1}$,

$$\Pr_{\mathbf{P}} \left[ \mathbf{P}(x) \neq \mathrm{OR}_{m_1}(x) \right] \leq \varepsilon < \frac{1}{2^{m_1}}.$$

By a union bound, there is some polynomial $P$ in the support of the probability distribution underlying $\mathbf{P}$ that agrees with the function $\mathrm{OR}_{m_1}$ everywhere. However, the unique multilinear polynomial representing $\mathrm{OR}_{m_1}$ has degree $m_1$. Hence, we see that $\deg(\mathbf{P}) \geq \deg(P) = m_1 = \Omega(\log(1/\varepsilon))$ concluding the proof of (4). $\qquad\square$

We now prove the lower bound on $\mathrm{pdeg}_\varepsilon(h)$ from Theorem 4.

**Lemma 28.** *Assume $B(h) \geq 1$. Then, $\varepsilon \in [2^{-n}, 1/3]$,*

$$\mathrm{pdeg}_\varepsilon(h) = \Omega \left( \frac{\sqrt{B(h)\log(1/\varepsilon)} + \log(1/\varepsilon)}{\log n} \right).$$

*Proof.* As in Lemma 26, we can assume that $\varepsilon \in [2^{-n/10}, 1/5]$.

By Observation 2, we have $b := B(h) \leq \lceil n/3 \rceil$ and further, that either $\mathrm{Spec}\, h(b-1) = 1$ or $\mathrm{Spec}\, h(n-b+1) = 1$. We assume that $\mathrm{Spec}\, h(b-1) = 1$ (the other case is similar).

Let $m = \lfloor n/6 \rfloor$ and $t = \lceil b/3 \rceil$. Note that $t \leq m$.

For $i \in [0, t-1]$, define $h_i \in s\mathcal{B}_{m+t}$ by

$$h_i(x) = h(x1^{b-t+i}0^{n-m-b-i}).$$

Then for every $i \in [0, t-1]$, we have $\mathrm{Spec}\, h_i = y_i 1 0^{m+i+1}$, for some $y_i \in \{0,1\}^{t-1-i}$. By standard linear algebra, it follows that the vector $1^t 0^{m+1}$ is in the span of the vectors $\mathrm{Spec}\, h_i$ $(i \in [0, t-1])$. It follows that we can write $1 - \mathrm{Thr}_{m+t}^t = \sum_{i=0}^{t-1} \alpha_i h_i$ for some choice of $\alpha_0, \ldots, \alpha_{t-1} \in \mathbb{F}$. So we have

$$\mathrm{pdeg}_\varepsilon(\mathrm{Thr}_{m+t}^t) \leq \max_i \mathrm{pdeg}_{\varepsilon/t}(h_i) \leq \mathrm{pdeg}_{\varepsilon/t}(h) \leq O(\log b) \cdot \mathrm{pdeg}_\varepsilon(h).$$

Lemma 27 now implies the lower bound. $\qquad\square$

## 4.3 Lower bound on $\mathrm{pdeg}_\varepsilon(f)$

We start with a slightly weaker lower bound on $\mathrm{pdeg}_\varepsilon(f)$ that is independent of $h$.

**Lemma 29.** *For any $\varepsilon \in [2^{-n}, 1/3]$,*

$$\mathrm{pdeg}_\varepsilon(f) \geq \begin{cases} \Omega\left( \frac{\sqrt{n\log(1/\varepsilon)}}{\log^2 n} \right) & \text{if } \mathrm{per}(g) > 1 \text{ and not a power of } p, \\ \Omega\left( \frac{\min\{\sqrt{n\log(1/\varepsilon)}, \mathrm{per}(g)\}}{\log^2 n} \right) & \text{if } \mathrm{per}(g) \text{ a power of } p. \end{cases}$$

*Proof.* Similar to Lemma 26, we may assume that $\varepsilon \in [2^{-n/100}, 1/5]$.

The proof of this lemma splits into two cases depending on the magnitude of $b := \mathrm{per}(g)$. Let $n_1 = n - 2\lceil n/3 \rceil$.

