Deterministic Approximate Counting $\mathbb{F}_2$ Polynomials Via Correlation-based Fourier Bounds

Szymon Snoeck, Sam Wang

April 2024

1 Background

Multivariate $\mathbb{F}_2$ polynomials form one of the most basic yet powerful computational models. Each polynomial amounts to a parity of conjunctions and can be written as:

$$p(x_1, \ldots, x_n) = \bigoplus_{S \in [n]} \alpha_S \bigwedge_{i \in S} x_i$$

Where each $\alpha_S \in \{0, 1\}$. Despite its simple presentation, $\mathbb{F}_2$ polynomials can compute all functions. Indeed every $f : \{-1, 1\}^n \to \{0, 1\}$ is computed by:

$$p(x_1, \ldots, x_n) = \bigoplus_{\alpha \in \{-1,1\}^n} f(\alpha) \bigwedge_{i \in [n]} (1 + \alpha_1 + x_i)$$

Hence, we must consider some natural restrictions on the complexity of the polynomials. One natural restriction is to bound the sparsity of the polynomial (i.e. bound the number of conjunctions being paritied). The other natural restriction is to bound the fan-in of the conjunctions, which coincides with the degree of the polynomial.

From an unconditional derandomization perspective, we are interested in deterministically estimating the expected values of these polynomials (or equivalently estimating the number of roots) to make progress on the **BPP = P** question. The first people to study $\mathbb{F}_2$ polynomials from a complexity-theoretic perspective were Ehrenfeucht and Karpinski who in 1990 proved that counting the number of roots of $\mathbb{F}_2$ polynomials was #P-complete for degree ≥ 3 [EK90]. Later that same year, [KL90] developed a randomized relative-error polynomial-time approximation algorithm for the number of roots of sparse $\mathbb{F}_2$ polynomials which gave hope that there might be an efficient deterministic algorithm. Luby, Velickovic, and Wigderson [LVW93] gave the first deterministic approximate counter (DAC) for sparse $\mathbb{F}_2$ polynomials. Their result was based on a modification of the Nisan-Wigderson "hardness versus randomness" paradigm which turns correlation bounds against a sufficiently strong class into a pseudorandom generator (PRG) against a weaker class. Notably, [LVW93]'s result generalized to any function consisting of a symmetric gate over conjunctions (i.e. SYM$\circ$AND).

Fourteen years later and using the same Nisan-Wigderson framework, this result was improved by [Vio07] who gave an ϵ-PRG for size-S SYM$\circ$AC^0_d circuits with seed length $2^{O(\sqrt{\log(S/\epsilon)})}$, thus expanding the class fooled while maintaining the same runtime. The key insight of [Vio07] was that correlation bounds against the larger class SYM$\circ$AND$\circ$OR translates into PRGs against SYM$\circ$AND via the Nisan-Wigderson framework. In a pattern that continued with future work, the correlation bounds were proved via random restrictions, a method dating back to the work of [Hås86] and [Ajt83]. [Vio07] utilized the famous [Hås86] switching lemma for this task. [LS11] subsequently improved on this correlation bound using [Ajt83]'s multi-switching lemma only at the cost of the size of the circuit that could be fooled. Hence [LS11]'s work produced an ϵ-PRG with seed length $2^{O(\log(S)/\log\log(S))} + (\log(1/\epsilon))^{2+o(1)}$. A significant improvement in the ϵ dependence at the cost of a dramatically worse dependence on the circuit size. Some seven years later, Servedio and Tan in a series of two papers, [ST18a] and [ST18b], managed to achieve the best of both worlds using the more recently proved Håstad's multi switching lemma [Hås14]. This allowed them to achieve a seed length of $2^{O(\sqrt{\log S})} + \text{polylog}(1/\epsilon)$ against the class SYM $\circ$ AC^0_d. Notice the retention of the size dependence demonstrated in [LVW93] and [Vio07] while keeping the ϵ dependence in line with the result of [LS11]. Moreover, in [ST18b] it was shown that the ϵ dependence is optimal up to a polynomial factor, and improving on the S-dependence would require "groundbreaking new lower bounds against low-degree $\mathbb{F}_2$ polynomials and ACC^0 circuits." Hence, it is unlikely that further improvement is possible using the Nisan-Widgerson framework.

In terms of PRGs for the class $\mathbb{F}_2^{(d,n)}$, all $\mathbb{F}_2$ polynomials over n variables of degree at most d, the first PRG was given by [NN93] for the $d = 1$ case with seed length $O(\log n)$ and error $1/n$ which is optimal (see [Alo+90]). To achieve this result, [NN93] introduced the small-bias generator which became the key ingredient to future attempts at fooling $\mathbb{F}_2^{(d,n)}$. After fourteen years with almost no progress, [BV07] introduced a novel approach: sum together d small-biased generators to fool $\mathbb{F}_2^{(d,n)}$. Using Gowers norms to analyze the resulting generator, it was unconditionally proved that this technique fooled $\mathbb{F}_2^{(d,n)}$ for $d \leq 3$. The higher degree case was also conditionally proved in [BV07] under the assumption that the *Inverse Conjecture for the Gowers norm* held. However, this conjecture was proved to be false in the general setting by [GT07]. The first proper unconditional result was by [Lov08] who demonstrated that the sum of 2^d small-biased generators fools $\mathbb{F}_2^{(d,n)}$. This proof unlike that of [BV07] did not rely on the theory of Gowers norms, and the result's degree dependence was still quite poor. The latest and current state-of-the-art PRG for $\mathbb{F}_2^{(d,n)}$ came only a year later. [Vio08] revisited the theory of Gowers Norms and using the"squaring trick" (see [Vio22]) proved that the sum of d small-biased generators fools $\mathbb{F}_2^{(d,n)}$. The resulting seed length is $O(d\log(n) + d \cdot 2^d \log(1/\epsilon))$ which falls just short of fooling $\mathbb{F}_2^{(\log(n),n)}$. It is still an open question whether $\mathbb{F}_2^{(\log(n),n)}$ can be fooled even with the most forgiving parameters:

Open Question 1. *[Vio22] Is there a PRG with seed length $\frac{n}{2}$ that fools polynomials of degree $\log n$ with error $\frac{1}{3}$?*

More recently, a new approach to this problem has been developing. [Cha+18a] introduced the polarizing random walks framework, a novel way to create explicit PRGs based on Fourier tails. Given a boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we define the $\mathcal{L}_1$-norm of the k-th Fourier level as

$$\mathcal{L}_{1,k}(f) = \sum_{S \subseteq [n] : |S| = k} |\widehat{f(S)}|.$$

Further, we extend this notion to families of functions, $\mathcal{F}$:

$$\mathcal{L}_{1,k}(\mathcal{F}) = \max_{f \in \mathcal{F}} \mathcal{L}_{1,k}(f).$$

In particular,

$$\mathcal{L}_{1,k}(\mathbb{F}_2^{(d,n)}) = \mathcal{L}_{1,k}(d).$$

[Cha+18a] proved the following theorem:

Theorem 1. *[Cha+18a] Let $\mathcal{F}$ be a family of boolean functions that is closed under restrictions and that*

$$\mathcal{L}_{1,k}(\mathcal{F}) \leq a \cdot b^k$$

for all $k \in [n]$. Then for any $\epsilon > 0$, there exists an explicit PRG for $\mathcal{F}$ with error ϵ and seed length $b^2 \cdot \mathrm{polylog}(an/\epsilon)$.

In an attempt to create a PRG for $\mathbb{F}_2^{(d,n)}$, [Cha+18a] proved the following tail bound for $\mathbb{F}_2^{(d,n)}$ via an inductive argument on the degree and Fourier level:

Theorem 2. *[Cha+18a] For all $k \in [n]$, $\mathcal{L}_{1,k}(d) \leq (k2^{3d})^k$.*

However, this result is too weak to develop a non-trivial PRG under [Cha+18a]'s incarnation of the framework. Two years later, the polarizing random walks framework was modified by [Cha+20]. [Cha+20] proved that a bound on $\max_{f \in \mathcal{F}} |\sum_{S:|S|=k} \widehat{f(S)}|$ for some k alongside a bound on the $\mathcal{L}_1$-norm of the Fourier levels less than k implies the existence of an explicit PRG. This stronger framework allows the bounds from [Cha+18a] to give a PRG for $\mathbb{F}_2^{(d,n)}$ with seed length $2^{O(d)}\mathrm{polylog}(n/\epsilon)$. Though the seed length is worse than that of [Vio08], this is still a major success of the framework as a proof of concept. It was able to take quite unsharp bounds and harness them to create near state-of-the-art results.

Another approach via the polarizing random walks framework was developed by [Cha+18b]. Instead of requiring broad Fourier bounds on many levels, [Cha+18b] proved that decent results could still be obtained if only the $\mathcal{L}_1$-norm of the second level was bound:

Theorem 3. *[Cha+18b] Let $\mathcal{F}$ be a family of boolean functions that is closed under restrictions such that*

$$\mathcal{L}_{1,2}(\mathcal{F}) \leq t.$$

Then, for any $\epsilon > 0$, there exists an explicit PRG for $\mathcal{F}$ with error ϵ and seed length $\mathrm{poly}(t, \log n, 1/\epsilon)$.

