

On polynomial approximations over $\mathbb{Z}/2^k\mathbb{Z}^*$

Abhishek Bhrushundi[†] Prahladh Harsha[‡] Srikanth Srinivasan[§]

Abstract

We study approximation of Boolean functions by low-degree polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$. More precisely, given a Boolean function $F : \{0, 1\}^n \rightarrow \{0, 1\}$, define its k -lift to be $F_k : \{0, 1\}^n \rightarrow \{0, 2^{k-1}\}$ by $F_k(x) = 2^{k-F(x)} \pmod{2^k}$. We consider the fractional agreement (which we refer to as $\gamma_{d,k}(F)$) of F_k with degree d polynomials from $\mathbb{Z}/2^k\mathbb{Z}[x_1, \dots, x_n]$.

Our results are the following:

- Increasing k can help: We observe that as k increases, $\gamma_{d,k}(F)$ cannot decrease. We give two kinds of examples where $\gamma_{d,k}(F)$ actually increases. The first is an infinite family of functions F such that $\gamma_{2d,2}(F) - \gamma_{3d-1,1}(F) \geq \Omega(1)$. The second is an infinite family of functions F such that $\gamma_{d,1}(F) \leq \frac{1}{2} + o(1)$ — as small as possible — but $\gamma_{d,3}(F) \geq \frac{1}{2} + \Omega(1)$.
- Increasing k doesn't always help: Adapting a proof of Green [*Comput. Complexity*, 9(1):16–38, 2000], we show that irrespective of the value of k , the Majority function Maj_n satisfies

$$\gamma_{d,k}(\text{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

In other words, polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ for large k do not approximate the majority function any better than polynomials over $\mathbb{Z}/2\mathbb{Z}$.

We observe that the model we study subsumes the model of *non-classical polynomials* in the sense that proving bounds in our model implies bounds on the agreement of non-classical polynomials with Boolean functions. In particular, our results answer questions raised by Bhowmick and Lovett [*In Proc. 30th Computational Complexity Conf., pages 72–87, 2015*] that ask whether non-classical polynomials approximate Boolean functions better than classical polynomials of the same degree.

*A preliminary version of this paper appeared in *Proc. 34th Annual Symp. on Theoretical Aspects of Comp. Science (STACS), 2017* [BHS17].

[†]Department of Computer Science, Rutgers University, USA. abhishek.bhr@cs.rutgers.edu. Work done while the author was visiting the Tata Institute of Fundamental Research. Research supported in part by UGC-ISF grant 6-2/2014(IC).

[‡]Tata Institute of Fundamental Research, India. prahladh@tifr.res.in. Research supported in part by UGC-ISF grant 6-2/2014(IC).

[§]Department of Mathematics, Indian Institute of Technology, Bombay, India. srikanth@math.iitb.ac.in.

1 Introduction

Many lower bound results in circuit complexity are proved by showing that any small sized circuit in a given circuit class can be approximated by a function from a simple computational model (e.g., small depth circuits by low-degree polynomials) and subsequently showing that this is not possible for some suitable “hard function”.

A classic case in point is the work of Razborov [Raz87] which shows lower bounds for $AC^0[\oplus]$, the class of constant depth circuits made up of AND, OR and \oplus gates. Razborov shows that any small $AC^0[\oplus]$ circuit C can be well approximated by a low-degree multivariate polynomial $Q(x_1, \dots, x_n) \in \mathbb{F}_2[x_1, \dots, x_n]$ in the sense that

$$\Pr_{x \sim \{0,1\}^n} [Q(x) \neq C(x)] = o(1).$$

The next step in the proof is to show that the hard function, on the other hand, does not have any such approximation. Razborov does this for a suitable symmetric function, Smolensky [Smo87] for the MOD_q function (for constant odd q), and Szegedy [Sze89] and Smolensky [Smo93] for the Majority function Maj_n on n bits.

Given the importance of the above lower bound, polynomial approximations in other domains and metrics have been intensely investigated and have resulted in interesting combinatorial constructions and error-correcting codes [Gro00, Efr12], learning algorithms [LMN93, KS04] and more recently in the design of algorithms for combinatorial problems [Wil14, AWY15] as well.

To describe the model of polynomial approximation considered in this paper, we first recall the Razborov [Raz87] model of polynomial approximation. Given a Boolean function $F : \{0,1\}^n \rightarrow \{0,1\}$ and degree $d \leq n$, Razborov considers the largest γ such that there is a degree d polynomial $Q \in \mathbb{F}_2[x_1, \dots, x_n]$ that has agreement at least γ with F (i.e., $\Pr_x [Q(x) = F(x)] \geq \gamma$). Call this $\gamma_d(F)$. In this notation, Szegedy [Sze89] and Smolensky’s [Smo93] results for the Majority function can be succinctly stated as

$$\gamma_d(\text{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

We consider a generalization of the above model to rings $\mathbb{Z}/2^k\mathbb{Z}$ in the following simple manner. To begin with, we consider the ring $\mathbb{Z}/4\mathbb{Z}$. Given a Boolean function F , let $F_2 : \{0,1\}^n \rightarrow \{0,2\} \subseteq \mathbb{Z}/4\mathbb{Z}$ be the 2-lift of F defined as $F_2(x) := 2^{2-F(x)}$ (i.e., $F_2(x) := 0$ if $F(x) = 0$ and $F_2(x) := 2$ otherwise). Once again, we can define $\gamma_{d,2}(F)$ to be the largest γ such that there exists a degree d polynomial $Q_2 \in \mathbb{Z}/4\mathbb{Z}[x_1, \dots, x_n]$ that has agreement γ with F_2 . Note that $\gamma_{d,2}(F) \geq \gamma_d(F)$ since if, for instance, $Q(x) = x_1x_2 + x_3 \in \mathbb{F}_2[x_1, \dots, x_n]$ has agreement γ with F , then $Q_2 := 2(x_1x_2 + x_3) \in \mathbb{Z}/4\mathbb{Z}[x_1, \dots, x_n]$ also has the same agreement γ with F_2 . Hence, proving upper bounds for $\gamma_{d,2}(F)$ is at least as hard as proving upper bounds for $\gamma_d(F)$.

More generally, we can extend these definitions to $\gamma_{d,k}(F)$, the agreement of F_k , the k -lift of F , defined as $F_k(x) = 2^{k-F(x)} \pmod{2^k}$, with degree d polynomials from $\mathbb{Z}/2^k\mathbb{Z}[x_1, \dots, x_n]$. It is not hard to show that $\gamma_{d,k+1}(F) \geq \gamma_{d,k}(F)$ and hence as k increases, the problem of proving upper bounds on $\gamma_{d,k}(F)$ can only get harder.

Our motivation for this model comes from a recent work of Bhowmick and Lovett [BL15], who study the maximum agreement between *non-classical polynomials* of degree d and a Boolean function F , which is similar to $\gamma_{d,d}(F)$ (see Section 5 for an exact translation between the above model

and non-classical polynomials). In particular, non-classical polynomials of degree d can be considered as a subset of the degree d polynomials in $\mathbb{Z}/2^d\mathbb{Z}[x_1, \dots, x_n]$. With respect to correlation¹, Bhowmick and Lovett showed that there exist non-classical polynomials (and hence polynomials in $\mathbb{Z}/2^d\mathbb{Z}[x_1, \dots, x_n]$) of logarithmic degree that have very good correlation with the Maj_n function. With respect to agreement, they show that low-degree non-classical polynomials can only have *small* agreement with the Majority function. Their results stated in our language, imply that

$$\gamma_{d,d}(\text{Maj}_n) \leq \frac{1}{2} + \frac{O(d \cdot 2^d)}{\sqrt{n}}.$$

In particular, if $d = \Omega(\log n)$, this result unfortunately does not give any non-trivial bound on the maximum agreement between non-classical polynomials of degree d and the Maj_n function. Bhowmick and Lovett, however, conjectured that this result could be improved and left open the question of whether non-classical polynomials of degree d can do any better than classical polynomials of the same degree in approximating the Majority function. More generally, they informally conjectured that although non-classical polynomials achieve better correlation with Boolean functions than their classical counterparts, they possibly do not approximate Boolean functions any better than classical polynomials. Our work stems from trying to answer these questions.

1.1 Our results

We prove the following results about agreement of Boolean functions with polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$:

1. We explore whether there exist Boolean functions for which agreement can increase by increasing k . In particular, do there exist Boolean F such that $\gamma_{d,k}(F) > \gamma_{d,1}(F)$?

It is not hard to show that this is impossible for $d = 1$. Further, it can be shown that if $\gamma_{d,k}(F) > 1 - \frac{1}{2^d}$, then $\gamma_{d,k}(F) = \gamma_{d,1}(F)$. Keeping this in mind, the first place where we can expect larger k to show better agreement is $\gamma_{2,2}$ vs. $\gamma_{2,1}$. Our first result shows that there are indeed separating examples in the regime.

- (a) Fix $d \in \mathbb{N}$ to be any power of 2. For infinitely many n , there exists a Boolean function $F : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $\gamma_{3d-1,1}(F) \leq 5/8 + o(1)$ but $\gamma_{2d,2}(F) \geq 3/4 - o(1)$.

Note that since F is Boolean, $\gamma_{d,k}(F) \geq 1/2$ for any d, k . We then ask if there exist Boolean functions F such that $\gamma_{d,1}(F)$ is more or less the trivial bound of $1/2$, while $\gamma_{d',k}(F)$ is significantly larger for $d' \leq d$ and some $k > 1$. In this context, we show the following result.

- (b) Fix any $\ell \geq 2$. For large enough n , there is a Boolean function $F : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $\gamma_{2^\ell-1,1}(F) \leq 1/2 + o(1)$ but $\gamma_{d,3}(F) \geq 9/16 - o(1)$, for $d = 2^{\ell-1} + 2^{\ell-2} \leq 2^\ell - 1$.

2. We show that for Maj_n , the majority function on n bits, and any $d, k \in \mathbb{Z}^+$,

$$\gamma_{d,k}(\text{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}},$$

¹The correlation between $F, G : \{0, 1\}^n \rightarrow \mathbb{Z}/2^k\mathbb{Z}$ is defined to be $\mathbb{E}_x[\omega^{F(x)-G(x)}]$ where ω is the primitive 2^k th root of unity in \mathbb{C} . If F, G are $\{0, 2^{k-1}\}$ -valued, then this quantity is exactly $2\gamma - 1$ where γ is the agreement between F and G . Otherwise, however, it does not measure agreement.

² by adapting a proof due to Green [Gre00] of a result on the approximability of the parity function by low-degree polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$ for prime $p \neq 2$.

Coupled with the observation that the class of polynomials over rings $\mathbb{Z}/2^k\mathbb{Z}$ subsumes the class of non-classical polynomials, part (b) of the first result provides a counterexample to an informal conjecture of Bhowmick and Lovett [BL15] that, for any Boolean function F , non-classical polynomials of degree d do not approximate F any better than classical polynomials of the same degree, and the second result confirms their conjecture that non-classical polynomials do not approximate the Majority function any better than classical polynomials.

1.2 Organisation

We start with some preliminaries in Section 2. In Section 3, we show some separation results. Next, in Section 4, we prove upper bounds for $\gamma_{d,k}(\text{Maj}_n)$. Finally, in Section 5, we discuss how our model relates to non-classical polynomials, answering questions raised by Bhowmick and Lovett.

2 Preliminaries

For $x \in \{0,1\}^n$, $|x|$ denotes the Hamming weight of x , and for $i \geq 0$, $|x|_i$ is the $(i+1)$ th least significant bit of $|x|$ in base 2. For $d \in \mathbb{N}$, we use $\{0,1\}_{\leq d}^n$ (resp. $\{0,1\}_{=d}^n$) to denote the set of elements in $\{0,1\}^n$ of Hamming weight at most d (resp. exactly d). We use \mathcal{F}_n to denote the collection of all Boolean functions defined on $\{0,1\}^n$.