Assume first that $b \leq \lfloor n_1/3 \rfloor$. Define $f' \in s\mathcal{B}_{n_1}$ by $f'(x) = f(x0^{\lceil n/3 \rceil}1^{\lceil n/3 \rceil})$. By our choice of the function $g$, the function $f'$ is also a function with period $b$. Now, the proof of Lemma 26 shows that $\mathrm{pdeg}_\varepsilon(f')$ is $\widetilde{\Omega}(\sqrt{n \log(1/\varepsilon)})$ if $b > 1$ and not a power of $p$, and $\widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, b\})$ if $b$ is a power of $p$. The same lower bound immediately applies to $f$ also, and hence the lemma is proved in this case.

From now on, we assume that $b > m := \lfloor n_1/3 \rfloor$. In particular, this implies that there is no periodic symmetric function in $s\mathcal{B}_n$ with period at most $m$ that agrees with $f$ on inputs of weight in $I := [\lceil n/3 \rceil, \lfloor 2n/3 \rfloor]$. Thus, for each $k \in [m]$, there exist $r_k \in I$ such that $r_k + k \in I$ and $\mathrm{Spec}\, f(r_k) \neq \mathrm{Spec}\, f(r_k + k)$.

Now, define a set of functions from $s\mathcal{B}_{3m}$ as follows. For each $i, j \in [m+1, 2m]$ with $i < j$, define $f_{i,j}(x)$ as follows. Set $k = j - i$. For $r_k$ as defined above, let

$$f_{i,j}(x) = f(x1^{r_k - i}0^{n - 3m - r_k + i}). \tag{5}$$

The parameters above are chosen so that $\mathrm{Spec}\, f_{i,j}(i) = \mathrm{Spec}\, f(r_k)$ and $\mathrm{Spec}\, f_{i,j}(j) = \mathrm{Spec}\, f(r_k+k)$. Consequently, we have $\mathrm{Spec}\, f_{i,j}(i) \neq \mathrm{Spec}\, f_{i,j}(j)$. Let $u_{i,j} : [m+1, 2m] \rightarrow \{0,1\}$ denote the restriction of the function $\mathrm{Spec}\, f_{i,j}$ to the interval $[m+1, 2m]$. We denote by $\mathcal{U}$ the set $\{u_{i,j}, 1 - u_{i,j} \mid m+1 \leq i < j \leq 2m\}$.

By Lemma 23, we know that there is a subset $\mathcal{U}' \subseteq \mathcal{U}$ such that $s := |\mathcal{U}'| = O(\log m)$ and $\prod_{u \in \mathcal{U}'} u$ has support $\{a\}$ for some $a \in [m+1, 2m]$. Assume that $\mathcal{U}' = \{\hat{u}_{i_1, j_1}, \ldots, \hat{u}_{i_s, j_s}\}$ where each $\hat{u}_{i_t, j_t}$ is either $u_{i_t, j_t}$ or $1 - u_{i_t, j_t}$ for $t \in [s]$.

Define $\hat{f}_{i_t, j_t}$ to be $f_{i_t, j_t}$ if $\hat{u}_{i_t, j_t} = u_{i_t, j_t}$, and $1 - f_{i_t, j_t}$ otherwise. Let $\mathcal{F}' = \{\hat{f}_{i_t, j_t} \mid t \in [s]\}$. It follows from the properties of $\mathcal{U}'$ that $G := \prod_{F \in \mathcal{F}'} F \in s\mathcal{B}_{3m}$ satisfies

$$\mathrm{Spec}\, G(a') = \begin{cases} 1 & \text{if } a' = a, \\ 0 & \text{if } a' \in [m+1, 2m] \setminus \{a\}. \end{cases}$$

We will now use $G$ to construct the Majority function on $\Theta(m)$ inputs. We assume that $a \leq 3m/2$ (the other case is similar). Let $m_1 = \lfloor m/2 \rfloor$ and define $G_i \in s\mathcal{B}_{m_1}$ for $i \in [0, m_1]$ by

$$G_i(x) = G(x1^{a-i}0^{3m-a+i-m_1}).$$

Note that $\mathrm{Spec}\, G_i = y_i 10^{m_1 - i - 1}$ for some $y_i \in \{0,1\}^i$. In particular, the vectors $\mathrm{Spec}\, G_i \in \{0,1\}^{m_1}$ are linearly independent and hence span $\mathbb{F}^{m_1+1}$. As a result, we can write $\mathrm{Spec}\, \mathrm{Maj}_{m_1} = \sum_{i=0}^{m_1} \alpha_i \cdot (\mathrm{Spec}\, G_i)$ for some choice of $\alpha_0, \ldots, \alpha_{m_1} \in \mathbb{F}$. Equivalently, we have $\mathrm{Maj}_{m_1} = \sum_{i=0}^{m_1} \alpha_i G_i$.