The main difference in the seed length between [Cha+18a] and [Cha+18b] is the loss of the logarithmic dependence on ϵ. While significant, this is a surprisingly small loss to pay to only have to worry about the second Fourier level and nothing else. Several conjectures have been put forth as a result of these Fourier tail-based frameworks. The two most important being:

Conjecture 1. *[Cha+18b]* $\mathcal{L}_{1,2}(d) \leq O(d^2)$.

Conjecture 2. *[Cha+20]* $\max_{g \in \mathbb{F}_2^{(d,n)}} \left| \sum_{S:|S|=k} \widehat{g(S)} \right| \leq 2^{o(dk)+O(k \log\log(n))}$ *for all* $k \leq O(\log(n))$.

If proved, these conjectures would positively answer [Vio22]'s **open question 1** and lead to groundbreaking progress on long-standing problems in derandomization. For example, proving **conjecture 1** would result in an explicit PRG against the class $\mathrm{AC}^0[\oplus]$ with poly-logarithmic seed length.

2 Notation and Definitions

For ease of notation, let $\mathrm{Cor}[f, d] = \max_{g \in \mathbb{F}_2^{(d,n)}} \mathrm{Cor}[f, g]$. Additionally, for a symmetric function $f : \{-1, 1\}^n \to \{-1, 1\}$, its output is only a function of the hamming weight thus we can arbitrarily write it as $f : \{0, \ldots, n\} \to \{-1, 1\}$ such that $f(|x|_H) = f(x)$ where $|\cdot|_H$ is the Hamming weight. Moreover, we can extend this concept to the Fourier spectrum (see **lemma 1**) of f by defining $\widehat{f(k)} = f(\widehat{S, |S|} = k)$.

In addition, define the *generalized majority function* as

$$\mathrm{Maj}_{k,n}(x) = \mathrm{Sign}\left[\sum_{S:|S|=k} \chi_S(x)\right].$$

Note $\mathrm{Maj}_{1,n}$ is the regular majority function.

Here are some algebraic notions that will appear later in the paper:

Definition 1. *A family of functions $\mathcal{F}$ is **closed under permutation** if for all permutations $\sigma \in \Sigma_n$ (symmetric group), $f \in \mathcal{F} \implies f \circ \sigma \in \mathcal{F}$.*

Definition 2. *A family of functions $\mathcal{F}$ is **closed under negation** if for all negations $\pi \in \Pi_n \equiv \{(x_1, \ldots, x_n) \mapsto (\alpha_1 x_1, \ldots, \alpha_n x_n) : (\alpha_1, \ldots, \alpha_n) \in \{-1, 1\}^n\}$ (negation group), $f \in \mathcal{F} \implies f \circ \pi \in \mathcal{F}$.*

Definition 3. *Let $\Theta_n = \Sigma_n \circ \Pi_n = \{\theta = \sigma \circ \pi : \sigma \in \Sigma_n, \pi \in \Pi_n\}$ be **the group of permutations and negations**. Note it forms a group under composition.*

Definition 4. *Let $f \circ \Theta_n = \{f \circ \theta : \theta \in \Theta_n\}$ be the **orbit of** f, and let $\Theta_n \restriction_f = \{\theta \in \Theta_n : f \circ \theta = f\}$ be the **stabilizer group**. Let $\mathcal{F} \restriction_\theta = \{f \in \mathcal{F} : f \circ \theta = f\}$ be the **invariant set of** θ.*

Define

$$\Delta_n(x, k) = \sqrt{1 - \frac{x^2}{\binom{n}{k}}},$$

an expression that appears in many of our results. In addition, let $M_k(d)$ denote the maximum weight of a degree-d polynomial on the k-th Fourier level, the target of **Conjecture 2**:

$$M_k(d) = \max_{g \in \mathbb{F}_2^{(d,n)}} \left| \sum_{S:|S|=k} \widehat{g(S)} \right|.$$

3 Main Result

Our main result is a connection between correlation bounds against $\mathbb{F}_2^{(d,n)}$ and bounds on $\mathcal{L}_{1,k}(d)$. In particular, the goal is to use correlation bounds against symmetric functions that have near maximum $\mathcal{L}_{1,k}$ values to argue that no $g \in \mathbb{F}_2^{(d,n)}$ has a $\mathcal{L}_{1,k}(g)$ value that is too high. This is in an attempt to make progress on **conjectures 1 & 2**. Our contribution is that correlation bounds against $\mathbb{F}_2^{(d,n)}$ imply $\mathcal{L}_{1,k}(d)$ bounds:

Theorem 4 (Main Result)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric, then*

$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left[\mathrm{Cor}[f,d] - \Delta_n(\mathcal{L}_{1,k}(f), k) \right] \leq \mathcal{L}_{1,k}(d),$$
$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left[\mathrm{Cor}[f,d] + \Delta_n(\mathcal{L}_{1,k}(f), k) \right] \geq M_k(d).$$

We will present the proof for this result in **Section 5**. Notice that by simply rearranging terms, we get the following bounds on $\mathrm{Cor}(f,d)$:

Corollary 1 ($\mathcal{L}_{1,k}(d)$ bounds $\implies$ $\mathrm{Cor}(f,d)$ bounds)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric. Then,*

$$\frac{\mathcal{L}_{1,k}(f) M_k(d)}{\binom{n}{k}} - \Delta_n(\mathcal{L}_{1,k}(f), k) \leq \mathrm{Cor}[f,d] \leq \frac{\mathcal{L}_{1,k}(f)\mathcal{L}_{1,k}(d)}{\binom{n}{k}} + \Delta_n(\mathcal{L}_{1,k}(f), k).$$

4 Symmetric Functions

In this section, we lay the groundwork for how future researchers can potentially use these results and present many useful facts about symmetric polynomials.

4.1 Basics

Lemma 1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric. Then, $\forall S, S' \subseteq [n]$ such that $|S| = |S'|$, $\widehat{f(S)} = \widehat{f(S')}$.*

Proof. Suppose for the sake of contradiction that $f : \{-1,1\}^n \to \{-1,1\}$ is symmetric but $\exists S, S' \subseteq [n]$ such that $|S| = |S'|$ and $\widehat{f(S)} \neq \widehat{f(S')}$. Then,

$$\mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}[f(x)\chi_S(x)] \neq \mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}[f(x)\chi_{S'}(x)].$$

Which can be rewritten as:

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x)(\chi_S(x)-\chi_{S'}(x))] \neq 0.$$

Define $B, B' \subset \{-1,1\}^n$ such that $B = \{x \mid \chi_{S\setminus S'} = -1 \text{ and } \chi_{S'\setminus S} = 1\}$ and $B' = \{x \mid \chi_{S\setminus S'} = 1 \text{ and } \chi_{S'\setminus S} = -1\}$. Note that since $|S| = |S'|$, there exists a permutation $\sigma : \{-1,1\}^n \to \{-1,1\}^n$ such that $B = \sigma(B')$. Indeed, $|S| = |S'|$ implies $|B| = |B'|$. Then,

$$\begin{aligned} 0 &\neq \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x)(\chi_S(x)-\chi_{S'}(x))] \\ &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[-2f(x)\mathbb{1}_B + 2f(x)\mathbb{1}_{B'}] \\ &= 2\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(\sigma(x)) \mid B']\,\mathbb{P}[B'] - 2\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x) \mid B]\,\mathbb{P}[B] \\ &= 2\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x) \mid B](\mathbb{P}[B'] - \mathbb{P}[B]). \end{aligned}$$

However, since the distribution is uniform and $|B| = |B'|$, $\mathbb{P}[B'] = \mathbb{P}[B]$, establishing a contradiction. □

An immediate consequence of this lemma is that we can extend the concept of writing symmetric functions as $f : \{0, \ldots, n\} \to \{-1,1\}$ (where the input is the Hamming weight of some $x \in \{-1,1\}^n$) to the Fourier spectrum of f by defining $\widehat{f(k)} = \widehat{f(S, |S| = k)}$. Another consequence is the following useful corollary:

Corollary 2. *If $f : \{-1,1\}^n \to \{-1,1\}$ symmetric, then $\mathcal{L}_{2,k}(f) = \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}$.*

Proof. Since f is symmetric and the previous lemma:

$$\mathcal{L}_{2,k}(f) = \sum_{S:|S|=k} \widehat{f(S)}^2 = \binom{n}{k}\widehat{f(k)}^2 = \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}$$

□

Proposition 1. *For arbitrary $f : \{-1,1\}^n \to \{-1,1\}$, $\mathcal{L}_{1,k}(f) \leq \sqrt{\binom{n}{k}}$.*

Proof. By Parseval's theorem, $\|f\|_2 = 1$. Thus:

$$\max_{f:\{-1,1\}^n\to\{-1,1\}} \mathcal{L}_{1,k}(f) \leq \max_{f:\|f\|_2=1} \mathcal{L}_{1,k}(f)$$

The RHS is clearly maximized when all the Fourier weight is concentrated at the k-th level and distributed evenly. Hence:

$$\max_{f:\{-1,1\}^n\to\{-1,1\}} \mathcal{L}_{1,k}(f) \leq \sum_{S:|S|=k} \frac{1}{\sqrt{\binom{n}{k}}} = \sqrt{\binom{n}{k}}$$

□

4.2 Algebraic Properties of Symmetric Polynomials

Notice that $\Delta_n(\cdot,k)$ is minimized when $\mathcal{L}_{1,k}(f)$ is maximized. Thus, if $\mathcal{L}_{1,k}(f)=\sqrt{\binom{n}{k}}$, then $\Delta_n(\mathcal{L}_{1,k}(f),k)=0$, and our bounds in **Theorem 4** are tight. That said, the functions that achieve this equality have outputs not in $\{-1,1\}$ for most choices of n and k. In order to find the functions, $f:\{-1,1\}^n\to\{-1,1\}$, that minimize $\Delta_n(\cdot,k)$, We posit that

Conjecture 3. *Let $\mathcal{F}$ be a family of functions that is closed under permutation and negation. For each $k\in\{0,\dots,n\}$, the set of functions that maximize $\mathcal{L}_{1,k}(f)$ contains a symmetric function f generated by $f\circ\Theta_n\cup\bar{f}\circ\Theta_n$ (WLOG can replace Θ_n with Π_n).*

From here, we will use $\mathcal{F}$ to denote some subset of $\{f:\{-1,1\}^n\to\{-1,1\}\}$ that is closed under permutation and negation.