2.1 Elementary symmetric polynomials

Recall that for $t \geq 1$, the elementary symmetric polynomial of degree t over \mathbb{F}_2 , $S_t(x_1, \dots, x_n)$, is defined as $S_t(x_1, \dots, x_n) = \bigoplus_{1 \leq a_1 < \dots < a_t \leq n} x_{a_1} \dots x_{a_t}$. Here \bigoplus denotes addition modulo two. This may be interpreted as

$$S_t(x_1, \dots, x_n) = \binom{|x|}{t} \bmod 2. \quad (2.1)$$

A direct consequence of Lucas's theorem (see, e.g., [Knu97, Section 1.2.6, Ex. 10]) and Eq. (2.1) is the following:

Lemma 2.1. *For every $\ell \geq 0$, $S_{2^\ell}(x) = |x|_\ell$. More generally, $S_t(x) = \prod_i |x|_i$ where the product runs over all $i \geq 0$ such that the $(i+1)$ th least significant bit of the binary expansion of t is 1.*

The following result follows from the work of Green and Tao [GT09, Theorem 11.3], who build upon the ideas of Alon and Beigel [AB01].

Theorem 2.2 (Green-Tao [GT09], Alon-Beigel [AB01]). *Fix $\ell \geq 0$. Then, for every multilinear polynomial $P \in \mathbb{F}_2[x_1, \dots, x_n]$ of degree at most $2^\ell - 1$, we have $\Pr_{x \sim \{0,1\}^n} [S_{2^\ell}(x) = P(x)] \leq 1/2 + o(1)$.*

Theorem 2.2 has a nice corollary:

²The constant in the $O(\cdot)$ is an absolute constant.

Corollary 2.3. For every fixed $\ell \geq 0$, the functions $\{S_{2^i}(x)\}_{0 \leq i \leq \ell}$ are almost balanced and almost uncorrelated, i.e.

- $\forall 0 \leq i \leq \ell, |\Pr[S_{2^i}(x) = 0] - \Pr[S_{2^i}(x) = 1]| = o(1)$
- $\forall a_0, \dots, a_\ell \in \{0, 1\}, |\Pr[\bigwedge_{0 \leq i \leq \ell} (S_{2^i}(x) = a_i)] - \frac{1}{2^{\ell+1}}| = o(1)$.

Combining [Corollary 2.3](#) with [Lemma 2.1](#), we get another useful fact:

Lemma 2.4. Let x be uniformly distributed over $\{0, 1\}^n$. Then, for every fixed $r \geq 1$, the random variables $\{|x|_i\}_{0 \leq i \leq r-1}$ are almost uniform and almost r -wise independent i.e.

- $\forall 0 \leq i \leq r-1, |\Pr[|x|_i = 0] - \Pr[|x|_i = 1]| = o(1)$.
- $\forall (a_0, \dots, a_{r-1}) \in \{0, 1\}^r, |\Pr[(|x|_0, \dots, |x|_{r-1}) = (a_0, \dots, a_{r-1})] - \frac{1}{2^r}| = o(1)$.

2.2 Boolean functions and polynomials over $\mathbb{Z}/2^k\mathbb{Z}$

Given an $F \in \mathcal{F}_n$ and $k \geq 1$, we define the k -lift of F to be the function $F_k : \{0, 1\}^n \rightarrow \mathbb{Z}/2^k\mathbb{Z}$ defined as follows. For any $x \in \{0, 1\}^n$,

$$F_k(x) = \begin{cases} 0 & \text{if } F(x) = 0, \\ 2^{k-1} & \text{otherwise.} \end{cases}$$

For $d \in \mathbb{N}$ and $k \geq 1$, $\mathcal{P}_{d,k}$ will denote the set of multilinear polynomials of degree at most d over the ring $\mathbb{Z}/2^k\mathbb{Z}$.

For functions $F, G : D \rightarrow R$ for some finite domain D and range R , the *agreement* between F and G , denoted by $\text{agr}(F, G)$, is defined to be the fraction of inputs where they agree, i.e.,

$$\text{agr}(F, G) = \Pr_{x \sim D}[F(x) = G(x)].$$

We will consider how well multilinear polynomials of degree d can approximate Boolean functions in the above sense. More precisely, for any Boolean function $F \in \mathcal{F}_n$, we define

$$\gamma_{d,k}(F) = \max_{Q \in \mathcal{P}_{d,k}} \text{agr}(F_k, Q).$$

Following [\[Gop08\]](#), we call a set $I \subseteq \{0, 1\}^n$ an *interpolating set*³ for $\mathcal{P}_{d,k}$ if the only polynomial $P \in \mathcal{P}_{d,k}$ that vanishes at all points in I is zero everywhere. Formally, for any $P \in \mathcal{P}_{d,k}$,

$$(\forall x \in I \ P(x) = 0) \Rightarrow (\forall y \in \{0, 1\}^n \ P(y) = 0).$$

We now state a number of standard facts regarding Boolean functions and multilinear polynomials over $\mathbb{Z}/2^k\mathbb{Z}$. The omitted proofs are either easy or well-known.

Unless mentioned otherwise, let n, d, k be any integers satisfying $n \geq 1, d \geq 0, k \geq 1$.

Lemma 2.5. Any polynomial $Q \in \mathcal{P}_{d,k}$ satisfies the following:

1. (Schwartz-Zippel) If Q is non-zero, then $\Pr_{x \sim \{0,1\}^n}[Q(x) \neq 0] \geq \frac{1}{2^d}$.

³This is also called a *hitting set* in the literature.

2. Q is the zero polynomial iff $Q(x) = 0$ for all $x \in \{0, 1\}^n$.
3. (Möbius Inversion) Say $Q(x) = \sum_{|S| \leq d} c_S x_S$, where $c_S \in \mathbb{Z}/2^k\mathbb{Z}$ and x_S denotes $\prod_{i \in S} x_i$. Then, $c_S = \sum_{T \subseteq S} (-1)^{|S|-|T|} Q(1_T)$ where $1_T \in \{0, 1\}^n$ is the characteristic vector of T .
4. ($\{0, 1\}_{\leq d}^n$ is an interpolating set) Q vanishes at all points in $\{0, 1\}^n$ iff Q vanishes at all points of $\{0, 1\}_{\leq d}^n$. By shifting the origin to any point of $\{0, 1\}^n$, the same is true of any Hamming ball of radius d in $\{0, 1\}^n$.

Proof. Point 1: Write Q as $Q(x) = 2^\ell \cdot Q'(x)$, where $\ell < k$ is the largest power of 2 that divides the GCD of the coefficients of Q . Projecting Q' to a non-zero polynomial over $\mathbb{Z}/2\mathbb{Z}$ by dropping all its coefficients modulo 2 and applying the standard Schwartz-Zippel lemma over $\mathbb{Z}/2\mathbb{Z}$ completes the proof.

Point 2 follows from point 1, and point 4 from point 3. □

Lemma 2.6. Fix any $F \in \mathcal{F}_n$.

1. $\gamma_{d,k}(F) \geq \frac{1}{2}$.
2. $\gamma_{d,k+1}(F) \geq \gamma_{d,k}(F)$.
3. $\gamma_{d,k}(F) > 1 - \frac{1}{2^d} \Rightarrow \gamma_{d,k}(F) = \gamma_{d,1}(F)$.
4. $\gamma_{1,k}(F) = \gamma_{1,1}(F)$.

Proof. Point 1 is trivial since there is a constant polynomial that has agreement at least $\frac{1}{2}$ with F_k .

Point 2: Say $P \in \mathcal{P}_{d,k}$ has agreement α with F_k . Then, $2 \cdot P$ (interpreted naturally as a polynomial in $\mathcal{P}_{d,k+1}$) has agreement α with F_{k+1} .

For point 3, consider a polynomial $Q \in \mathcal{P}_{d,k}$ that achieves the maximum agreement $\alpha > 1 - \frac{1}{2^d}$ with F_k . Let $Q' \in \mathcal{P}_{d,1}$ be the polynomial obtained from Q by dropping all its co-efficients modulo 2^{k-1} . Note that for any x , $Q(x) \in \{0, 2^{k-1}\}$ implies that $Q'(x) = 0$ (in the ring $\mathbb{Z}/2^{k-1}\mathbb{Z}$). Hence, the probability that Q' is zero is at least $\alpha > 1 - \frac{1}{2^d}$. [Lemma 2.5](#) point 1 implies that Q' must be the zero polynomial. Equivalently, all of the coefficients of Q are divisible by 2^{k-1} and hence Q can be naturally identified with $2^{k-1} \cdot Q''$ for some $Q'' \in \mathbb{Z}/2\mathbb{Z}[x_1, \dots, x_n]$. It is easy to check that $\text{agr}(Q'', F_1) = \alpha$ and hence we have $\gamma_{d,1}(F) \geq \gamma_{d,k}(F)$. On the other hand, from point 2, we already know that $\gamma_{d,k}(F) \leq \gamma_{d,1}(F)$. Hence we are done.

Point 4 follows from points 1 and 3. □

3 Some separation results

3.1 Symmetric functions as separating examples

We know from [Theorem 2.2](#) that, for every fixed $\ell \geq 2$, $\gamma_{2^\ell-1,1}(S_{2^\ell}) \leq \frac{1}{2} + o(1)$. In contrast, the main result of this section shows that

Theorem 3.1. For every fixed $\ell \geq 2$, $\gamma_{d,3}(S_{2^\ell}) \geq \frac{9}{16} - o(1)$, where $d = 2^{\ell-1} + 2^{\ell-2}$.

Notice that $2^{\ell-1} + 2^{\ell-2} \leq 2^\ell - 1$ for $\ell \geq 2$. This implies that, for $\ell \geq 2$, $S_{2^\ell}(x)$ is an example of a function F for which there exist $k, d \in \mathbb{N}$ such that $\gamma_{d,1}(F) \leq \frac{1}{2} + o(1)$ but $\gamma_{d',k}(F) \geq \frac{1}{2} + \Omega(1)$ for some $d' \leq d$.

Proof of Theorem 3.1. Lemma 2.1 from Section 2 tells us that $S_{2^\ell}(x) = |x|_\ell$. Thus, $S_{2^\ell,3}(x) \in \mathbb{Z}/8\mathbb{Z}[x_1, \dots, x_n]$, the 3-lift of $S_{2^\ell}(x)$, is given by

$$S_{2^\ell,3}(x) = \begin{cases} 4 & \text{if } |x|_\ell = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

Fix d to be $2^{\ell-1} + 2^{\ell-2}$. Consider the polynomial $P(x) = \sum_{T \in \binom{[n]}{d}} \prod_{i \in T} x_i$ in $\mathbb{Z}/8\mathbb{Z}[x_1, \dots, x_n]$. To prove the theorem, it suffices to show that

$$\Pr_{x \sim \{0,1\}^n} [P(x) = S_{2^\ell,3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

Clearly, $P(x) = \binom{|x|}{d} \pmod{8}$, and

$$P(x) = \begin{cases} 0 & \text{if } 8 \mid \binom{|x|}{d} \\ 4 & \text{if } 4 \mid \binom{|x|}{d} \text{ but } 8 \nmid \binom{|x|}{d} \end{cases} \quad (3.2)$$

The following theorem due to Kummer (see, e.g., [Knu97, Section 1.2.6, Ex. 11]) determines the largest power of a prime that divides a binomial coefficient.

Theorem 3.2 (Kummer). *Let p be a prime and $N, M \in \mathbb{N}$ such that $N \geq M$. Suppose r is the largest integer such that $p^r \mid \binom{N}{M}$. Then r is equal to the number of borrows required when subtracting M from N in base p .*

Let $B(x)$ be the number of borrows required when subtracting d from $|x|$. Rewriting Eq. (3.2) in terms of $B(x)$ using Kummer's theorem, we get

$$P(x) = \begin{cases} 4 & \text{if } B(x) = 2 \\ 0 & \text{if } B(x) \geq 3 \end{cases} \quad (3.3)$$

We will need the following lemma.