This implies that

$$\mathrm{pdeg}_\varepsilon(\mathrm{Maj}_{m_1}) \leq \max_{i \in [0, m_1]} \mathrm{pdeg}_{\varepsilon/m_1}(G_i) \leq \mathrm{pdeg}_{\varepsilon/m}(G).$$

The function $G$ in turn is a product of $s = O(\log m)$ many functions from $\mathcal{F}'$. By (5), each $f' \in \mathcal{F}'$ satisfies $\mathrm{pdeg}_\delta(f') \leq \mathrm{pdeg}_\delta(f)$ for any $\delta > 0$. Hence, we have

$$\begin{aligned} \mathrm{pdeg}_\varepsilon(\mathrm{Maj}_{m_1}) &\leq \mathrm{pdeg}_{\varepsilon/m}(G) \\ &\leq O(\log m) \cdot \max_{m+1 \leq i < j \leq 2m, f' \in \mathcal{F}'} \mathrm{pdeg}_{\varepsilon/(m \log m)}(f') \\ &\leq O(\log m) \cdot \mathrm{pdeg}_{\varepsilon/(m \log m)}(f) \leq O(\log^2 m) \cdot \mathrm{pdeg}_\varepsilon(f). \end{aligned}$$

As $\varepsilon \geq 2^{-n/100} \geq 2^{-m_1}$, Lemma 8 implies that $\mathrm{pdeg}_\varepsilon(\mathrm{Maj}_{m_1}) = \Omega(\sqrt{m_1 \log(1/\varepsilon)}) = \Omega(\sqrt{n \log(1/\varepsilon)})$. Along with the above inequality, this implies the desired lower bound on $\mathrm{pdeg}_\varepsilon(f)$. $\quad\square$

We are now ready to prove the final lower bound on $\mathrm{pdeg}_\varepsilon(f)$.

**Lemma 30.** *For any $\varepsilon \in [2^{-n}, 1/3]$,*

$$
\mathrm{pdeg}_\varepsilon(f) \geq
\begin{cases}
\widetilde{\Omega}(\sqrt{n \log(1/\varepsilon)}) & \text{if } \mathrm{per}(g) > 1 \text{ and not a power of } p, \\
\widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g)\}) & \text{if } \mathrm{per}(g) \text{ a power of } p \text{ and } B(h) = 0, \\
\widetilde{\Omega}(\min\{\sqrt{n \log(1/\varepsilon)}, \mathrm{per}(g) & \text{otherwise.} \\
\quad + \sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon)\})
\end{cases}
$$

*Proof.* Lemma 29 already implies the result in the case that any of the following conditions hold.

- $\mathrm{per}(g)$ is not a power of $p$, or

- $\mathrm{per}(g)$ is a power of $p$ and $\mathrm{per}(g) \geq \sqrt{n \log(1/\varepsilon)}$, or

- $B(h) = 0$.

So from now, we assume that $\mathrm{per}(g)$ is a power of $p$ upper-bounded by $\sqrt{n \log(1/\varepsilon)}$ and that $B(h) \geq 1$. In this case, Lemma 29 shows that $\mathrm{pdeg}(f) \geq \widetilde{\Omega}(\mathrm{per}(g))$. On the other hand, since $B(h) \leq n$ and $\varepsilon \geq 2^{-n}$, the lower bound we need to show is $\widetilde{\Omega}(\mathrm{per}(g) + \sqrt{B(h) \log(1/\varepsilon)} + \log(1/\varepsilon))$. By Lemma 28, it suffices to show a lower bound of $\widetilde{\Omega}(\mathrm{per}(g) + \mathrm{pdeg}_\varepsilon(h))$.

The analysis splits into two simple cases based on the relative magnitudes of $\mathrm{per}(g)$ and $\mathrm{pdeg}_\varepsilon(h)$.

Assume first that $\mathrm{pdeg}_\varepsilon(h) \leq 4 \cdot \mathrm{per}(g)$. In this case, we are trivially done, because we already have $\mathrm{pdeg}(f) = \widetilde{\Omega}(\mathrm{per}(g))$, which is $\widetilde{\Omega}(\mathrm{pdeg}(g) + \mathrm{pdeg}_\varepsilon(h))$ as a result of our assumption.