If the conjecture is true, then the family of functions that maximize $\mathcal{L}_{1,k}(\cdot)$ for some $k\in[0,n]$ can be exactly found:

Theorem 5. *If **conjecture 3** holds for $\mathcal{F}=\{f:\{-1,1\}^n\to\{-1,1\}\}$, then* $\arg\max_{f\in\mathcal{F}}\mathcal{L}_{1,k}(f)=\mathrm{Maj}_{n,k}$.

Proof. By **conjecture 3**,

$$\begin{aligned}\max_{f:\{-1,1\}^n\to\{-1,1\}}\mathcal{L}_{1,k}(f)&=\max_{f\text{ symmetric}}\mathcal{L}_{1,k}(f)\\&=\max_{f\text{ symmetric}}\sum_{S:|S|=k}|\mathop{\mathbb{E}}_x[f(x)\chi_S(x)]|.\end{aligned}$$

By **lemma 1**, the above equals

$$\begin{aligned}\max_{f\text{ symmetric}}\Big|\sum_{S:|S|=k}\mathop{\mathbb{E}}_x[f(x)\chi_S(x)]\Big|&=\max_{f\text{ symmetric}}\Big|\mathop{\mathbb{E}}_x[f(x)(\sum_{S:|S|=k}\chi_S(x))]\Big|\\&=\mathop{\mathbb{E}}_x[\Big|\sum_{S:|S|=k}\chi_S(x)\Big|].\end{aligned}$$

Thus $\arg\max_f\mathcal{L}_{1,k}(f)=\mathrm{Maj}_{k,n}=\mathrm{Sign}(\sum_{S:|S|=k}\chi_S)$. □

Since $\mathbb{F}_2^{(d,n)}$ is closed under permutation and negation, **conjecture 3** would also imply that $\mathcal{L}_{1,k}(d)$ is maximized by a symmetric function. Proving this conjecture true would be a massive step towards bounding $\mathcal{L}_{1,k}(d)$ as it would significantly simplify the problem. Consider the problem of bounding $\mathcal{L}_{1,k}(\log n)$. A sharp bound for $k=2$ or a weaker bound for $k>O(\log n)$ would imply PRGs for $\log n$-degree $\mathbb{F}_2$ polynomials which would positively answer **open question 1** [Vio22]. If **conjecture 3** holds, then we only need to bound $\mathcal{L}_{1,k}(\cdot)$ for symmetric functions, of which there are relatively few. Since symmetric functions are agnostic to all permutations of inputs, for every degree $0,\dots,d$, either all the coefficients are present or all of them are absent. There are $d+1$ such binary choices, so there are only 2^{d+1} symmetric functions in $\mathbb{F}_2^{(d,n)}$ (see **section 7.1** for more details). Thus for $d=\log n$, there are only $2n$ functions to check which is far smaller than:

$$|\mathbb{F}_2^{(\log(n),n)}| = 2^{\sum_{m=0}^{\log(n)} \binom{n}{m}}.$$

Further, the simple form symmetric functions take a when expressed as $\mathbb{F}_2$ polynomials makes it easier for degree bounds and $\mathcal{L}_{1,k}$ bounds to communicate as mathematical concepts.

We were unable to prove or disprove **conjecture 3**. However, the intuition as to why **conjecture 3** may be true stems from the following two facts:

Proposition 2.

1. *If $\exists \theta \in \Theta_n$ such that two functions f, g abide $f = g \circ \theta$, then $\mathcal{L}_{1,k}(f) = \mathcal{L}_{1,k}(g)$ for all $k \in \{0, \dots, n\}$. Moreover, $\mathcal{L}_{1,k}(f) = \mathcal{L}_{1,k}(\bar{f})$.*
2. *[FKN02] Let $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ be a unit vector. If $\|\alpha\|_1 > 1 - \delta$, then the variance of $\alpha_1, \dots, \alpha_n$ is at most 2δ.*

The first fact hints at a deeper algebraic relation between functions that have the same $\mathcal{L}_{1,k}$ weights:

Proposition 3.

1. *$f \sim g \iff \exists \theta \in \Theta_n$ such that $f = g \circ \theta$ is an equivalence relation. Thus $\mathcal{F}/\sim$ is a partition of $\mathcal{F}$ into groups of equal $\mathcal{L}_{1,k}$ weights.*
2. *If $f \sim g$, then $\mathbb{E}_{x\sim\{-1,1\}^n}[f] = \mathbb{E}_{x\sim\{-1,1\}^n}[g]$. However, the converse does not always hold.*

Thus, the equivalence relation breaks up functions first according to their number of roots and then further sub-divides them according to their $\mathcal{L}_{1,k}$ weights. We can define this in the language of a group action, $\circ : \mathcal{F} \times \Theta_n \to \mathcal{F}$. From the definition of orbit, it is clear $\mathcal{F}/\sim$ is just the set of orbits. Moreover, through this group action lens, we get several interesting results:

Proposition 4.

1. *For all $f \in \mathcal{F}$, $\Theta_n \upharpoonright_f$ is a subgroup of Θ_n.*
2. *The group action $\circ : \mathcal{F} \times \Theta_n \to \mathcal{F}$ is faithful and not transitive for $\mathcal{F} = \mathbb{F}_2^{(d,n)}$, $d \geq 1$ or $\{f : \{-1,1\}^n \to \{-1,1\}\}$.*
3. *$|\Theta_n \upharpoonright_f| \cdot |f \circ \Theta_n| = n! \cdot 2^n$.*
4. *If $f \sim g$, then $\Theta_n \upharpoonright_f$ is isomorphic to $\Theta_n \upharpoonright_g$.*
5. *$|\ \mathcal{F}/\sim\ | = \frac{1}{n!2^n} \sum_{\theta \in \Theta_n} |\ \mathcal{F} \upharpoonright_\theta\ |$.*

Though none of these results are directly applicable to proving the conjecture, we feel that understanding this behavior more deeply is an important step to proving the conjecture and answering other important problems. For one, knowing how to estimate the expected value of a representative for each class in $\mathcal{F}/\sim$ would allow you to form a DAC for all $f \in \mathcal{F}$ via **proposition 3.2**. Additionally, popular classes such as $\mathrm{AC}^0_{s,d}$, $\mathbb{F}_2^{(d,n)}$, k-CNFs, and k-DNFs, are all closed under permutation and negation. Thus, they can be described as the union of a small (at least when compared to the total number of functions) number of orbits. Most importantly, these equivalence classes have important connections to the $\mathcal{L}_{1,k}$ values achievable by a family of functions as seen in **proposition 2.1**. However, there is much more to explore in terms of this connection than we had time to do. One important open question we would like to see answered is:

Open Question 2. *Let $f, g : \{-1,1\}^n \to \{-1,1\}$. Is it true that $\mathcal{L}_{1,k}(f) = \mathcal{L}_{1,k}(g)$ for all $k \in \{0,\dots,n\}$ implies $f \sim g$ or $f \sim \bar{g}$?*

4.3 Correlation Bounds on Symmetric Polynomials

The other part of the conjecture states that the set $\arg\max_{f\in\mathcal{F}} \mathcal{L}_{1,k}(f)$ contains a symmetric function. Via Parseval's theorem, if $f : \{-1,1\}^n \to \{-1,1\}$ then $\|f\|_2 = 1$. Thus, by **proposition 2.2**, $\mathcal{L}_1(f)$ can be high if and only if the magnitudes of all the coefficients are approximately equal. Symmetric functions fit this behavior quite well (see **lemma 1**). Furthermore,

Proposition 5. $|\sum_{S:|S|=k} \widehat{f(S)}\ |$ *is maximized when f is symmetric.*

Proof.

$$\begin{aligned}|\sum_{S:|S|=k} \widehat{f(S)}| &= |\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x)(\sum_{S:|S|=k} \chi_S)]| \\ &\leq \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|\sum_{S:|S|=k} \chi_S|].\end{aligned}$$

Where exact inequality is only achieved if $f = \mathrm{Maj}_{k,n}$. □

Moreover, equality in **proposition 1** is only achieved by a symmetric function $f : \{-1,1\}^n \to \mathbb{R}$ (given that $\|f\|_2 = 1$). Thus, it seems reasonable that f symmetric maximize $\mathcal{L}_{1,k}$.