Lemma 3.3. $P(x) = S_{2^\ell,3}(x)$ if either

1. $|x|_{\ell-2} = 0$, or
2. $(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1, 0, 0, 0)$.

Proof. Since $d = 2^{\ell-1} + 2^{\ell-2}$, all the bits of d except $d_{\ell-1}$ and $d_{\ell-2}$ are zero. Thus, when subtracting d from $|x|$, no borrows are required by the bits $|x|_i$, $0 \leq i \leq \ell - 3$.

Using the above observation, the reader can verify that when $(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1, 0, 0, 0)$ the number of borrows required is at least 3 i.e. $B(x) \geq 3$, which in turn implies that $P(x) = 0$. Since $|x|_\ell = 0$, $S_{2^\ell,3}(x) = 0$. This proves the second part of the lemma.

To prove the first part, suppose $|x|_{\ell-2} = 0$. Since $d_{\ell-1} = d_{\ell-2} = 1$, it follows that both $|x|_{\ell-2}$ and $|x|_{\ell-1}$ will need to borrow when subtracting d from $|x|$. As argued before, no borrows are required by the bits before (i.e. less significant than) $|x|_{\ell-2}$, and thus the total number of borrows required by the bits $|x|_i, 0 \leq i \leq \ell-1$, is 2.

Note that the bit $|x|_{\ell-1}$ borrows from $|x|_{\ell}$. Consider the following case analysis:

- Case $|x|_{\ell} = 1$: $|x|_{\ell}$ will not need to borrow since $d_{\ell} = 0$. In fact, none of the bits after (i.e. more significant than) $|x|_{\ell}$ will need to borrow, and thus $B(x) = 2$. This implies that $P(x) = 4$. We also have $S_{2^{\ell},3}(x) = 4$ and hence $P(x) = S_{2^{\ell},3}(x)$.
- Case $|x|_{\ell} = 0$: $|x|_{\ell}$ will require a borrow and this means $B(x) \geq 3$. This implies that $P(x) = 0$. Since $|x|_{\ell} = 0$, it follows that $P(x) = S_{2^{\ell},3}(x)$.

This completes the proof. □

By [Lemma 3.3](#), we have

$$\Pr[P(x) = S_{2^{\ell},3}(x)] \geq \Pr[|x|_{\ell-2} = 0] + \Pr[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_{\ell}, |x|_{\ell+1}) = (1, 0, 0, 0)] \quad (3.4)$$

Using [Lemma 2.4](#) from [Section 2](#), we have

$$\begin{aligned} \Pr[|x|_{\ell-2} = 0] &\geq \frac{1}{2} - o(1) \\ \Pr[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_{\ell}, |x|_{\ell+1}) = (1, 0, 0, 0)] &\geq \frac{1}{16} - o(1) \end{aligned}$$

which, together with [Eq. \(3.4\)](#), implies

$$\Pr[P(x) = S_{2^{\ell},3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

□

3.2 A separation at $k = 2$

Let $d \in \mathbb{N}$ be any power of 2. In this section, we show that there are functions F for which $\gamma_{2d,2}(F) > \gamma_{3d-1,1}(F)$.

Theorem 3.4. *For large enough n , there exists a function $F \in \mathcal{F}_{2n}$ such that $\gamma_{2d,2}(F) \geq \frac{3}{4} - o(1)$ but $\gamma_{3d-1,1}(F) \leq \frac{5}{8} + o(1)$.*

In particular, we see that $\gamma_{2,2}(F) > \gamma_{2,1}(F)$. This result is notable, since it shows that there is a separation at the first place where it is possible to have one (Recall that $\gamma_{1,k}(F) = \gamma_{1,1}(F)$ for any $F \in \mathcal{F}_n$ by [Lemma 2.6](#)).

Let us begin the proof of [Theorem 3.4](#). We first define a family of Boolean functions on $\{0, 1\}^{2n}$. We denote the $2n$ variables by x_1, \dots, x_n and y_1, \dots, y_n . We use $\binom{|x|}{d}$ to denote the d th elementary symmetric polynomial from the ring $\mathbb{Z}/4\mathbb{Z}[x_1, \dots, x_n]$, i.e., $\binom{|x|}{d} = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} x_i$.⁴

We will need the following easy corollary of [Theorem 3.2](#).

⁴We distinguish between $\binom{|x|}{d}$ and $S_d(x)$ since the former is from $\mathbb{Z}/4\mathbb{Z}[x_1, \dots, x_n]$ and latter a polynomial in $\mathbb{F}_2[x_1, \dots, x_n]$.

Corollary 3.5. *Let d be a power of 2. Then, for $N \geq d$, the highest power of 2 dividing $\binom{N}{d}$ is equal to the highest power of 2 dividing $\lfloor \frac{N}{d} \rfloor$.*

Let $S = \{(x, y) \mid \binom{|x|}{d}, \binom{|y|}{d} \equiv 1 \pmod{2}\}$. Given any function $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$, we define the Boolean function $F_H(x_1, \dots, x_n, y_1, \dots, y_n)$ as follows:

$$F_H(x, y) = \begin{cases} 0 & \text{if } \binom{|x|}{d} \cdot \binom{|y|}{d} \equiv 0 \pmod{4}, \\ 1 & \text{if } \binom{|x|}{d} \cdot \binom{|y|}{d} \equiv 2 \pmod{4}, \\ H(x, y) & \text{otherwise.} \end{cases}$$

Define $P(x, y) = \binom{|x|}{d} \cdot \binom{|y|}{d} \in \mathbb{Z}/4\mathbb{Z}[x_1, \dots, x_n, y_1, \dots, y_n]$. Note that $F_H(x, y)$ is defined so that its 2-lift agrees with $P(x, y)$ on points (x, y) where $P(x, y) \in \{0, 2\}$. Also [Corollary 3.5](#) implies that the following is an alternate equivalent definition of F_H in terms of elementary symmetric polynomials modulo 2.

$$F_H(x, y) = \begin{cases} 0 & \text{if } S_d(x) = S_d(y) = 0, \\ S_{2d}(y) & \text{if } S_d(x) = 1 \text{ and } S_d(y) = 0, \\ S_{2d}(x) & \text{if } S_d(x) = 0 \text{ and } S_d(y) = 1, \\ H(x, y) & \text{otherwise.} \end{cases} \quad (3.5)$$

We now begin the proof of [Theorem 3.4](#). First of all, let us note that for any choice of H , we have:

Lemma 3.6. $\gamma_{2d,2}(F_H) \geq \frac{3}{4} - o(1)$.

Proof. Consider the polynomial $P(x, y) \in \mathcal{P}_{2d,2}$ defined above. From [Eq. \(3.5\)](#), it follows that the probability that $P(x, y) \neq F_{H,2}(x, y)$ ⁵ is less than or equal to the probability that $S_d(x) = S_d(y) = 1$, which is $\frac{1}{4} + o(1)$ by [Corollary 2.3](#). This gives the claim. \square

The main lemma is the following.

Lemma 3.7. *Say $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$ is chosen uniformly at random. Then,*

$$\Pr_H[\gamma_{3d-1,1}(F_H) > \frac{5}{8} + o(1)] = o(1).$$

This will prove [Theorem 3.4](#). We will prove the above lemma in the following subsection.

3.3 Proof of [Lemma 3.7](#)

The outline of the proof is as follows. Fix any polynomial $Q \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n]$ of degree at most $3d - 1$. We need to show that $\text{agr}(F_H, Q) \leq \frac{5}{8} + o(1)$ for a random $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$. The fact that H is random ensures that any Q cannot agree with H on significantly more than half the inputs in S . For inputs outside S , we need a more involved argument, following Alon and Beigel [[AB01](#)]. We show that for any Q we can find somewhat large sets I and J of x and y variables respectively such that when we set the variables outside $I \cup J$, we obtain a polynomial that is symmetric in the variables of $I \cup J$. This is a Ramsey theoretic argument á la Alon-Beigel [[AB01](#)].

⁵ $F_{H,2}$ denotes the 2-lift of F_H .

Following this argument, we only need to prove the agreement upper bound for Q that is symmetric in x and y variables. This can be done by reduction to a constant-sized problem, as we show below. A careful computation to solve the constant-sized problem will finish the proof.

We begin with some notation that will be useful in the proof. Throughout, we work with disjoint sets of x -variables and y -variables of equal size and consider polynomials over these variables. Let the x -variables be $\{x_1, \dots, x_n\}$ and the y -variables be $\{y_1, \dots, y_n\}$. For $I, J \subseteq [n]$, the set of \mathbb{F}_2 -polynomials Q over the variables $\{x_i \mid i \in I\}$ and $\{y_j \mid j \in J\}$ is denoted $\mathbb{F}_2[x_I, y_J]$. Similarly, we use $Q \in \mathbb{F}_2[x_I]$ to denote the fact that Q is a polynomial only over the variables $\{x_i \mid i \in I\}$.

We use $\mathcal{A}_{I,J}$ to denote Boolean assignments $\sigma : \{x_i \mid i \notin I\} \cup \{y_j \mid j \notin J\} \rightarrow \{0, 1\}$. Given $F : \mathbb{F}_2^{2n} \rightarrow \mathbb{F}_2$ and $\sigma \in \mathcal{A}_{I,J}$, we use $F|_\sigma \in \mathbb{F}_2[x_I, y_J]$ to denote its natural restriction to the variables indexed by $I \cup J$.

We say that $Q \in \mathbb{F}_2[x_I, y_J]$ is (x, y) -symmetric if it is a linear combination of the polynomials in the set $\{S_{d_1}(x_I) \cdot S_{d_2}(y_J) \mid d_1, d_2 \in \mathbb{N}\}$. We note that being (x, y) -symmetric depends on the sets I, J under consideration. This will be implicit when used.

Given a multilinear monomial m over the x and y -variables, its *multidegree* is defined to be (i, j) if m multiplies i x -variables and j y -variables. Let $\mathcal{D} = \{(i, j) \mid i + j \leq 3d - 1\}$ be the set of multidegrees of monomials of degree at most $3d - 1$. We order \mathcal{D} in ascending order according to $i + j$, i.e., fix a total ordering \preceq of \mathcal{D} such that if $i_1 + j_1 < i_2 + j_2$, then $(i_1, j_1) \preceq (i_2, j_2)$ ⁶. Let (i_0, j_0) be the largest element in the ordering \preceq . We will define $\text{multdeg}(Q)$ to be the largest (w.r.t. \preceq) multidegree of a monomial that has a non-zero coefficient in Q . For (i, j) such that $i + j \leq 3d - 1$, we say that a polynomial $Q \in \mathbb{F}_2[x_I, y_J]$ is $(x, y; i, j)$ -symmetric if we can write Q as

$$Q = Q_1 \oplus Q_2 \tag{3.6}$$

where $\text{multdeg}(Q_1) \preceq (i, j)$ and Q_2 is (x, y) -symmetric. Note that if $(i, j) = (i_0, j_0)$, then any polynomial of degree at most $3d - 1$ is $(x, y; i_0, j_0)$ -symmetric, since we can take $Q_1 = Q$ and $Q_2 = 0$.

We also need the following variant of the function F_H defined above. Call a function $\Phi \in \mathbb{F}_2[x_1, \dots, x_n]$ (resp. $\Psi \in \mathbb{F}_2[y_1, \dots, y_n]$) d -simple w.r.t. x (resp. w.r.t. y) if it is a linear combination of symmetric polynomials in x (resp. in y) of degree *strictly* less than d . Equivalently, we can say that $\Phi(x)$ only depends upon $|x|_0, \dots, |x|_{\lg d - 1}$, and similarly for $\Psi(y)$ w.r.t. y .