Now assume that $\mathrm{pdeg}_\varepsilon(h) > 4 \cdot \mathrm{per}(g)$. We know that $f = g \oplus h$ and hence $h = f \oplus g$. Hence, we have

$$
\mathrm{pdeg}_\varepsilon(h) \leq 2(\mathrm{pdeg}_{\varepsilon/2}(f) + \mathrm{pdeg}_{\varepsilon/2}(g)) \leq O(\mathrm{pdeg}_\varepsilon(f)) + 2 \cdot \mathrm{per}(g),
$$

where the first inequality is a consequence of Fact 6 item 2 and the second is a consequence of Fact 6 item 1 and Theorem 3. The above yields

$$
\mathrm{pdeg}_\varepsilon(f) = \Omega((\mathrm{pdeg}_\varepsilon(h) - 2 \cdot \mathrm{per}(g))) = \Omega(\mathrm{pdeg}_\varepsilon(h)) = \Omega(\mathrm{per}(g) + \mathrm{pdeg}_\varepsilon(h)).
$$

This finishes the proof. $\quad\square$

# References

[AW15]      Josh Alman and Ryan Williams. Probabilistic polynomials and hamming nearest neighbors. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 136–150. IEEE, 2015.

[Bei93]     Richard Beigel. The polynomial method in circuit complexity. *[1993] Proceedings of the Eigth Annual Structure in Complexity Theory Conference*, pages 82–95, 1993.

[BHMS18]    Siddharth Bhandari, Prahladh Harsha, Tulasimohan Molli, and Srikanth Srinivasan. On the Probabilistic Degree of OR over the Reals. In Sumit Ganguly and Paritosh Pandya, editors, *38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2018)*, volume 122 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:12, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[Bra10]     Mark Braverman. Polylogarithmic independence fools $AC^0$ circuits. *J. ACM*, 57(5), 2010.

[BRS91]     Richard Beigel, Nick Reingold, and Daniel A. Spielman. The perceptron strikes back. In *Proceedings of the Sixth Annual Structure in Complexity Theory Conference, Chicago, Illinois, USA, June 30 - July 3, 1991*, pages 286–291, 1991.

[BW87]      Bettina Brustmann and Ingo Wegener. The complexity of symmetric functions in bounded-depth circuits. *Inf. Process. Lett.*, 25(4):217–219, 1987.

[DP09]      Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[FKPS85]    Ronald Fagin, Maria M. Klawe, Nicholas Pippenger, and Larry J. Stockmeyer. Bounded-depth, polynomial-size circuits for symmetric functions. *Theor. Comput. Sci.*, 36:239–250, 1985.

[HS16]      Prahladh Harsha and Srikanth Srinivasan. On Polynomial Approximations to $AC^0$. In Klaus Jansen, Claire Mathieu, José D. P. Rolim, and Chris Umans, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[JJJ80]     David Lawrence Johnson, David Leroy Johnson, and SS Johnson. *Topics in the theory of group presentations*, volume 42. Cambridge University Press, 1980.

[LN90]      Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[Lu01]      Chi-Jen Lu. An exact characterization of symmetric functions in qAC$^0$[2]. *Theoretical Computer Science*, 261(2):297–303, 2001.

[Luc78]   Edouard Lucas. Théorie des fonctions numériques simplement périodiques. *American Journal of Mathematics*, pages 289–321, 1878.

[MNV16]   Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12(1):1–17, 2016.

[O'D14]   Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, New York, NY, USA, 2014.

[Pat92]   Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 468–474. ACM, 1992.

[Raz87]   Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematicheskie Zametki*, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987).

[Smo87a]  Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82. ACM, 1987.

[Smo87b]  Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[Smo93a]  Roman Smolensky. On representations by low-degree polynomials. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 130–138. IEEE, 1993.

[Smo93b]  Roman Smolensky. On representations by low-degree polynomials. In *FOCS*, pages 130–138, 1993.

[Tar93]   Jun Tarui. Probablistic polynomials, $AC^0$ functions, and the polynomial-time hierarchy. *Theoretical Computer Science*, 113(1):167–183, 1993.

[Wil14]   Richard Ryan Williams. The polynomial method in circuit complexity applied to algorithm design (invited talk). In *34th International Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

[ZBT93]   Zhi-Li Zhang, David A Mix Barrington, and Jun Tarui. Computing symmetric functions with and/or circuits and a single majority gate. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 535–544. Springer, 1993.