Assuming **conjecture 3** holds, the prime candidates to use our main result (**theorem 1**) on are the $\mathrm{Maj}_{k,n}$ functions (see **theorem 2**). [Smo93] and [Vio21] proved that $\mathrm{Cor}[\mathrm{Maj}_{1,n}, d] = \Theta(\frac{d}{\sqrt{n}})$. We observe that

Lemma 2. $\mathcal{L}_{1,1}(\mathrm{Maj}_{1,n}) = \sqrt{\frac{2n}{\pi}} + O(\frac{1}{\sqrt{n}})$.

Proof. First note that $\mathrm{Maj}_{1,n}$ is monotone. Hence by **proposition 2.21** in [ODo21]:

$$\widehat{\mathrm{Maj}_{1,n}}(i) = \mathrm{Inf}_i[\mathrm{Maj}_{1,n}] \equiv \Pr_{x\sim\{-1,1\}^n}[\mathrm{Maj}_{1,n}(x) \neq \mathrm{Maj}_{1,n}(x^{\oplus i})],$$

where $x^{\oplus i}$ is x with the ith bit flipped. Continuing,

$$\begin{aligned}\Pr_{x\sim\{-1,1\}^n}[\mathrm{Maj}_{1,n}(x) \neq \mathrm{Maj}_{1,n}(x^{\oplus i})] &= \frac{1}{2^n}\left(\binom{n-1}{\lfloor (n-1)/2 \rfloor} + \binom{n-1}{\lceil (n-1)/2 \rceil}\right) \\ &= \frac{1}{2^{n-1}}\binom{n-1}{(n-1)/2}.\end{aligned}$$

Applying Stirling's approximation,

$$\widehat{\mathrm{Maj}_{1,n}}(i) = \frac{1}{2^{n-1}}\binom{n-1}{(n-1)/2} = \sqrt{\frac{2}{n\pi}} + O\left(\frac{1}{n^{3/2}}\right).$$

Therefore via **lemma 1**,

$$\mathcal{L}_{1,1}(\mathrm{Maj}_{1,n}) = n \cdot \widehat{\mathrm{Maj}_{1,n}}(i) = \sqrt{\frac{2n}{\pi}} + O\left(\frac{1}{\sqrt{n}}\right).$$

□

Combining these results, we can now form an application of our main result:

Theorem 6.

$$\mathcal{L}_{1,1}(d) \geq O(d - \sqrt{n}),$$
$$M_k(d) \leq O(d + \sqrt{n}).$$

Corollary 3. *If* ***conjecture 3*** *holds, then*

$$O(d + \sqrt{n}) \geq \mathcal{L}_{1,1}(d) \geq O(d - \sqrt{n}).$$

Since the maximizer would be symmetric. As a whole, these bounds are nontrivial since the constant multiplying the $\sqrt{n}$ factor is small, however, the $\sqrt{n}$ term is still very much hurting these bounds. It is the result of the $\Delta_n(\mathcal{L}_{1,k}(f), k)$ term in **theorem 1**. Since $\lim_{n\to\infty} \mathcal{L}_{1,k}(\mathrm{Maj}_{1,k}) \neq 1$, the term does not disappear as n grows. In attempts to improve the bound, we tried to replace the $\Delta_n(\mathcal{L}_{1,k}(f), k)$ term but were unable to do so in a way that did more than improve the constant multiplying $\sqrt{n}$. We leave it as an open question to the reader whether this result can be improved enough to make the $\Delta_n(\mathcal{L}_{1,k}(f), k)$ term $o(1)$. There is indeed some hope this may be possible as [Cha+20] proved the bound $\mathcal{L}_{1,1}(d) \leq O(d)$ via a dimension counting result inspired by [Smo93].

4.4 Empirical Studies of $\mathcal{L}_{1,2}(d)$

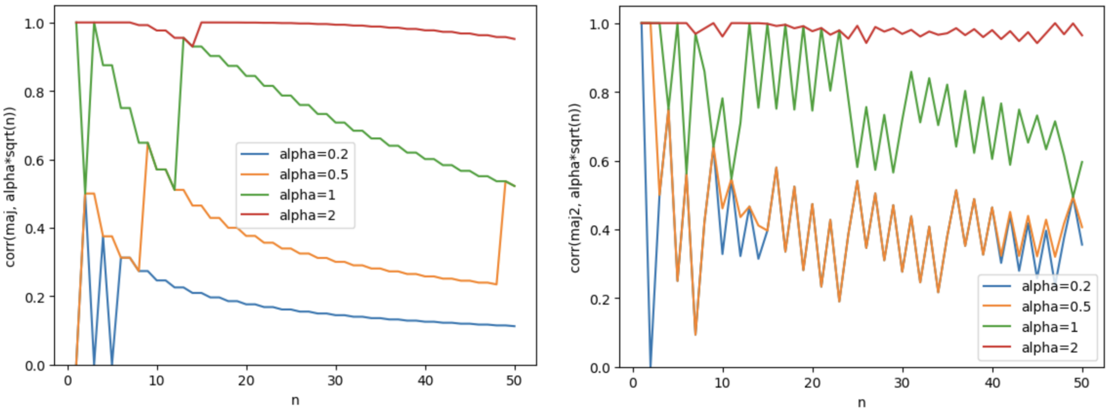
One of the earlier conjectures was that $\mathcal{L}_{1,2}(d) \leq O(d^2)$ (**Conjecture 1**). [Cha+20] proved that $\mathcal{L}_{1,1}(d) \leq O(d)$, so it seems possible that **Conjecture 1** might hold too. We studied this question from an empirical perspective by analyzing the $\mathrm{Maj}_{1,n}$ and $\mathrm{Maj}_{2,n}$ functions. By **corollary 1**,

$$\mathrm{Cor}[\mathrm{Maj}_{1,n}, d] = \Theta\left(\frac{d}{\sqrt{n}}\right),$$

and if $\mathcal{L}_{1,2}(d) \leq O(d^2)$, then

$$\mathrm{Cor}[\mathrm{Maj}_{2,n}, d] = \Theta\left(\frac{d^2}{n}\right).$$

Therefore, if **conjecture 1** is true, then correlation plots should remain steady for for $d = O(\sqrt{n})$. From empirical studies, this does appear to be the case.



(a) $\mathrm{Maj}_{1,n}$

(b) $\mathrm{Maj}_{2,n}$

Figure 1: Plots of $\mathrm{Cor}[\mathrm{Maj}, \alpha\sqrt{n}]$ against n. Notice that for larger α, both graphs stay mostly constant for all n.

5 Proof of Main Result

Recall our main result:

Theorem 3 (Main Result)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric, then*

$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)}\left[\mathrm{Cor}[f,d] - \Delta_n(\mathcal{L}_{1,k}(f),k)\right] \leq \mathcal{L}_{1,k}(d),$$
$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)}\left[\mathrm{Cor}[f,d] + \Delta_n(\mathcal{L}_{1,k}(f),k)\right] \geq M_k(d).$$

We will prove this theorem in two parts, first analyzing the lower bound, then the upper bound.

Proposition 6 (Main Result Lower Bound)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric, then*

$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)}\left(\mathrm{Cor}[f,d] - \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}}\right) \leq \max_{g \in \mathbb{F}_2^{(d,n)}} \mathcal{L}_{1,k}(g).$$

Proof. Let $g \in \mathbb{F}_2^{(d,n)}$. Using the definition of correlation,

$$\begin{aligned}
\mathrm{Cor}[f,g] &= \left| \sum_{S \subseteq [n]} \widehat{f(S)}\widehat{g(S)} \right| \\
&\leq \left| \sum_{S:|S|=k} \widehat{f(S)}\widehat{g(S)} \right| + \left| \sum_{S:|S|\neq k} \widehat{f(S)}\widehat{g(S)} \right| \\
&\leq \frac{\mathcal{L}_{1,k}(f)}{\binom{n}{k}} \left| \sum_{S:|S|=k} \widehat{g(S)} \right| + \|g'\|_2 \|f'\|_2,
\end{aligned}$$

where f' is f with the k-th Fourier level zeroed out. Using the preceding corollary and Parseval's theorem,

$$\begin{aligned}
\mathrm{Cor}[f,g] &\leq \frac{\mathcal{L}_{1,k}(f)}{\binom{n}{k}} \left| \sum_{S:|S|=k} \widehat{g(S)} \right| + \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}} \\
&\leq \frac{\mathcal{L}_{1,k}(f)}{\binom{n}{k}} \mathcal{L}_{1,k}(g) + \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}}.
\end{aligned}$$

Rearranging the terms, we get that

$$\mathcal{L}_{1,k}(g) \geq \frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left(\mathrm{Cor}[f,g] - \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}} \right).$$