Given pairs of polynomials $\Phi = (\Phi_1, \Phi_2) \in \mathbb{F}_2[x_1, \dots, x_n] \times \mathbb{F}_2[x_1, \dots, x_n]$, and $\Psi = (\Psi_1, \Psi_2) \in \mathbb{F}_2[y_1, \dots, y_n] \times \mathbb{F}_2[y_1, \dots, y_n]$, such that Φ_1, Φ_2 and Ψ_1, Ψ_2 are d -simple w.r.t. x and y respectively, define

$$F_{H, \Phi, \Psi}(x, y) = \begin{cases} 0 & \text{if } S_d(x) = \Phi_1(x), S_d(y) = \Psi_1(y), \\ S_{2d}(y) \oplus \Psi_2(y) & \text{if } S_d(x) = 1 \oplus \Phi_1(x) \text{ and } S_d(y) = \Psi_1(y), \\ S_{2d}(x) \oplus \Phi_2(x) & \text{if } S_d(x) = \Phi_1(x) \text{ and } S_d(y) = 1 \oplus \Psi_1(y), \\ H(x, y) & \text{otherwise.} \end{cases} \tag{3.7}$$

Also, define $S_{\Phi, \Psi} = \{(x, y) \mid S_d(x) = 1 \oplus \Phi_1(x), S_d(y) = 1 \oplus \Psi_1(y)\}$.

With the notation above, we are ready to state a claim that generalizes [Lemma 3.7](#).

Lemma 3.8. *Fix any $d \in \mathbb{N}$. For $(i, j) \in \mathcal{D}$ and $\varepsilon \in (0, 1)$, there is an $n(i, j, \varepsilon)$ such that given any $n \geq n(i, j, \varepsilon)$, for uniformly random $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$, we have for any $\Phi = (\Phi_1, \Phi_2) \in \mathbb{F}_2[x_1, \dots, x_n]^2$*

⁶If $(i_1, j_1) \neq (i_2, j_2)$, but $i_1 + j_1 = i_2 + j_2$, then the relation between (i_1, j_1) and (i_2, j_2) is fixed in an arbitrary manner.

and $\Psi = (\Psi_1, \Psi_2) \in \mathbb{F}_2[y_1, \dots, y_n]^2$ such that Φ_1, Φ_2 are d -simple w.r.t. x , and Ψ_1, Ψ_2 are d -simple w.r.t. y ,

$$\Pr_H[\exists Q \text{ of degree } \leq 3d - 1 \text{ that is } (x, y; i, j)\text{-symmetric s.t. } \text{agr}(F_{H, \Phi, \Psi}, Q) \geq \frac{5}{8} + \varepsilon] \leq \varepsilon. \quad (3.8)$$

The statement of the above lemma for $\Phi = \Psi = (0, 0)$ and $(i, j) = (i_0, j_0)$ implies [Lemma 3.7](#) since in this case $F_{H, \Phi, \Psi} = F_H$ and as noted above, any polynomial Q of degree at most $3d - 1$ is $(x, y; i_0, j_0)$ -symmetric.

The proof of [Lemma 3.8](#) is by induction on the order \preceq . The base case is the case when $(i, j) = (0, 0)$, the minimal element of the ordering \preceq .

Throughout, the parameter d is a fixed integer power of 2.

3.3.1 Base case of the induction: $i = j = 0$

In this case, by [Eq. \(3.6\)](#), it is clear that Q is an (x, y) -symmetric polynomial. We show in this case that bounding $\text{agr}(F_{H, \Phi, \Psi}, Q)$ reduces (for most H) to bounding the correlations between functions on 5 inputs. A simple computation solves this problem.

Fix any bits $\phi, \psi \in \{0, 1\}^2$. Define the Boolean function f on 5 variables $a_1, b_1, a_2, b_2, z \in \mathbb{F}_2$ as follows. Notice the similarity to [Eqs. \(3.5\)](#) and [\(3.7\)](#).

$$f_{\phi, \psi}(a_1, a_2, b_1, b_2, z) = \begin{cases} 0 & \text{if } a_1 = \phi_1, b_1 = \psi_1, \\ b_2 \oplus \psi_2 & \text{if } a_1 = 1 \oplus \phi_1 \text{ and } b_1 = \psi_1, \\ a_2 \oplus \phi_2 & \text{if } a_1 = \phi_1 \text{ and } b_1 = 1 \oplus \psi_1, \\ z & \text{otherwise.} \end{cases} \quad (3.9)$$

Call a polynomial $q \in \mathbb{F}_2[a_1, a_2, b_1, b_2, z]$ *relevant* if q is a linear combination of monomials from the set $\{1, a_1, a_2, b_1, b_2, a_1 b_1\}$. Let \mathcal{R} denote the set of relevant polynomials. Note that relevant polynomials do not involve the variable z .

Given $H, Q' \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n]$ and Φ, Ψ as in the statement of [Lemma 3.8](#), we define $\text{agr}_S(H, Q')$ to be $\Pr_{(x, y) \in S_{\Phi, \Psi}}[H(x, y) = Q'(x, y)]$.

We say that $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$ is ε -hard if for any $Q' \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n]$ of degree at most $3d - 1$, we have $|\text{agr}_S(H, Q') - \frac{1}{2}| \leq \varepsilon$.

We need the following property of a random $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$.

Lemma 3.9. *For any $\varepsilon > 0$, there is an $n_0(\varepsilon) \in \mathbb{N}$ such that if $n \geq n_0(\varepsilon)$, then for $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$ chosen uniformly at random $\Pr_H[H \text{ not } \varepsilon\text{-hard}] \leq \varepsilon$.*

Proof. The proof is a trivial union bound. The number of polynomials Q' of degree at most $3d - 1$ is at most $2^{\sum_{i=0}^{3d-1} \binom{2n}{i}}$ (there are $\sum_{i=0}^{3d-1} \binom{2n}{i} \leq (2n)^{3d}$ many possible monomials each has 2 possible coefficients). For each such Q' , the expected number of locations $x \in S_{\Phi, \Psi}$ where $H(x) \neq Q'(x)$ is $|S_{\Phi, \Psi}|/2$. By a Chernoff bound, the probability that this number is not in the range $[|S_{\Phi, \Psi}|/2 - \varepsilon|S_{\Phi, \Psi}|, |S_{\Phi, \Psi}|/2 + \varepsilon|S_{\Phi, \Psi}|]$ is $\exp(-\Omega(\varepsilon^2|S_{\Phi, \Psi}|))$. By [Corollary 2.3](#), it follows that $|S_{\Phi, \Psi}| = \Omega(2^{2n})$, and hence the above probability can be upper bounded by $\exp(-\Omega(\varepsilon^2 2^{2n}))$. A union bound over all the possible Q' tells us that with probability $1 - \exp(-(2n)^{3d} - \Omega(\varepsilon^2 2^{2n}))$ over the choice of H , every Q' of degree at most $3d - 1$ satisfies $\text{agr}_S(H, Q') \in [\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon]$. In particular, for any ε , a large enough n will ensure that the probability that H is not ε -hard is at most ε . \square

We will prove the following lemmas.

Lemma 3.10. Fix any Φ, Ψ as in the statement of [Lemma 3.8](#). For any $\varepsilon > 0$, there is an $n(0, 0, \varepsilon) \in \mathbb{N}$ such that for any $n \geq n(0, 0, \varepsilon)$

$$\Pr_H[\exists Q \text{ of degree } \leq 3d - 1 \text{ that is } (x, y)\text{-symmetric s.t. } \text{agr}(F_{H, \Phi, \Psi}, Q) > \max_{q \in \mathcal{R}, \phi, \psi} \text{agr}(f_{\phi, \psi}, q) + \varepsilon] \leq \varepsilon.$$

Lemma 3.11. $\max_{q \in \mathcal{R}, \phi, \psi} \text{agr}(f_{\phi, \psi}, q) \leq \frac{5}{8}$.

The above lemmas clearly prove [Eq. \(3.8\)](#) in the case $i = j = 0$, which completes the base case.

Proof of [Lemma 3.10](#). We choose $n(0, 0, \varepsilon)$ during the course of the proof. First of all, we will assume that $n(0, 0, \varepsilon) \geq n_0(\varepsilon/2)$, so that we have

$$\Pr_H[H \text{ not } \varepsilon/2\text{-hard}] \leq \varepsilon/2. \quad (3.10)$$

We now show that when H is $\varepsilon/2$ -hard, then for any Q that is (x, y) -symmetric of degree at most $3d - 1$, we have

$$\text{agr}(F_{H, \Phi, \Psi}, Q) \leq \max_{q \in \mathcal{R}, \phi, \psi} \text{agr}(f_{\phi, \psi}, q) + \varepsilon. \quad (3.11)$$

This will prove the lemma. Fix any $\varepsilon/2$ -hard H for the remainder of the lemma.

Since Q is (x, y) -symmetric, it follows that we can write

$$Q = \bigoplus_{d_1, d_2: d_1 + d_2 \leq 3d - 1} \gamma_{d_1, d_2} S_{d_1}(x) \cdot S_{d_2}(y) \quad (3.12)$$

for some choice of the γ_{d_1, d_2} s from \mathbb{F}_2 .

Let z be a new variable taking values in \mathbb{F}_2 . We now define $F \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n, z]$ as follows. We set $F(x, y) = F_{H, \Phi, \Psi}(x, y)$ for all $(x, y) \notin S_{\Phi, \Psi}$ and $F(x, y) = z$ for all $(x, y) \in S_{\Phi, \Psi}$. Note that we have $\text{agr}_S(F, Q) = \frac{1}{2}$ since Q does not depend on the random variable z . Further, since H is $\varepsilon/2$ -hard, we know that $|\text{agr}_S(F_{H, \Phi, \Psi}, Q) - \frac{1}{2}| \leq \varepsilon/2$.

In particular, we see that $|\text{agr}_S(F_{H, \Phi, \Psi}, Q) - \text{agr}_S(F, Q)| \leq \varepsilon/2$. Since F and $F_{H, \Phi, \Psi}$ agree outside $S_{\Phi, \Psi}$, this implies that

$$|\text{agr}(F_{H, \Phi, \Psi}, Q) - \text{agr}(F, Q)| \leq \varepsilon/2. \quad (3.13)$$

So to upper bound $\text{agr}(F_{H, \Phi, \Psi}, Q)$, we upper bound $\text{agr}(F, Q)$. Assume $d = 2^\ell$.

Consider Q . By [Eq. \(3.12\)](#), we have

$$Q(x, y) = \bigoplus_{d_1, d_2: d_1 + d_2 \leq 3d - 1} \gamma_{d_1, d_2} S_{d_1}(x) \cdot S_{d_2}(y) = \bigoplus_{\substack{A, B \subseteq \{0, \dots, \ell + 1\}: \\ \sum_{i \in A} 2^i + \sum_{j \in B} 2^j \leq 3d - 1}} \gamma_{A, B} \prod_{i \in A} |x|_i \cdot \prod_{j \in B} |y|_j \quad (3.14)$$

where the $\gamma_{A, B}$ s are in \mathbb{F}_2 , $|x|_0, \dots, |x|_{\ell + 1}$ being the $\ell + 2$ least significant bits of $|x|$ (and similarly for y) and we have used [Lemma 2.1](#) for the final equality above.

Now consider F . By the definition of F above, [Eq. \(3.7\)](#), and once again using [Lemma 2.1](#), we have

$$F(x, y, z) = \begin{cases} 0 & \text{if } |x|_\ell = \Phi_1(x), |y|_\ell = \Psi_1(y), \\ |y|_{\ell + 1} \oplus \Psi_2(y) & \text{if } |x|_\ell = 1 \oplus \Phi_1(x) \text{ and } |y|_\ell = \Psi_1(y), \\ |x|_{\ell + 1} \oplus \Phi_2(y) & \text{if } |x|_\ell = \Phi_1(x) \text{ and } |y|_\ell = 1 \oplus \Psi_1(y), \\ z & \text{otherwise,} \end{cases} \quad (3.15)$$

where Φ_1, Φ_2 , being d -simple, are functions of $|x|_0, \dots, |x|_{\ell-1}$, and similarly, Ψ_1, Ψ_2 are functions of $|y|_0, \dots, |y|_{\ell-1}$.

Let $\alpha_0, \dots, \alpha_{\ell+1}, \beta_0, \dots, \beta_{\ell+1}$ be $2(\ell+2)$ new variables. Define $q' \in \mathbb{F}_2[\alpha_0, \dots, \alpha_{\ell+1}, \beta_0, \dots, \beta_{\ell+1}]$ by replacing $|x|_i$ by α_i and $|y|_j$ by β_j in Eq. (3.14) above. That is,

$$q'(\alpha_0, \dots, \alpha_{\ell+1}, \beta_0, \dots, \beta_{\ell+1}) = \bigoplus_{\substack{A, B \subseteq \{0, \dots, \ell+1\}: \\ \sum_{i \in A} 2^i + \sum_{j \in B} 2^j \leq 3d-1}} \gamma_{A,B} \prod_{i \in A} \alpha_i \cdot \prod_{j \in B} \beta_j \quad (3.16)$$

and similarly define $f' \in \mathbb{F}_2[\alpha_0, \dots, \alpha_{\ell+1}, \beta_0, \dots, \beta_{\ell+1}, z]$ by replacing $|x|_i$ by α_i and $|y|_j$ by β_j for each $i, j \in \{0, \dots, \ell+1\}$ in the definition of F above. We have

$$\begin{aligned} \text{agr}(F, Q) &= \Pr_{x,y,z} [F(x, y, z) = Q(x, y)] \\ &= \Pr_{x,y,z} [f'(|x|_0, \dots, |x|_{\ell+1}, |y|_0, \dots, |y|_{\ell+1}, z) = q'(|x|_0, \dots, |x|_{\ell+1}, |y|_0, \dots, |y|_{\ell+1})] \end{aligned}$$

By Lemma 2.4, we know that if $n(0, 0, \varepsilon)$ is large enough, then for uniformly random $x, y \in \mathbb{F}_2^n$, the tuples $(|x|_0, \dots, |x|_{\ell+1})$ and $(|y|_0, \dots, |y|_{\ell+1})$ are $\varepsilon/4$ -close to the uniform distribution (in statistical distance) over $\mathbb{F}_2^{\ell+2}$. Note also that x, y, z are mutually independent. From this, it easily follows that the final expression in the above display is $\frac{\varepsilon}{2}$ -close to $\text{agr}(f', q')$. Thus, we get

$$|\text{agr}(F, Q) - \text{agr}(f', q')| \leq \frac{\varepsilon}{2}. \quad (3.17)$$

Therefore, we analyze $\text{agr}(f', q')$. Conditioning on any setting τ of $\alpha_0, \dots, \alpha_{\ell-1}, \beta_0, \dots, \beta_{\ell-1}$, we see that the functions $\Phi_1, \Phi_2, \Psi_1, \Psi_2$ (being d -simple) are fixed to some constants in \mathbb{F}_2 and hence f' simplifies to a polynomial $f''(\alpha_\ell, \alpha_{\ell+1}, \beta_\ell, \beta_{\ell+1})$. Similarly, q' simplifies to some $q''(\alpha_\ell, \alpha_{\ell+1}, \beta_\ell, \beta_{\ell+1})$. Further, note that by the constraints on sets A and B in Eq. (3.16), q'' must be a linear combination of monomials from the set $\{1, \alpha_\ell, \alpha_{\ell+1}, \beta_\ell, \beta_{\ell+1}, \alpha_\ell \beta_\ell\}$. Renaming variables $\alpha_\ell, \alpha_{\ell+1}, \beta_\ell, \beta_{\ell+1}$ to a_1, a_2, b_1, b_2 respectively, we see that $\text{agr}(f'', q'') \leq \max_{q \in \mathcal{R}, \phi, \psi \in \mathbb{F}_2^2} \text{agr}(f_{\phi, \psi}, q)$. Since this is true of any τ , the same upper bound holds for $\text{agr}(f', q')$ as well.

Combined with Eq. (3.17) and Eq. (3.13), this yields $\text{agr}(F_{H, \Phi, \Psi}, Q) \leq \max_{q \in \mathcal{R}, \phi, \psi} \text{agr}(f_{\phi, \psi}, q) + \varepsilon$. Since this is true for every $\varepsilon/2$ -hard function H , and the probability that a random H is $(\varepsilon/2)$ -hard is at least $1 - \varepsilon/2$, we are done. \square

Proof of Lemma 3.11. We prove the statement by a simple case analysis.

The first case is that the relevant polynomial $q \in \mathcal{R}$ depends on at least one among $\{a_2, b_2\}$. Without loss of generality, we assume that q depends on a_2 . Then, by the definition of \mathcal{R} , we can write $q = a_2 \oplus q'$ where $q' \in \mathbb{F}_2[a_1, b_1, b_2]$. Consider any setting of (a_1, b_1, b_2, z) such that $(a_1, b_1) \in \{(\phi_1, \psi_1), (1 \oplus \phi_1, \psi_1), (1 \oplus \phi_1, 1 \oplus \psi_1)\}$. Under this restriction, $f_{\phi, \psi}$ is a constant function whereas q is a non-constant linear function depending on a_2 . Hence, when $(a_1, b_1) \in \{(\phi_1, \psi_1), (1 \oplus \phi_1, \psi_1), (1 \oplus \phi_1, 1 \oplus \psi_1)\}$, $f_{\phi, \psi}$ and q can agree on at most half the inputs. Thus we get that $\text{agr}(f_{\phi, \psi}, q) \leq \frac{5}{8}$.

The second case is that q depends on neither a_2 nor b_2 . In this case, consider any setting of $(a_1, b_1) \in \{(\phi_1, 1 \oplus \psi_1), (1 \oplus \phi_1, \psi_1), (1 \oplus \phi_1, 1 \oplus \psi_1)\}$. Under each of these restrictions, q computes the constant function (recall that q does not depend on z) whereas $f_{\phi, \psi}$ is a non-constant linear function. Thus, as before, we get that $\text{agr}(f_{\phi, \psi}, q) \leq \frac{5}{8}$. This proves the lemma. \square

3.3.2 The induction case

We now induct. Let $(i, j) \in \mathcal{D}$ be non-minimal and let (i', j') be its predecessor w.r.t. \preceq . Assume [Lemma 3.8](#) for $(x, y; i', j')$ -symmetric polynomials. We now prove it for $(x, y; i, j)$ -symmetric polynomials.

We will need the following basic Ramsey-theoretic statement. It is a straightforward generalization (to hypergraphs) of the fact that any large enough bipartite graph contains large bipartite independent sets or complete bipartite subgraphs. Unfortunately we could not find exactly this statement in the literature, so we provide a proof of the statement in [Appendix A](#).

Let I and J be disjoint sets of size n each. A function $c : \binom{I}{i} \times \binom{J}{j} \rightarrow \{0, 1\}$ is said to be an (i, j) -colouring of (I, J) . (Recall that $\binom{I}{i}$ denotes the collection of all i -sized subsets of I .)

Lemma 3.12. *For any $i, j \in \mathbb{N}$ and any $r \in \mathbb{N}$, there is an $n_R(i, j, r) \geq 2^r \in \mathbb{N}$ such that for any $n \geq n_R(i, j, r)$, any disjoint n -sets I, J and any (i, j) -colouring c of (I, J) , there are sets $I' \subseteq I$ and $J' \subseteq J$ with $|I'| = |J'| = r$ such that the restriction c' of c to $\binom{I'}{i} \times \binom{J'}{j}$ is a constant function.*

We now prove the inductive case of [Lemma 3.8](#). Let $Q \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n]$ be $(x, y; i, j)$ -symmetric. By [Eq. \(3.6\)](#), we have

$$Q = Q_1 \oplus Q_2 = Q'_1 \oplus Q_{i,j} \oplus Q_2 \quad (3.18)$$

where Q_2 is (x, y) -symmetric, Q_1 has multidegree at most (i, j) , $Q_{i,j}$ is the part of Q_1 of multidegree exactly (i, j) , and Q'_1 is the part of multidegree strictly less than (i, j) (i.e. at most (i', j')).

Use $Q_{i,j}$ to define an (i, j) -colouring c of $([n], [n])$ as follows. For $A \in \binom{[n]}{i}$ and $B \in \binom{[n]}{j}$, we define $c(A, B)$ to be the coefficient of the monomial $\prod_{s \in A} x_s \cdot \prod_{t \in B} y_t$ in $Q_{i,j}$. Applying [Lemma 3.12](#) with $r = n(i', j', \varepsilon/4)$, we see that if $n \geq n_R(i, j, r)$, then there are $I, J \in \binom{[n]}{r}$ and $\alpha \in \mathbb{F}_2$ such that for all $A \in \binom{I}{i}$ and $B \in \binom{J}{j}$, we have $c(A, B) = \alpha$.

Assume that Q is as in [Eq. \(3.18\)](#) and $n \geq n_R(i, j, r)$. We find I, J as above. For any setting $\sigma \in \mathcal{A}_{I, J}$, we can write the polynomial $Q_{i,j}|_\sigma$ as $Q'_{i,j} \oplus Q''_{i,j}$ where $Q'_{i,j}$ is the part of degree $i + j$, and $Q''_{i,j}$ has degree strictly less than $i + j$.

Observe that

$$Q'_{i,j} = \bigoplus_{A \in \binom{I}{i}, B \in \binom{J}{j}} c(A, B) \prod_{s \in A} x_s \prod_{t \in B} y_t = \alpha \cdot S_i(x_I) S_j(y_J),$$

and is an (x, y) -symmetric polynomial (on the remaining variables x_I, y_J). Hence, by [Eq. \(3.18\)](#), we get

$$Q|_\sigma = Q'_1|_\sigma \oplus Q''_{i,j} \oplus Q'_{i,j} \oplus Q_2|_\sigma.$$

As observed above, $Q'_{i,j}$ is (x, y) -symmetric. Further, it is easily checked that any restriction of an (x, y) -symmetric polynomial continues to be (x, y) -symmetric on the remaining variables. Hence, $Q_2|_\sigma$ is also (x, y) -symmetric. Further, note that $Q'_1|_\sigma$ has multidegree at most (i', j') . Also, by definition, the degree of $Q''_{i,j}$ is strictly less than $i + j$ and hence the multidegree of $Q''_{i,j}$ is at most (i', j') . Altogether, this implies that $Q|_\sigma$ is a sum of an (x, y) -symmetric polynomial (i.e. $Q'_{i,j} \oplus Q_2|_\sigma$) and a polynomial of multidegree at most (i', j') (i.e. $Q'_1|_\sigma \oplus Q''_{i,j}$). Thus, $Q|_\sigma$ is an $(x, y; i', j')$ -symmetric polynomial on r x -variables and r y -variables, where $r = n(i', j', \varepsilon/4)$.

Now, we analyze $\text{agr}(F_{H,\Phi,\Psi}, Q)$. Note that choosing a random function $H : \{0, 1\}^{2n} \rightarrow \{0, 1\}$ is the same as choosing each of its restrictions $H|_\sigma : \{0, 1\}^{2r} \rightarrow \{0, 1\}$ independently and uniformly at random.