Taking the max over $g \in \mathbb{F}_2^{(d,n)}$,

$$\max_{g \in \mathbb{F}_2^{(d,n)}} \mathcal{L}_{1,k}(g) \geq \frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left(\mathrm{Cor}[f,d] - \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}} \right).$$

□

Proposition 7 (Main Result Upper Bound)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be symmetric, then:*

$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left(\mathrm{Cor}[f,d] + \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}} \right) \geq \left| \sum_{S:|S|=k} \widehat{g(S)} \right|$$

Main Result Upper Bound Proof. Let $g \in \mathbb{F}_2^{(d,n)}$. Again applying the definition of correlation,

$$\begin{aligned}
\mathrm{Cor}[f,g] &= \left|\sum_{S\subseteq[n]} \widehat{f(S)}\widehat{g(S)}\right| \\
&\geq \left|\sum_{S:|S|=k} \widehat{f(S)}\widehat{g(S)}\right| - \left|\sum_{S:|S|\neq k} \widehat{f(S)}\widehat{g(S)}\right| \\
&\geq \frac{\mathcal{L}_{1_k}(f)}{\binom{n}{k}} \left|\sum_{S:|S|=k} \widehat{g(S)}\right| - \|g\|_2\|f'\|_2.
\end{aligned}$$

Using Parseval's theorem,

$$\mathrm{Cor}[f,g] \geq \frac{\mathcal{L}_{1_k}(f)}{\binom{n}{k}} \left|\sum_{S:|S|=k} \widehat{g(S)}\right| - \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}}$$

Rearranging terms, we get that

$$\frac{\binom{n}{k}}{\mathcal{L}_{1_k}(f)} \left(\mathrm{Cor}[f,g] + \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}}\right) \geq \left|\sum_{S:|S|=k} \widehat{g(S)}\right|.$$

To finish the proof, simply take the max of both sides over $g \in \mathbb{F}_2^{(d,n)}$ as in the proof of the lower bound.

□

6 Conclusion and future directions

Answering **conjectures 1 & 2** ([Cha+18b] [Cha+20]) is a tough challenge however we are hopeful that they will be proved true. Early attempts by [Vio21] to disprove them have already failed (though they did come quite close, implying that **conjecture 1** is possibly tight). After our attempts to prove them, we believe the best path forward is to analyze the functions that maximize $\mathcal{L}_{1,k}$. In particular, we believe that symmetric functions maximize $\mathcal{L}_{1,k}$ for any suitably robust class of functions. To study this problem, it will be essential to understand the transformations that preserve $\mathcal{L}_{1,k}$ i.e. $\theta \in \Theta_n$. Understanding the algebraic properties of this will give deeper insight into the meaning of $\mathcal{L}_{1,k}$ and hopefully prove **conjecture 3**. This would be a great step forward towards answering **open question 1** [Vio22] as there are only a few symmetric functions of degree less than $\log(n)$ and they all have simple representations. Moreover, the impact of proving **conjecture 3** would extend far beyond $\mathbb{F}_2$ polynomials as many classes are closed under permutation and negation. Understanding exactly which functions maximize $\mathcal{L}_{1,k}$ would surely usher in several new PRGs for various classes through [Cha+18a] polarizing random walks framework.

7 Appendix

7.1 Symmetric Functions in $\mathbb{F}_2$

Theorem 4. *There are only 2^{d+1} symmetric functions in $\mathbb{F}_2^{(d,n)}$.*

Definition 5. *Let $x^S = \bigwedge_{i\in S} x_i$, and $e_m = \bigoplus_{S:|S|=m} x^S$*

Proof. To prove the theorem, it is sufficient to show that all symmetric functions are a linear combination (over $\mathbb{F}_2$) of $e_0, e_1, \ldots, e_n$. Thus when restricted to $f \in \mathbb{F}_2^{(d,n)}$ there are only 2^{d+1} unique linear combinations of $e_0, \ldots, e_d$. Indeed, any linear combination of $e_0, e_1, \ldots, e_n$ must be symmetric since each e_m is symmetric. Moreover, this includes all symmetric functions as there are 2^{n+1} unique linear combinations and 2^{n+1} symmetric functions (there are $n+1$ possibly hamming weights and thus 2^{n+1} possible labeling schemes). Hence all symmetric functions are equivalent to a linear combination of $e_0, e_1, \ldots, e_n$.

□

7.2 Proof of Proposition 2.1

Proposition 2.

1. *If $\exists\theta \in \Theta_n$ such that two functions f, g abide $f = g\circ\theta$, then $\mathcal{L}_{1,k}(f) = \mathcal{L}_{1,k}(g)$ for all $k \in \{0, \ldots, n\}$.*
2. *[FKN02] Let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ be a unit vector. If $\|\alpha\|_1 > 1-\delta$, then the variance of $\alpha_1, \ldots, \alpha_n$ is at most 2δ.*

Proof. Since $f = g\circ\theta$:

$$f(x) = g\circ\theta = \sum_{S\subseteq[n]} \widehat{g(S)}\chi_S \circ\theta(x)$$

Thus it is sufficient to show that $\chi_S\circ\theta = \pm\chi_{S'}$ for some $S' \subseteq [n]$ such that $|S'| = |S|$. For all $\theta \in \Theta_n$, θ can be arbitrarily rewritten as $\sigma\circ\pi$ for some $\sigma \in \Sigma_n$ and $\pi \in \Pi_n$. It is clear to see that $\chi_S \circ\sigma = \chi_{\sigma(S)}$. Moreover, $\chi_S\circ\pi = \pm\chi_S$ where the sign negative if and only if π flips the sign of an odd number of variables in S. Hence, $\chi_S\circ\theta = \chi_S\circ\sigma\circ\pi = \pm\chi_{\sigma(S)}$.

□

7.3 Proof of Proposition 3

Proposition 3.

1. *$f \sim g \iff \exists\theta \in \Theta_n$ such that $f = g\circ\theta$ is an equivalence relation. Thus $\mathcal{F}/\sim$ is a partition of $\mathcal{F}$ into groups of equal $\mathcal{L}_{1,k}$ weights.*
2. *If $f \sim g$, then $\mathbb{E}_{x\sim\{-1,1\}^n}[f] = \mathbb{E}_{x\sim\{-1,1\}^n}[g]$. However, the converse does not always hold.*

Proof. First, we verify that $\sim$ satisfies the axioms of an equivalence relation:

1. (Reflexivity) $f \sim f$ since Θ_n contains an identity element.
2. (Symmetry) $f \sim g \iff f = g \circ \theta \iff f \circ \theta^{-1} = g \iff g \sim f$. Where we know $\theta^{-1} \in \Theta_n$ since if $\theta = \sigma \circ \pi$ then $\theta^{-1} = \pi^{-1} \circ \sigma^{-1} = \pi \circ \sigma^{-1}$.
3. (Transitivity) If $f \sim g, g \sim h$, then $f \circ \theta' = g = h \circ \theta \iff f \circ \theta' \circ \theta^{-1} = h \iff f \sim h$.

Now for the second statement. Since $f \sim g$, there exists a bijection $\theta \in \Theta_n$ such that $f = g \circ \theta$. Therefore, if $x \in \{-1,1\}^n$ such that $g(x) = 0$, then there exists a unique $x' = \theta^{-1}(x)$ such that $f(x') = g \circ \theta(x') = g(x) = 0$. Alternatively, $\mathbb{E}_{x\sim\{-1,1\}^n}[g] = \mathcal{L}_{1,0}(g) = \mathcal{L}_{1,0}(g \circ \theta) = \mathcal{L}_{1,0}(f) = \mathbb{E}_{x\sim\{-1,1\}^n}[f]$.

To see that the converse is not true, consider the simple counter-example:

$$f = x_1 \; g = x_1 \oplus x_2.$$

Clearly, $\mathbb{E}_{x\sim\{-1,1\}^n}[f] = \mathbb{E}_{x\sim\{-1,1\}^n}[g] = \frac{1}{2}$ yet $f \not\sim g$. □

7.4 Proof of Proposition 4

Proposition 4.

1. *For all $f \in \mathcal{F}$, $\Theta_n \restriction_f$ is a subgroup of Θ_n.*
2. *The group action $\circ : \mathcal{F} \times \Theta_n \to \mathcal{F}$ is faithful and not transitive for $\mathcal{F} = \mathbb{F}_2^{(d,n)}$, $d \geq 1$ or $\{f : \{-1,1\}^n \to \{-1,1\}\}$.*
3. *$|\Theta_n \restriction_f| \cdot |f \circ \Theta_n| = n! \cdot 2^n$.*
4. *If $f \sim g$, then $\Theta_n \restriction_f$ is isomorphic to $\Theta_n \restriction_g$.*
5. *$| \mathcal{F}/\sim | = \frac{1}{n!2^n} \sum_{\theta\in\Theta_n} | \mathcal{F} \restriction_\theta |$.*

Proof. **Proposition 4.1**:

Let $e \in \Theta_n$ be the identity element. Clearly, $\Theta_n \restriction_f$ contains e because $f \circ e = f$. If $\theta, \theta' \in \Theta_n \restriction_f$, then $f \circ \theta \circ \theta' = f$ thus $\theta \circ \theta' \in \Theta_n \restriction_f$. Finally, if $\theta \in \Theta_n \restriction_f$ then $f \circ \theta^{-1} = f \circ \theta \circ \theta^{-1} = f$ so $\theta^{-1} \in \Theta_n \restriction_f$.