We claim that for each σ , by the induction hypothesis, we have

$$\Pr_{H|_\sigma}[\text{agr}(F_{H,\Phi,\Psi}|_\sigma, Q|_\sigma) \geq \frac{5}{8} + \frac{\varepsilon}{4}] \leq \frac{\varepsilon}{4}. \quad (3.19)$$

Assuming the above, we show how to finish the proof. Let Y_Q denote the number of σ such that $\text{agr}(F_{H,\Phi,\Psi}|_\sigma, Q|_\sigma) \geq \frac{5}{8} + \frac{\varepsilon}{4}$. The random variable⁷ Y_Q is a sum of $2^{2(n-r)}$ independent 0-1 random variables with $\mathbf{E}[Y_Q] \leq 2^{2(n-r)} \cdot \frac{\varepsilon}{4}$. Thus, by the Chernoff bound, we have

$$\Pr_H[Y_Q \geq 2^{2(n-r)} \cdot \frac{\varepsilon}{2}] \leq \Pr_H[Y_Q - \mathbf{E}[Y_Q] \geq 2^{2(n-r)} \cdot \frac{\varepsilon}{4}] \leq \exp(-\Omega(\varepsilon^2 2^{2(n-r)})). \quad (3.20)$$

In the event that $Y_Q < 2^{2(n-r)} \cdot \frac{\varepsilon}{2}$, we have $\Pr_\sigma[\text{agr}(F_{H,\Phi,\Psi}|_\sigma, Q|_\sigma) \geq \frac{5}{8} + \frac{\varepsilon}{4}] \leq \frac{\varepsilon}{2}$. Hence, we see that in this case

$$\text{agr}(F_{H,\Phi,\Psi}, Q) = \mathbf{E}_\sigma[\text{agr}(F_{H,\Phi,\Psi}|_\sigma, Q|_\sigma)] \leq \left(\frac{5}{8} + \frac{\varepsilon}{4}\right) + \Pr_\sigma[\text{agr}(F_{H,\Phi,\Psi}|_\sigma, Q|_\sigma) \geq \left(\frac{5}{8} + \frac{\varepsilon}{4}\right)] \leq \frac{5}{8} + \frac{3\varepsilon}{4}. \quad (3.21)$$

In particular, the probability that there is any $Q \in \mathbb{F}_2[x_1, \dots, x_n, y_1, \dots, y_n]$ of degree at most $3d - 1$ such that $\text{agr}(F_{H,\Phi,\Psi}, Q) \geq \frac{5}{8} + \varepsilon$ can be upper bound bounded, using Eq. (3.20) and a union bound over all such Q , by

$$2^{(2n)^{3d-1}} \exp(-\Omega(\varepsilon^2 2^{n-r})) \leq \exp((2n)^{3d-1} - \Omega(\varepsilon^2 2^{n-r})) \leq \exp((2n)^{3d-1} - \Omega(\varepsilon^2 \frac{2^n}{n})) < \varepsilon.$$

Here, we have used the fact that the number of polynomials Q of degree at most $3d - 1$ is equal to the number of ways of choosing the coefficients (in \mathbb{F}_2) of $\binom{2n}{0} + \dots + \binom{2n}{3d-1} \leq (2n)^{3d-1}$ many monomials. The second inequality follows from the fact that $n \geq n_R(i, j, r) \geq 2^r$. The final inequality is true as long $n \geq n_1(\varepsilon)$ for some $n_1(\varepsilon) \in \mathbb{N}$.

Overall, we see that if we define $n(i, j, \varepsilon) = \max\{n_R(i, j, r), n_1(\varepsilon)\}$, then for any $n \geq n(i, j, \varepsilon)$, we have the statement of the lemma for $(x, y; i, j)$ -symmetric polynomials. This completes the induction.

It remains to prove Eq. (3.19). Fix any $\sigma \in \mathcal{A}_{I,J}$. Let $F' = F_{H,\Phi,\Psi}|_\sigma$. We use u and v to denote assignments to the variables indexed by I and J respectively and \tilde{u} and \tilde{v} to denote their natural completions to an assignment to all the variables (i.e. the other variables are assigned by σ). Assume that $d = 2^\ell$.

By the definition of $F_{H,\Phi,\Psi}$ in Eq. (3.7) and using Lemma 2.1, we have

$$F'(u, v, z) = \begin{cases} 0 & \text{if } |\tilde{u}|_\ell = \Phi_1(\tilde{u}), |\tilde{v}|_\ell = \Psi_1(\tilde{v}), \\ |\tilde{v}|_{\ell+1} \oplus \Psi_2(\tilde{v}) & \text{if } |\tilde{u}|_\ell = 1 \oplus \Phi_1(\tilde{u}) \text{ and } |\tilde{v}|_\ell = \Psi_1(\tilde{v}), \\ |\tilde{u}|_{\ell+1} \oplus \Phi_2(\tilde{u}) & \text{if } |\tilde{u}|_\ell = \Phi_1(\tilde{u}) \text{ and } |\tilde{v}|_\ell = 1 \oplus \Psi_1(\tilde{v}), \\ H|_\sigma(u, v) & \text{otherwise,} \end{cases} \quad (3.22)$$

where Φ_1, Φ_2 , being d -simple, are functions of $|\tilde{u}|_0, \dots, |\tilde{u}|_{\ell-1}$, and similarly, Ψ_1, Ψ_2 are functions of $|\tilde{v}|_0, \dots, |\tilde{v}|_{\ell-1}$.

⁷Note that Y_Q is a random variable since it depends on the random function H .

We would like to write the above in terms of the bits of $|u|$ and $|v|$. This is done as follows. Consider the case of $|\tilde{u}|_\ell$. Let $|\sigma_x|$ denote the number of 1s assigned by σ to the x variables. Note that $|\tilde{u}| = |u| + |\sigma_x|$, and hence it follows that the function $\Phi_1(\tilde{u}) = \Phi_1|_\sigma(u)$ is a function of $|u|_0, \dots, |u|_{\ell-1}$ and hence d -simple w.r.t. u ; similarly, $\Psi_1|_\sigma(\tilde{v}) = \Psi_1|_\sigma(v)$ is d -simple w.r.t. v . Similarly, we can also write $|\tilde{u}|_\ell = |u|_\ell \oplus \Phi'_1(u)$ for some d -simple $\Phi'_1(u)$ depending on σ ; also, $|\tilde{v}|_\ell = |v|_\ell \oplus \Psi'_1(v)$ for some d -simple $\Psi'_1(v)$ depending on σ .

Further elementary reasoning (left to the reader) allows us to deduce that there are d -simple $\Phi'_2(u), \Psi'_2(v)$ (depending on σ) such that

$$\begin{aligned} |\tilde{u}|_{\ell+1} &= |u|_{\ell+1} \oplus \Phi'_2(u) \text{ when } |\tilde{u}|_\ell = \Phi_1|_\sigma(u) \text{ and } |\tilde{v}|_\ell = 1 \oplus \Psi_1|_\sigma(v) \\ |\tilde{v}|_{\ell+1} &= |v|_{\ell+1} \oplus \Psi'_2(v) \text{ when } |\tilde{u}|_\ell = 1 \oplus \Phi_1|_\sigma(u) \text{ and } |\tilde{v}|_\ell = \Psi_1|_\sigma(v). \end{aligned}$$

The above along with [Eq. \(3.22\)](#) gives us

$$F'(u, v, z) = \begin{cases} 0 & \text{if } |u|_\ell = \Phi'_1(u) \text{ and } |v|_\ell = \Psi'_1(v), \\ |v|_{\ell+1} \oplus \Psi'_2(v) & \text{if } |u|_\ell = 1 \oplus \Phi'_1(u) \text{ and } |v|_\ell = \Psi'_1(v), \\ |u|_{\ell+1} \oplus \Phi'_2(u) & \text{if } |u|_\ell = \Phi'_1(u) \text{ and } |v|_\ell = 1 \oplus \Psi'_1(v), \\ H|_\sigma(u, v) & \text{otherwise,} \end{cases} \quad (3.23)$$

where for each $i \in [2]$, $\Phi'_i(u)$ satisfies $\Phi'_i = \Phi_i \oplus \Phi'_i$, and is hence d -simple w.r.t. u , and similarly $\Psi'_i = \Psi_i \oplus \Psi'_i$ is d -simple w.r.t. v . Hence, we see that $F' = F_{H, \Phi, \Psi}|_\sigma = F_{H|_\sigma, \Phi', \Psi'}$ (i.e. same as our hard function, but on $2r$ inputs). Using the fact that $r \geq n(i', j', \varepsilon/4)$, the induction hypothesis gives us

$$\Pr_{H|_\sigma}[\exists Q'' \in \mathbb{F}_2[x_L, y_J] \text{ of degree } \leq 3d - 1 \text{ and } (x, y; i', j')\text{-symmetric s.t. } \text{agr}(F', Q'') \geq \frac{5}{8} + \frac{\varepsilon}{4}] \leq \frac{\varepsilon}{4}.$$

In particular, since $Q|_\sigma$ is $(x, y; i', j')$ -symmetric, we have $\Pr_{H|_\sigma}[\text{agr}(F', Q|_\sigma) \geq \frac{5}{8} + \frac{\varepsilon}{4}] \leq \frac{\varepsilon}{4}$, which establishes [Eq. \(3.19\)](#) and completes the proof.

4 Upper bounds for $\gamma_{d,k}(\text{Maj}_n)$

In this section, we show an upper bound on $\gamma_{d,k}(\text{Maj}_n)$ where Maj_n denotes the Majority function on n bits⁸

Theorem 4.1. *For any $k \geq 1, d \in \mathbb{Z}^+, \gamma_{d,k}(\text{Maj}_n) \leq \frac{1}{2} + \frac{10d}{\sqrt{n}}$.*

The proof of [Theorem 4.1](#) presented below is an adaptation of techniques appearing in a work of Green [[Gre00](#)], who proved a similar result on the approximability of the parity function by polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$, for prime $p \neq 2$.

We will need some definitions and facts about $\mathcal{P}_{d,k}$.

We use π to denote the unique ring homomorphism from $\mathbb{Z}/2^k\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. Its kernel $\pi^{-1}(0) = \{a \in \mathbb{Z}/2^k\mathbb{Z} \mid 2^{k-1}a = 0\}$ is the set of non-invertible elements in $\mathbb{Z}/2^k\mathbb{Z}$.

We call a set $S \subseteq \{0, 1\}^n$ *forcing* for $\mathcal{P}_{d,k}$ if any polynomial $P \in \mathcal{P}_{d,k}$ that vanishes over S is forced to take a value in $\pi^{-1}(0)$ at all points $x \in \{0, 1\}^n$. Formally,

$$(\forall x \in S \ P(x) = 0) \Rightarrow (\forall y \in \{0, 1\}^n \ \pi(P(y)) = 0).$$

⁸We define the majority function as $\text{Maj}_n(x) = 1$ iff $|x| > n/2$.

Define the polynomial $\pi(P) \in \mathbb{Z}/2\mathbb{Z}[x_1, \dots, x_n]$ to be the polynomial obtained by applying the map π to each of the coefficients of P . Since a multilinear polynomial in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \dots, x_n]$ is the zero polynomial iff it vanishes at all points of $\{0, 1\}^n$ (by [Lemma 2.5](#)), we see that S is forcing iff $(\forall x \in S \ P(x) = 0) \Rightarrow \pi(P) = 0$.

Note that any interpolating set for $\mathcal{P}_{d,k}$ (see [Section 2](#) for the definition) is forcing for $\mathcal{P}_{d,k}$, but the converse need not be true.

We now adapt the proof of [Lemma 11](#) in [[Gre00](#)] to bound the size of forcing sets for $\mathcal{P}_{d,k}$.

Lemma 4.2. *If S is forcing for $\mathcal{P}_{d,k}$, then $|S| \geq |\{0, 1\}_{\leq d}^n| = \binom{n}{\leq d}$.*

Proof. Assume for the sake of contradiction that $S \subseteq \{0, 1\}^n$ is forcing for $\mathcal{P}_{d,k}$ and $|S| < \binom{n}{\leq d}$. The latter implies the existence of a non-zero multilinear polynomial $Q(x) \in \mathbb{Q}[x_1, \dots, x_n]$ of degree at most d satisfying $Q(x) = 0$ for all $x \in S$.

Let $Q'(x) = \sum_{T \subseteq [n]} c_T \prod_{i \in T} x_i$ be the polynomial in $\mathbb{Z}[x_1, \dots, x_n]$ obtained by first clearing out the denominators of the coefficients of $Q(x)$, followed by dividing the resulting polynomial by the GCD of all the coefficients. Finally, let $P(x) = \sum_{T \subseteq [n]} c'_T \prod_{i \in T} x_i$ be any polynomial in $\mathcal{P}_{d,k}$ satisfying $c'_T \equiv c_T \pmod{2^k}$. It follows that P is a non-zero polynomial of degree at most d such that $\pi(P) \neq 0$, since $\pi(P) = 0$ would imply that every coefficient of P (and thus every coefficient of Q') is divisible by two, which is impossible since the coefficients of Q' have no common divisor.

To complete the proof, observe that $P(x) = 0$ for all $x \in S$, and since S is forcing for $\mathcal{P}_{d,k}$, this implies that $\pi(P) = 0$, which is a contradiction. □

We now use [Lemma 4.2](#) to prove [Theorem 4.1](#).

Proof of [Theorem 4.1](#). We assume throughout that $1 \leq d \leq \frac{\sqrt{n}}{10}$; otherwise, there is nothing to prove. Let $\text{Maj}_{n,k} : \{0, 1\}^n \rightarrow \mathbb{Z}/2^k\mathbb{Z}$ be the k -lift of the Maj_n function. Let $P \in \mathcal{P}_{d,k}$ be arbitrary and let $S_P = \{x \in \{0, 1\}^n \mid P(x) = \text{Maj}_{n,k}(x)\}$. We want to show that $|S_P| \leq 2^n \cdot \left(\frac{1}{2} + \frac{10d}{\sqrt{n}}\right)$. We will argue by contradiction. So assume that $|S_P| > 2^n \cdot \left(\frac{1}{2} + \frac{10d}{\sqrt{n}}\right)$.

Let E_P be the complement of S_P , i.e. the set of points where P makes an error in computing $\text{Maj}_{n,k}$. We have $|E_P| < 2^n \left(\frac{1}{2} - \frac{10d}{\sqrt{n}}\right)$. We will try to find a degree D (for suitable $D \leq \lfloor n/2 \rfloor$) polynomial Q such that Q vanishes at all points in E_P but has the property that $Q(x)$ is a unit (i.e. $\pi(Q(x)) \neq 0$) for some $x \in \{0, 1\}^n$. To be able to do this, we need the fact that E_P is not forcing for $\mathcal{P}_{D,k}$. By [Lemma 4.2](#), if E_P is indeed forcing for $\mathcal{P}_{D,k}$, then

$$\begin{aligned} |E_P| &\geq \sum_{i=0}^D \binom{n}{i} = \left(\sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} \right) - \sum_{i=D+1}^{\lfloor n/2 \rfloor} \binom{n}{i} \\ &\geq 2^{n-1} - (\lfloor n/2 \rfloor - D) \cdot \binom{n}{\lfloor n/2 \rfloor} \\ &\geq 2^n \cdot \left(\frac{1}{2} - \frac{2(\lfloor n/2 \rfloor - D)}{\sqrt{n}} \right) = 2^n \cdot \left(\frac{1}{2} - \frac{4d}{\sqrt{n}} \right) \end{aligned}$$

where the last equality follows if we choose $D = \lfloor n/2 \rfloor - 2d$. This contradicts our upper bound on the size of $|E_P|$. Hence, E_P cannot be forcing for $\mathcal{P}_{D,k}$. In particular, we can find Q that vanishes on E_P and furthermore, $\pi(Q(x)) \neq 0$ for some $x \in \{0, 1\}^n$.

We now claim that $\pi(Q(x_0)) \neq 0$ for some x_0 of Hamming weight $> n/2$. To see this, consider the polynomial $Q_1 = \pi(Q)$. By construction of Q , we know that Q_1 is a non-zero polynomial of degree D . Hence, by [Lemma 2.5](#), Q_1 is non-zero when restricted to the Hamming ball of radius $D < n/2$ around the all 1s vector. In particular, this implies that there is an input x_0 of Hamming weight $> n/2$ where $Q_1(x_0)$ is non-zero and hence $\pi(Q(x_0)) \neq 0$, or equivalently $2^{k-1}Q(x_0) \neq 0$. Fix this x_0 for the remainder of the proof. Note that $x_0 \notin E_P$ since Q vanishes on E_P .

Now, consider the polynomial $R(x) = Q(x) \cdot P(x)$. We first show that $R(x) = 0$ for all x of Hamming weight $\leq n/2$. Consider any x of Hamming weight $\leq n/2$. If $x \in E_P$, then $R(x) = 0$ since $Q(x) = 0$. On the other hand, if $x \notin E_P$, then $P(x) = \text{Maj}_{n,k}(x) = 0$ since x has Hamming weight $\leq n/2$. Thus, R vanishes at all inputs of Hamming weight $\leq n/2$.

Since the degree of R is at most $\deg(Q) + \deg(P) = D + d = (\lfloor n/2 \rfloor - 2d) + d \leq \lfloor n/2 \rfloor - d$ and R vanishes at all inputs of $\{0, 1\}_{\leq n/2}^n$, this implies (by [Lemma 2.5](#)) that R must be 0 everywhere. However, at x_0 , $R(x_0) = Q(x_0)P(x_0) = Q(x_0)\text{Maj}_{n,k}(x_0) = 2^{k-1}Q(x_0) \neq 0$. This yields the desired contradiction. \square

5 Connection to non-classical polynomials

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the one dimensional torus. Observing that the additive structure of \mathbb{F}_2 is isomorphic to the additive subgroup $\{0, 1/2\} < \mathbb{T}$, we can think of a Boolean function $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ as a function $F : \mathbb{F}_2^n \rightarrow \{0, 1/2\}$, and conversely, a map $F : \mathbb{F}_2^n \rightarrow \{0, 1/2\}$ as a Boolean function.

Tao and Ziegler [[TZ12](#)] give a characterization of non-classical polynomials as follows:

Definition 5.1 (Tao and Ziegler [[TZ12](#)]). *A function $F : \mathbb{F}_2^n \rightarrow \mathbb{T}$ is a non-classical polynomial of degree $\leq d$ if and only if it has the following form:*

$$F(x_1, \dots, x_n) = \alpha + \sum_{0 \leq e_1, \dots, e_n \leq 1, k \geq 1: \sum_i e_i + (k-1) \leq d} \frac{c_{e_1, \dots, e_n, k} x_1^{e_1} \dots x_n^{e_n}}{2^k} \pmod{1}$$

Here $\alpha \in \mathbb{T}$, and $c_{e_1, \dots, e_n, k} \in \{0, 1\}$ are uniquely determined. α is called the shift of F , and the largest k such that $c_{e_1, \dots, e_n, k} \neq 0$ for some $(e_1, \dots, e_n) \in \{0, 1\}^n$ is called the depth of F .

Since we are interested in the agreement of a non-classical polynomial with Boolean ($\{0, 1/2\}$ -valued) functions, we will only consider polynomials with shift $\alpha = \frac{A}{2^k}$, where k is the depth of the polynomial and $A \in \{0, \dots, 2^k - 1\}$.

Remark 5.2. *Classical polynomials are non-classical polynomials with $\alpha \in \{0, 1/2\}$ and depth = 1. It is easy to see that every classical polynomial corresponds to a Boolean function. It is also not hard to show that every Boolean function can be represented as a classical polynomial.*

The following lemma relates our model to non-classical polynomials:

Lemma 5.3. *Let F be a Boolean function, and $d, k \in \mathbb{Z}^+$, $d \geq k$.*

1. *If there is a non-classical polynomial P of degree d and depth k satisfying $\text{agr}(F, P) = \gamma$, then there is a $P' \in \mathcal{P}_{d,k}$ satisfying $\text{agr}(F_k, P') = \gamma$, where F_k is the k -lift of F .*
2. *If there is a $P \in \mathcal{P}_{d,k}$ satisfying $\text{agr}(F_k, P) = \gamma$, then there is a non-classical polynomial P' of degree $\leq d + k - 1$ and depth k satisfying $\text{agr}(F, P') = \gamma$.*

Proof. Fix F , d , and k for the rest of the proof.

Proof of 1: Let P be a non-classical polynomial of degree d and depth k with $\text{agr}(F, P) = \gamma$. It is not hard to verify that P can be written in the following form (See, e.g., proof of Lemma 2.2 in [BL15]):

$$P(x) = \frac{P''(x)}{2^k} \pmod{1}$$

where $P''(x) \in \mathbb{Z}[x_1, \dots, x_n]$ is of degree d .

Suppose $P''(x) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i$. Choose $P' \in \mathcal{P}_{d,k}$, $P'(x) = \sum_{S \subseteq [n]} c'_S \prod_{i \in S} x_i$, satisfying

$$\forall S \subseteq [n], c'_S \equiv c_S \pmod{2^k}.$$

By our choice of P' , we have that, for every $x \in \{0, 1\}^n$ and $a \in \{0, \dots, 2^k - 1\}$,

$$P(x) = \frac{a}{2^k} \Leftrightarrow P'(x) = a.$$

It follows that $\text{agr}(F_k, P') = \gamma$.

Proof of 2: Let $P \in \mathcal{P}_{d,k}$ such that $\text{agr}(F_k, P) = \gamma$. Using arguments similar to above, we can find a $P'' \in \mathbb{Z}[x_1, \dots, x_n]$ of degree d such that $P''(x) \equiv P(x) \pmod{2^k}$, for all $x \in \{0, 1\}^n$.

Define $P'(x)$ as

$$P'(x) = \frac{P''(x)}{2^k} \pmod{1}.$$

By comparing to the form in Definition 5.1, it is easy to see that $P'(x)$ is a non-classical polynomial of degree at most $d + k - 1$ and depth k . Furthermore, we have that, for all $x \in \{0, 1\}^n$ and $a \in \{0, \dots, 2^k - 1\}$,

$$P(x) = a \Leftrightarrow P'(x) = \frac{a}{2^k}.$$

This completes the proof. □

The first part of Lemma 5.3 implies the following corollary of Theorem 4.1:

Corollary 5.4. *Let $F : \mathbb{F}_2^n \rightarrow \mathbb{T}$ be a non-classical polynomial of degree d . Then,*

$$\Pr_{x \sim \mathbb{F}_2^n} [\text{Maj}_n(x) = F(x)] \leq \frac{1}{2} + O\left(\frac{d}{\sqrt{n}}\right).$$

This proves a conjecture of Bhowmick and Lovett [BL15] that non-classical polynomials of degree d do not approximate the Majority function any better than classical polynomials of the same degree.

The following is a consequence of Theorem 2.2 and the first part of Lemma 5.3:

Corollary 5.5. *Let $\ell \geq 2$. Then, for every classical polynomial $P : \mathbb{F}_2^n \rightarrow \mathbb{T}$ of degree $\leq 2^\ell - 1$,*

$$\Pr_{x \sim \mathbb{F}_2^n} [P(x) = S_{2^\ell}(x)] \leq \frac{1}{2} + o(1).$$

On the other hand, the second part of Lemma 5.3 and Theorem 3.1 imply

Corollary 5.6. *For every $\ell \geq 2$, there is a non-classical polynomial $F : \mathbb{F}_2^n \rightarrow \mathbb{T}$ of degree $\leq 2^{\ell-1} + 2^{\ell-2} + 2$ and depth 3 such that*

$$\Pr_{x \sim \mathbb{F}_2^n} [F(x) = S_{2^\ell}(x)] \geq \frac{9}{16} - o(1)$$

Noting that $2^{\ell-1} + 2^{\ell-2} + 2 < 2^\ell$ for $\ell \geq 4$, [Corollary 5.5](#) and [Corollary 5.6](#) imply the following:

Theorem 5.7. *There is a Boolean function $F : \mathbb{F}_2^n \rightarrow \{0, 1/2\}$ and $d \geq 1$, such that for every classical polynomial P of degree at most d , we have*

$$\Pr_{x \sim \mathbb{F}_2^n} [F(x) = P(x)] \leq \frac{1}{2} + o(1),$$

but there is a non-classical polynomial P' of degree $d' \leq d$ satisfying

$$\Pr_{x \sim \mathbb{F}_2^n} [F(x) = P'(x)] \geq \frac{1}{2} + \Omega(1).$$

This provides a counterexample to an informal conjecture of Bhowmick and Lovett [[BL15](#)] that, for any Boolean function F , non-classical polynomials of degree d do not approximate F any better than classical polynomials of the same degree.