Proposition 4.2: $\circ : \mathcal{F} \times \Theta_n \to \mathcal{F}$ is faithful if and only if $f \circ \theta = f$ for all $f \in \mathcal{F}$ implies $\theta = e$ (identity). Note that dictator functions $x_1, \ldots x_n$ are contained in $\mathbb{F}_2^{(d,n)} \subseteq \{f : \{-1,1\}^n \to \{-1,1\}\}$. Thus $f \circ \theta = f$ for all $f \in \mathcal{F}$ implies $x_i \circ \theta = x_i$ for all $i \in [n]$. Hence θ cannot negate nor swap any of the coordinates and therefore $\theta = e$.

$\circ : \mathcal{F} \times \Theta_n \to \mathcal{F}$ is transitive if $\forall f, g \in \mathcal{F}$ there exists $\theta \in \Theta_n$ such that $f \circ \theta = g$. However, this is not the case as $1, x_1 \in \mathcal{F}$ and there does not exist $\theta \in \Theta_n$ such that $1 \circ \theta = x_1$.

Proposition 4.3: By theorem 14.11 in [Jud21], $|f \circ \Theta_n| = [\Theta_n : \Theta_n \restriction_f]$. Thus by Lagrange's theorem,

$$|f \circ \Theta_n||\Theta_n \restriction_f| = |\Theta_n|.$$

To finish, note that $|\theta_n| = |\Sigma_n| \cdot |\Pi_n| = n! \cdot 2^n$.

Proposition 4.4: Since $f \sim g$, there exists $\theta \in \Theta_n$ such that $f = g \circ \theta$. Let $\vartheta \in \Theta_n \restriction_g$, then:

$$f \circ \theta^{-1} \circ \vartheta \circ \theta = g \circ \vartheta \circ \theta = g \circ \theta = f$$

Thus we can create the map $\Phi : \Theta_n \restriction_g \to \Theta_n \restriction_f$ such that $\Phi(\vartheta) = \theta^{-1} \circ \vartheta \circ \theta$. It is a homomorphism. Indeed:

$$\Phi(\vartheta \circ \vartheta') = \theta^{-1} \circ \vartheta \circ \vartheta' \circ \theta = \theta^{-1} \circ \vartheta \circ \theta \circ \theta^{-1} \vartheta' \circ \theta = \Phi(\vartheta) \circ \Phi(\vartheta')$$

To finish the proof, it is shown that Φ is a bijection. Suppose $\Phi(\vartheta) = \Phi(\vartheta')$ then:

$$\theta^{-1} \circ \vartheta \circ \theta = \theta^{-1} \circ \vartheta' \circ \theta \iff \vartheta = \vartheta'$$

Since θ is a bijection. To see that the map is subjective, fix some $\vartheta \in \Theta_n \restriction_f$. Then $\Phi(\theta \circ \vartheta \circ \theta^{-1}) = \vartheta$, and $\theta \circ \vartheta \circ \theta^{-1} \in \Theta_n \restriction_g$ since $g \circ \theta \circ \vartheta \circ \theta^{-1} = f \circ \vartheta \circ \theta^{-1} = f \circ \theta^{-1} = g$.

Proposition 4.5: By Burnside's Counting Theorem (see 14.3 in [Jud21]):

$$|\mathcal{F}/\sim| = \frac{1}{|\Theta_n|} \sum_{\theta \in \Theta_n} |\mathcal{F} \restriction_\theta|$$

To finish, simply apply the fact that $|\theta_n| = |\Sigma_n| \cdot |\Pi_n| = n! \cdot 2^n$. □

7.5 Proof of Theorem 6

Theorem 5 (Theorem 6 Restated)**.**

$$\mathcal{L}_{1,1}(d) \geq O(d - \sqrt{n}),$$
$$M_k(d) \leq O(d + \sqrt{n}).$$

Proof. From **theorem 4**,

$$\mathcal{L}_{1,k}(d) \geq \frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \left(\mathrm{Cor}[f, d] - \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}} \right).$$

Plugging in $k = 1$, $\text{Cor}[\text{Maj}_{1,n}, d] = \Theta(\frac{d}{\sqrt{n}})$, and the results of **lemma 2**, we get that

$$\begin{aligned}
\mathcal{L}_{1,1}(d) &\geq \frac{n}{\sqrt{\frac{2n}{\pi}} + O(\frac{1}{\sqrt{n}})}\left(\Theta(\frac{d}{\sqrt{n}}) - \sqrt{1 - \frac{(\sqrt{\frac{2n}{\pi}} + O(\frac{1}{\sqrt{n}}))^2}{n}}\right) \\
&= O(\sqrt{n})\left(\Theta(\frac{d}{\sqrt{n}}) - \sqrt{1 - \frac{(\frac{2n}{\pi} + O(1))}{n}}\right) \\
&= O(d) - O(\sqrt{n})\sqrt{1 - \frac{2}{\pi} - O\left(\frac{1}{n}\right)} \\
&\geq O(d - \sqrt{n}).
\end{aligned}$$

To prove the other inequality, start with the other bound from **theorem 1**:

$$\frac{\binom{n}{k}}{\mathcal{L}_{1,k}(f)}\left(\text{Cor}[f, d] + \sqrt{1 - \frac{\mathcal{L}_{1,k}(f)^2}{\binom{n}{k}}}\right) \geq \left|\sum_{S:|S|=k} \widehat{g(S)}\right|.$$

Again, plugging in the correlation bound for majority and $k = 1$,

$$\begin{aligned}
|M_k(d)| &\leq \frac{n}{\sqrt{\frac{2n}{\pi}} + O(\frac{1}{\sqrt{n}})}\left(\Theta(\frac{d}{\sqrt{n}}) + \sqrt{1 - \frac{(\sqrt{\frac{2n}{\pi}} + O(\frac{1}{\sqrt{n}}))^2}{n}}\right) \\
&= O(\sqrt{n})\left(\Theta(\frac{d}{\sqrt{n}}) + \sqrt{1 - \frac{2}{\pi} - O(\frac{1}{n})}\right) \\
&\leq O(d - \sqrt{n}).
\end{aligned}$$

□

7.6 $\mathcal{L}_1$ Lower Bound for $\mathbb{F}_2^{(d,n)}$

Definition 6. $\mathcal{L}_1(f) = \sum_{S\subseteq[n]} |\widehat{f(S)}|$

Theorem 7 ($\mathcal{L}_1$ Lower Bound)**.** *Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be symmetric such that $|\widehat{f(k)}| \geq |\widehat{f(i)}|$ for all $i \in [n] \setminus k$. Then,*

$$\frac{\text{Cor}[f, d]\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \leq \max_{g\in\mathbb{F}_2^{(d,n)}} \mathcal{L}_1(g).$$

Proof. Let $g \in \mathbb{F}_2^{(d,n)}$. Then,

$$\mathrm{Cor}[f,d] = \big|\sum_{S\subseteq[n]} \widehat{f(S)}\widehat{g(S)}\big| = \widehat{f(k)}\big|\sum_{S\subseteq[n]} \frac{\widehat{f(S)}}{\widehat{f(k)}}\widehat{g(S)}\big|.$$

Since $\widehat{f(k)} \geq \widehat{f(i)}$ for all $i \in [n] \setminus k$,

$$\mathrm{Cor}[f,g] \leq \widehat{f(k)}\big|\sum_{S\subseteq[n]} \widehat{g(S)}\big| \leq \frac{\mathcal{L}_{1,k}(f)\mathcal{L}_1(g)}{\frac{n}{k}},$$

so

$$\iff \frac{\mathrm{Cor}[f,g]\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \leq \mathcal{L}_1(g).$$

Maximizing both sides over $g \in \mathbb{F}_2^{(d,n)}$,

$$\frac{\mathrm{Cor}[f,d]\binom{n}{k}}{\mathcal{L}_{1,k}(f)} \leq \max_{g\in\mathbb{F}_2^{(d,n)}} \mathcal{L}_1(g).$$

□

8 Graveyard

This section is dedicated to discussing previous approaches we have tried and the difficulties/roadblocks we experienced.

8.1 Modular Root Counting

For $f \in \mathbb{F}_2^{(d,n)}$, define $N(f)$ to the number of roots of f (for $f : \{-1,1\}^n \to \{-1,1\}$, x is a root if $f(x) = 1$) and let $N_r(f) = N(f) \bmod r$. The value of $N_r(f)$ has been extensively studied from an algebraic perspective. Famously, Chevalley and Warning proved that if f is a polynomial of degree $< n$ over some field $\mathbb{F}_q$ of characteristic p then $N_p(f) = 0$ [Sch76]. This theorem was subsequently improved by Ax who showed that if $k = \lceil \frac{n-\deg(f)}{\deg(f)} \rceil$ then $N_{q^k}(f) = 0$ [Wan02]. Katz extended this result to systems of equations thus creating the Ax-Katz theorem. Others, such as Monroe and Monroe, further improved the theorem. Moreover, [Wan08] and [GGL08] both gave algorithms for computing $N_{p^k}(f)$ (in our case $p = 2$ since we consider a field of characteristic 2) that ran in polynomial time with respect to n and the number of terms in the polynomial.

We hoped that by using these efficient modular counting algorithms, we might be able to estimate $N(f)$. This notion was founded on results such as [Sch95]. In 1995, Schoof produced an algorithm that counted the roots of elliptical curves over finite fields by computing $N_r(f)$ for small primes and recovering $N(f)$ via the Chinese remainder theorem. This approach unfortunately does not carry over to the class of $\mathbb{F}_2^{(d,n)}$ as it is NP-hard to determine if $N_r(f) = 0$ when r is not a power of 2. Thus dashing hopes at recreating Schoof's algorithm. Moreover, [Wan08] and [GGL08] algorithms for computing $N_{p^k}(f)$ have exponential dependence on k making them hard to scale. This dependence was proved to be optimal by [GGL08] which is unsurprising since for $k = n$, $N_{p^k}(f) = N(f)$.

8.2 Sparsification

Sparsification is the technique of reducing the number of terms of an $\mathbb{F}_2$ polynomial while maintaining the number of roots. It remains largely unexplored in the literature of $\mathbb{F}_2$ polynomials. The closest result we were able to find was [CAO11] which discusses degree reductions that preserved the number of roots. However, the results only hold when the number of terms is less than the number of variables which is quite limiting. That said, sparsification has proved to be successful for various other models of computation. For example, [GMR12] used a modification of the celebrated sunflower lemma to sparsify DNFs, creating a state-of-the-art DAC for DNFs. However, this technique has several issues when applied to $\mathbb{F}_2$ polynomials. For one, [GMR12] relies on creating sandwich approximators by replacing a group of terms with sparse upper and lower bounding functions. This is substantially harder for $\mathbb{F}_2$ polynomials as PARITY is not monotonic unlike OR. Thus, upper and lower bounding individual terms in a $\mathbb{F}_2$ polynomial does not allow for the creation of sandwich approximators. In fact, we prove that the conditions for this to happen are incredibly strict. For a function $f : \{0,1\}^n \to \{0,1\}$, define $R_f = \{x \in \{0,1\}^n \mid f(x) = 0\}$ to be the set of roots. Then:

Theorem 8. *Let C be an $\mathbb{F}_2$ polynomial. Suppose $f = C \oplus h$ and $g = C \oplus h'$ where h, h' are non-empty conjunctions such that $h < h'$ (i.e. $h(x) \leq h'(x)$ for all $x \in \{0,1\}^n$ and $h \neq h'$), then $f \leq g \iff R_h \subseteq R_C$.*

Lemma 3. *(i) $R_{f\oplus g} = (R_f \cap R_g) \cup (R_{1\oplus f} \cap R_{1\oplus g})$.*

(ii) $R_{f\cdot g} = (R_f \cup R_g)$.

Proof of lemma 2. (i) $f \oplus g = 0 \iff f = 0, g = 0$ or $f = 1, g = 1$. Thus $R_{f\oplus g} = (R_f \cap R_g) \cup (R_f^c \cap R_g^c)$, To finish note that $R_f^c = R_{\bar{f}} = R_{1\oplus f}$.

(ii) $f \cdot g = 0 \iff f = 0$ or $g = 0$ thus $R_{f\cdot g} = R_f \cup R_g$. □

Proof of theorem 31. $(\implies)$ $R_g \subseteq R_f \iff (R_C \cap R_{h'}) \cup (R_{1\,\oplus\,C} \cap R_{1\,\oplus\,h'}) \subseteq (R_C \cap R_h) \cup (R_{1\,\oplus\,C} \cap R_{1\,\oplus\,h})$. Since R_C and $R_{1\,\oplus\,C}$ are disjoint it follows that the above implies:

(1) $(R_C \cap R_{h'}) \subseteq (R_C \cap R_h)$

(2) $(R_{1\,\oplus\,C} \cap R_{1\,\oplus\,h'}) \subseteq (R_{1\,\oplus\,C} \cap R_{1\,\oplus\,h})$

(1) holds since $h < h' \iff R_{h'} \subseteq R_h$. (2) holds if and only if $R_{1\,\oplus\,C} \subseteq R_{1\,\oplus\,h} \iff R_h \subseteq R_C$.

$(\impliedby)$ $R_h \subseteq R_C \iff C \leq h$, thus $C \leq h \leq h'$. Therefore, $f = C \oplus h = hC \oplus h = h\bar{C}$ and $g = C \oplus h' = h'C \oplus h' = h'\bar{C}$. Since $h \leq h'$, it follows that $f = h\bar{C} \leq h'\bar{C} = g$. □

In other words, replacing h with an upper bound only upper bounds the whole function if C shares all of h's roots which is incredibly unlikely in the general case. This was a problem we were able to solve via a modification of the traditional sandwich approximator framework:

Definition 7 (Hot Dog Approximators)**.** *Let $S \subseteq \{0,1\}^n$.* $\mathbf{f_l}, \mathbf{f_u}$ ***are δ-hot dog approximators of* f *on* S *if*:**

(i) $\forall x \in S$, $f_l(x) \leq f(x) \leq f_u(x)$ and $\forall x \in S^c$, $f_u(x) \leq f(x) \leq f_l(x)$

(ii) $\mathbb{E}[(f_u - f_l)\mathbb{1}_S] \leq \delta$ and $\mathbb{E}[(f_l - f_u)\mathbb{1}_{S^c}] \leq \delta$

We dubbed f_l, f_u hot dog approximators as they bound the "contents of our sandwich", f, on both sides, much like a hot dog bun. Recall f, g, and C as defined in the previous theorem. Simply knowing $h < h'$ holds, we have that $f(x) \leq g(x)$ holds on the set $S = R_C$ and $g(x) \leq f(x)$ on the set $S^c = R_{1 \oplus C}$ which makes g an upper sandwich approximator. Hence [GMR12]'s technique of replacing terms by sparse lower and upper bounds can be used to create hot dog approximators. Moreover, like sandwich approximators, fooling hot dog approximators implies fooling the original function:

Theorem 9. *Suppose f_l, f_u are δ-hot dog approximators of f on S, and let X be an ϵ-PRG for f_l, f_u. Then, X $2(\delta + \epsilon)$-fools f.*

Proof.

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f(X)\mathbb{1}_S] + \mathbb{E}[f(X)\mathbb{1}_{S^c}] \\
&\leq \mathbb{E}[f_u(X)\mathbb{1}_S] + \mathbb{E}[f_l(X)\mathbb{1}_{S^c}] \\
&\leq \mathbb{E}[f_u(\mathcal{U})\mathbb{1}_S] + \mathbb{E}[f_l(\mathcal{U})\mathbb{1}_{S^c}] + 2\epsilon \\
&\leq \mathbb{E}[f(\mathcal{U})\mathbb{1}_S] + \mathbb{E}[f(\mathcal{U})\mathbb{1}_{S^c}] + 2\epsilon + 2\delta \\
&= \mathbb{E}[f(\mathcal{U})] + 2\epsilon + 2\delta
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f(X)\mathbb{1}_S] + \mathbb{E}[f(X)\mathbb{1}_{S^c}] \\
&\geq \mathbb{E}[f_u(X)\mathbb{1}_S] + \mathbb{E}[f_l(X)\mathbb{1}_{S^c}] \\
&\geq \mathbb{E}[f_u(\mathcal{U})\mathbb{1}_S] + \mathbb{E}[f_l(\mathcal{U})\mathbb{1}_{S^c}] - 2\epsilon \\
&\geq \mathbb{E}[f(\mathcal{U})\mathbb{1}_S] + \mathbb{E}[f(\mathcal{U})\mathbb{1}_{S^c}] - 2\epsilon - 2\delta \\
&= \mathbb{E}[f(\mathcal{U})] - 2\epsilon - 2\delta
\end{aligned}$$

□

Thus, the first challenge of applying [GMR12]'s technique to $\mathbb{F}_2$ polynomials was solved. The second and larger issue was finding/proving a suitable forbidden sub-family lemma. [GMR12] uses the sunflower lemma, a forbidden sub-family style lemma, to argue that large DNFs must contain a nicely approximated sub-structure. We were unable to find such a lemma that would work for $\mathbb{F}_2$ polynomials. Even if we did find such a lemma, the problem remains far harder than in the case of DNFs since the PRGs for $\mathbb{F}_2$ polynomials have worse runtimes than those for DNFs. Therefore, the sparsification would need to be even more dramatic than that of [GMR12].