6 Acknowledgements

We would like to thank David Barrington for taking the time to explain Szegedy’s [[Sze89](#)] result to us, Arkadev Chattopadhyay for referring us to Green’s result [[Gre00](#)], and Swagato Sanyal for helpful discussions. We are also grateful to the anonymous reviewers for their detailed and helpful comments. In particular, we thank an anonymous reviewer for STACS 2017 who pointed out an error in the induction case of a previous proof of [Lemma 3.8](#).

References

- [AB01] NOGA ALON and RICHARD BEIGEL. *Lower bounds for approximations by low degree polynomials over \mathbb{Z}_m* . In *Proc. 16th IEEE Conf. on Computational Complexity*, pages 184–187. 2001. doi:10.1109/CCC.2001.933885. 3, 8
- [AWY15] AMIR ABOUD, RICHARD RYAN WILLIAMS, and HUACHENG YU. *More applications of the polynomial method to algorithm design*. In *Proc. 26th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 218–230. 2015. doi:10.1137/1.9781611973730.17. 1
- [BHS17] ABHISHEK BHRUSHUNDI, PRAHLADH HARSHA, and SRIKANTH SRINIVASAN. *On polynomial approximations over $\mathbb{Z}/2^k\mathbb{Z}$* . In BRIGITTE VALLÉE and HERIBERT VOLLMER, eds., *Proc. 34th Annual Symp. on Theoretical Aspects of Comp. Science (STACS)*, volume 66 of *LIPIcs*. Schloss Dagstuhl, 2017. 0
- [BL15] ABHISHEK BHOWMICK and SHACHAR LOVETT. *Nonclassical polynomials as a barrier to polynomial lower bounds*. In *Proc. 30th Computational Complexity Conf.*, pages 72–87. 2015. arXiv:1412.4719, doi:10.4230/LIPIcs.CCC.2015.72. 1, 3, 18, 19

- [Efr12] KLIM EFREMENKO. *3-query locally decodable codes of subexponential length*. SIAM J. Comput., 41(6):1694–1703, 2012. (Preliminary version in *41st STOC*, 2009). [eccc:TR08-069](#), [doi:10.1137/090772721](#). 1
- [Gop08] PARIKSHIT GOPALAN. *Query-efficient algorithms for polynomial interpolation over composites*. SIAM J. Comput., 38(3):1033–1057, 2008. (Preliminary version in *17th SODA*, 2006). [doi:10.1137/060661259](#). 4
- [Gre00] FREDERIC GREEN. *A complex-number Fourier technique for lower bounds on the mod- m degree*. Comput. Complexity, 9(1):16–38, 2000. (Preliminary version in *12th STACS*, 1995). [doi:10.1007/PL00001599](#). 3, 15, 16, 19
- [Gro00] VINCE GROLMUSZ. *Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs*. Combinatorica, 20(1):71–86, 2000. [doi:10.1007/s004930070032](#). 1
- [GT09] BEN JOSEPH GREEN and TERENCE TAO. *The distribution of polynomials over finite fields, with applications to the Gowers norm*. Contributions to Discrete Mathematics, 4(2), 2009. [arXiv:0711.3191](#). 3
- [Knu97] DONALD ERVIN KNUTH. *Fundamental Algorithms*, volume I of *The Art of Computer Programming*. Addison-Wesley, 3rd edition, 1997. 3, 6
- [KS04] ADAM R. KLIVANS and ROCCO A. SERVEDIO. *Learning DNF in time $2^{O(n^{1/3})}$* . J. Comput. Syst. Sci., 68(2):303–318, 2004. (Preliminary version in *33rd STOC*, 2001). [doi:10.1016/j.jcss.2003.07.007](#). 1
- [LMN93] NATHAN LINIAL, YISHAY MANSOUR, and NOAM NISAN. *Constant depth circuits, Fourier transform, and learnability*. J. ACM, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). [doi:10.1145/174130.174138](#). 1
- [Raz87] ALEXANDER A. RAZBOROV. *Mathematicheskii Zametki*, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). [doi:10.1007/BF01137685](#). 1
- [Smo87] ROMAN SMOLENSKY. *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*. In *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82. 1987. [doi:10.1145/28395.28404](#). 1
- [Smo93] ———. *On representations by low-degree polynomials*. In *Proc. 34th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 130–138. 1993. [doi:10.1109/SFCS.1993.366874](#). 1
- [Sze89] MARIO SZEGEDY. *Algebraic Methods in Lower Bounds for Computational Models with Limited Communication*. Ph.D. thesis, University of Chicago, 1989. 1, 19
- [TZ12] TERENCE TAO and TAMAR ZIEGLER. *The inverse conjecture for the Gowers norm over finite fields in low characteristic*. Ann. Comb., 16(1):121–188, 2012. [arXiv:1101.1469](#), [doi:10.1007/s00026-011-0124-3](#). 17
- [Wil14] RYAN WILLIAMS. *New algorithms and lower bounds for circuits with linear threshold gates*. In *Proc. 46th ACM Symp. on Theory of Computing (STOC)*, pages 194–202. 2014. [arXiv:1401.2444](#), [doi:10.1145/2591796.2591858](#). 1

A Proof of Lemma 3.12

Proof. Note that the constraint $n_R(i, j, r) \geq 2^r$ is easy to satisfy since if the latter part of the lemma holds for some $n_R(i, j, r) < 2^r$, then it continues to be the case for $n_R(i, j, r) = 2^r$. So we ignore the constraint $n_R(i, j, r) \geq 2^r$ for the rest of the proof.

We prove by induction the following stronger statement. For any $i, j \in \mathbb{N}$ and any $r_0, s_0, r_1, s_1 \in \mathbb{N}$, there is an $m_R(i, j; r_0, s_0, r_1, s_1) \in \mathbb{N}$ such that for any $n \geq m_R(i, j; r_0, s_0, r_1, s_1)$, any disjoint n -sets I, J and any (i, j) -colouring c of (I, J) , one of the following holds.

- There are sets $I' \subseteq I$ and $J' \subseteq J$ with $|I'| = r_0$ and $|J'| = s_0$ such that the restriction c_1 of c to $\binom{I'}{i} \times \binom{J'}{j}$ is the constant 0 function.
- There are sets $I' \subseteq I$ and $J' \subseteq J$ with $|I'| = r_1$ and $|J'| = s_1$ such that the restriction c_1 of c to $\binom{I'}{i} \times \binom{J'}{j}$ is the constant 1 function.

Setting $r_0 = r_1 = s_0 = s_1 = r$ above clearly yields the lemma.

The proof is by induction on $\min\{i, j\}$. Note that the statement is trivial when $i = j = 0$, since a $(0, 0)$ -colouring is by definition a constant function. So we can take $m_R(0, 0; r_0, s_0, r_1, s_1) = \max\{r_0, s_0, r_1, s_1\}$ for any $r_0, r_1, s_0, s_1 \in \mathbb{N}$.

Now consider the case when $\min\{i, j\} = 0$ and $\max\{i, j\} \geq 1$; w.l.o.g. assume $j = \max\{i, j\}$. In this case, the function c is essentially a colouring of $\binom{J}{j}$ and hence the statement of the lemma reduces to the case of the standard Ramsey theorem for j -uniform hypergraphs. Thus, we know that $m_R(i, j; r_0, s_0, r_1, s_1)$ exists in this case. This completes the base case.

For the induction, assume the statement for any $r_0, s_0, r_1, s_1 \in \mathbb{N}$ and any (i', j') with $\min\{i', j'\} < k$ for some $k \geq 1$. Consider the case of (i, j) such that $\min\{i, j\} = k$. Assume w.l.o.g. that $i = \min\{i, j\} \geq 1$. We now proceed by induction on $t = r_0 + s_0 + r_1 + s_1$.

The base case of the induction is when $\min\{r_0, s_0, r_1, s_1\} = 0$, which is trivial as $i, j \geq 1$. For the induction case, assume that $\min\{r_0, s_0, r_1, s_1\} \geq 1$ and we have the statement for smaller values of t . W.l.o.g. assume that $r_0 = \min\{r_0, s_0, r_1, s_1\} \geq 1$.

By the induction hypotheses, we know the existence of

$$m_1 = \max\{m_R(i, j; r_0 - 1, s_0, r_1, s_1), m_R(i, j; r_0, s_0, r_1 - 1, s_1)\}$$

and $m_2 = m_R(i - 1, j; m_1, m_1, m_1, m_1)$. We claim that $m_R(i, j; r_0, s_0, r_1, s_1) = m_2 + 1$ has the required properties.

To see this, consider any (i, j) -colouring c of (I, J) with $|I| = |J| \geq m_2 + 1$. Fix an arbitrary $a \in I$ and $b \in J$. Note that for $I_1 = I \setminus \{a\}$ and $J_1 = J \setminus \{b\}$, we obtain a $(i - 1, j)$ colouring c_a of (I_1, J_1) by setting $c_a(A, B) = c(A \cup \{a\}, B)$. Since $|I_1| = |J_1| \geq m_2$, we know that there exist $I_2 \subseteq I_1$ and $J_2 \subseteq J_1$ of size m_1 each such that the restriction of c_a to $\binom{I_2}{i-1} \times \binom{J_2}{j}$ is a constant. Equivalently, there is an $\alpha \in \mathbb{F}_2$ such that for each $A \in \binom{I_2}{i-1}$ and $B \in \binom{J_2}{j}$, we have $c(A \cup \{a\}, B) = \alpha$.

Assume $\alpha = 0$. Now, consider the restriction c_2 of c to $\binom{I_2}{i} \times \binom{J_2}{j}$. Since $|I_2| = |J_2| = m_1 \geq m_R(i, j; r_0 - 1, s_0, r_1, s_1)$, we see that there exist $I'' \subseteq I_2$ and $J'' \subseteq J_2$ satisfying one of the following.

- $|I''| = r_0 - 1$ and $|J''| = s_0$, and the restriction c'' of c to $\binom{I''}{i} \times \binom{J''}{j}$ is the constant 0 function.
- $|I''| = r_1$ and $|J''| = s_1$, and the restriction c'' of c to $\binom{I''}{i} \times \binom{J''}{j}$ is the constant 1 function.

In the former case, we can take $I' = I'' \cup \{a\}$ and $J' = J''$ to prove the inductive case. Note that since c_a is the constant 0 function on $\binom{I''}{i-1} \times \binom{J''}{j}$, the restriction c' of c to $\binom{I'}{i} \times \binom{J'}{j}$ is also the constant 0 function.

In the latter case, we just take $I' = I''$ and $J' = J''$, since we are guaranteed that the restriction of c to $\binom{I'}{i} \times \binom{J'}{j}$ is the constant 1 function.

In the case that $\alpha = 1$, we repeat the same proof except that we use the fact that $m_1 \geq m_R(i, j; r_0, s_0, r_1 - 1, s_1)$ to prove that there exist $I'' \subseteq I_2$ and $J'' \subseteq J_2$ satisfying one of the following.

- $|I''| = r_0$ and $|J''| = s_0$, and the restriction c'' of c to $\binom{I''}{i} \times \binom{J''}{j}$ is the constant 0 function.
- $|I''| = r_1 - 1$ and $|J''| = s_1$, and the restriction c'' of c to $\binom{I''}{i} \times \binom{J''}{j}$ is the constant 1 function.

This proves the inductive case, and hence completes the proof. \square