Despite the setbacks and challenges, we still believe this is a route worth further exploration in the future. In particular, we think progress could be made on finding transformations that reduce the number of terms of a polynomial while maintaining the number of roots. [CAO11] is an example of how this could be done. Furthermore, we showed that it does not take many terms to achieve all possible $N(f)$ values:

Theorem 10. *Any $N(f)$ can be achieved by a polynomial of size less than or equal to n.*

Proof. WLOG let $f : \{0,1\}^n \to \{0,1\}$. If the desired $N(f) = 2^n$, set $f = 0$. Otherwise, the desired $N(f)$ can be expressed in binary as $a_1 a_2 ... a_n$. Consider the following polynomial:

$$f(x; a_1 \dots a_n) = (1 \oplus a_1) \oplus a_n x_1 \dots x_n \oplus \bigoplus_{k \in [n-1]} (a_k \oplus a_{k+1}) x_1 \dots x_k$$

f has $N(f) = a_1 a_2 \dots a_n$ in binary. We proceed by induction on the digit. If $a_2, \dots, a_n = 0$ and $a_1 = 1$, then $N(f(x; 100 \dots 0)) = N(x_1) = 2^{n-1}$. If $a_1, \dots, a_n = 0$, then $N(f(x; 00 \dots 0)) = N(1) = 0$. Now assume that $N(f(x; a_1 \dots a_k 0 \dots 0)) = a_1 \dots a_k 0 \dots 0$, then:

$$N(f(x; a_1 \dots a_k 1 \dots 0)) = N(f(x; a_1 \dots a_k 0 \dots 0) \oplus x_1 \dots x_k \oplus x_1 \dots x_{k+1})$$
$$= N(f(x; a_1 \dots a_k 0 \dots 0) \oplus x_1 \dots x_k \bar{x}_{k+1})$$

Note that $R_f \subseteq R_{x_1 \dots x_k \bar{x}_{k+1}}$ and $R_{1 \oplus x_1 \dots x_k \bar{x}_{k+1}} \subseteq R_{1 \oplus f}$ since $x_1 \dots x_k \bar{x}_{k+1} = 1 \implies f = 1$. Thus:

$$N(f(x; a_1 \dots a_k 0 \dots 0) \oplus x_1 \dots x_k \bar{x}_{k+1}) = |(R_{x_1 \dots x_k \bar{x}_{k+1}} \cap R_f) \cup (R_{1 \oplus f}$$
$$\cap R_{1 \oplus x_1 \dots x_k \bar{x}_{k+1}})| = |R_f \cup R_{1 \oplus x_1 \dots x_k \bar{x}_{k+1}}| = |R_f| + |R_{1 \oplus x_1 \dots x_k \bar{x}_{k+1}}|$$
$$= a_1 \dots a_k 0 \dots 0 + \frac{1}{2^{k+1}} = a_1 \dots a_k 10 \dots 0$$

Thus the inductive step holds. To finish the proof, we note that f has size at most n since one term always cancels out. □

Therefore, one could potentially reduce the size of a polynomial to just n while retaining the exact number of roots which would be very nice from a derandomization perspective. From a Fourier perspective, this is the problem of reducing the degree of a polynomial as much as possible while retaining the $\mathcal{L}_1$-norm of the 0-th level.

8.3 Sensitivity-based DAC

[Dia+08] gives the first black-box randomized algorithm that is both query-efficient and time-efficient for testing if a function is a sparse $\mathbb{F}_2$ polynomial. At the heart of their argument was the claim that sparse polynomials become juntas over a small number of variables under a particular random restriction that targeted low-influence variables. This claim was the result of a simple observation:

Lemma 4. *[Dia+08] let $p : \{0,1\}^n \to \{0,1\}$ be an s-sparse polynomial. For any $\delta > 0$, there are at most $s \log(2s/\delta)$ variables x_i that have influence greater than δ.*

Proof from [Dia+08]. Any variable x_i with influence greater than δ must occur in some term of length at most $\log(2s/\delta)$. Otherwise, each occurrence of x_i would contribute less than δ/s to the influence of the i-th coordinate, and since there are at most s terms this would imply the influence of $x_i < s \cdot (\delta/s) = \delta$. Since at most $s \log(2s/\delta)$ distinct variables can occur in terms of length at most $\log(2s/\delta)$, the lemma follows. □

This suggests a simple DAC for s-sparse polynomials. Given an s-sparse polynomial, drop all terms of length greater than $\log(2s/\epsilon)$, thereby removing all the low influence variables. By the lemma, we are left with an $s\log(2s/\epsilon)$-junta that is ϵ-close to the original polynomial. Then, count the roots of the $s\log(2s/\epsilon)$-junta by testing on all $s\log(2s/\epsilon)$-bit strings. This number will be $2^n \cdot \epsilon$ close to the number of roots of the original polynomial thus giving us an ϵ-DAC. The runtime of this procedure is $O(2^{s\log(2s/\epsilon)}) = O((2s/\epsilon)^s)$ which is only non-trivial if s is near constant.

Despite the result being weak, we believe there is potential in exploiting the white-box access granted to DACs to evaluate the influence of variables and thereby estimate the number of roots. This is something we did not get a chance to fully explore but are hopeful that progress can be made.

References

[Ajt83] M. Ajtai. "Σ^1_1-Formulae on finite structures". In: *Annals of Pure and Applied Logic* 24.1 (1983), pp. 1–48. ISSN: 0168-0072. DOI: https://doi.org/10.1016/0168-0072(83)90038-6. URL: https://www.sciencedirect.com/science/article/pii/0168007283900386.

[Alo+90] N. Alon et al. "Simple construction of almost k-wise independent random variables". In: *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science.* 1990, 544–553 vol.2. DOI: 10.1109/FSCS.1990.89575.

[BV07] Andrej Bogdanov and Emanuele Viola. "Pseudorandom Bits for Polynomials". In: *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (2007), pp. 41–51. URL: https://api.semanticscholar.org/CorpusID:1142549.

[CAO11] WEI CAO. "A SPECIAL DEGREE REDUCTION OF POLYNOMIALS OVER FINITE FIELDS WITH APPLICATIONS". In: *International Journal of Number Theory* 07.04 (2011), pp. 1093–1102. DOI: 10.1142/S1793042111004277. eprint: https://doi.org/10.1142/S1793042111004277. URL: https://doi.org/10.1142/S1793042111004277.

[Cha+18a] Eshan Chattopadhyay et al. "Pseudorandom generators from polarizing random walks". In: *Proceedings of the 33rd Computational Complexity Conference.* CCC '18. San Diego, California: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. ISBN: 9783959770699.

[Cha+18b] Eshan Chattopadhyay et al. "Pseudorandom generators from the second Fourier level and applications to AC0 with parity gates". In: *Electron. Colloquium Comput. Complex.* TR18 (2018). URL: https://api.semanticscholar.org/CorpusID:52193953.

[Cha+20] Eshan Chattopadhyay et al. *Fractional Pseudorandom Generators from Any Fourier Level.* 2020. arXiv: 2008.01316 [cs.CC].

[Dia+08] Ilias Diakonikolas et al. "Efficiently Testing Sparse GF(2) Polynomials". In: *CoRR* abs/0805.1765 (2008). arXiv: 0805.1765. URL: http://arxiv.org/abs/0805.1765.

[EK90] Andrzej Ehrenfeucht and Marek Karpinski. *The Computational Complexity of (XOR, AND)-Counting Problems.* Tech. rep. 1990.

[FKN02] Ehud Friedgut, Gil Kalai, and Assaf Naor. "Boolean functions whose Fourier transform is concentrated on the first two levels". In: *Advances in Applied Mathematics* 29.3 (2002), pp. 427–437. ISSN: 0196-8858. DOI: https://doi.org/10.1016/S0196-8858(02)00024-6. URL: https://www.sciencedirect.com/science/article/pii/S0196885802000246.

[GGL08] Parikshit Gopalan, Venkatesan Guruswami, and Richard J. Lipton. "Algorithms for Modular Counting of Roots of Multivariate Polynomials". In: *Algorithmica* 50.4 (Apr. 2008), pp. 479–496. ISSN: 0178-4617.

[GMR12] Parikshit Gopalan, Raghu Meka, and Omer Reingold. "DNF Sparsification and a Faster Deterministic Counting Algorithm". In: *CoRR* abs/1205.3534 (2012). arXiv: 1205.3534. URL: http://arxiv.org/abs/1205.3534.

[GT07] Ben Green and Terence Tao. *The distribution of polynomials over finite fields, with applications to the Gowers norms.* 2007. arXiv: 0711.3191 [math.CO].

[Hås14] Johan Håstad. "On the Correlation of Parity and Small-Depth Circuits". In: *Electron. Colloquium Comput. Complex.* TR12 (2014). URL: https://api.semanticscholar.org/CorpusID:2830176.

[Hås86] J Håstad. "Almost optimal lower bounds for small depth circuits". In: *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing.* STOC '86. Berkeley, California, USA: Association for Computing Machinery, 1986, pp. 6–20. ISBN: 0897911938. DOI: 10.1145/12130.12132. URL: https://doi.org/10.1145/12130.12132.

[Jud21] T.W. Judson. *Abstract Algebra: Theory and Applications.* Orthogonal Publishing L3C, 2021. ISBN: 9781944325145. URL: https://books.google.com/books?id=_iaazgEACAAJ.

[KL90] Marek Karpinski and Michael Luby. *Approximating the Number of Solutions of a G F [2] Polynomial.* Tech. rep. 1990.

[Lov08] Shachar Lovett. "Unconditional pseudorandom generators for low degree polynomials". In: *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing.* STOC '08. Victoria, British Columbia, Canada: Association for Computing Machinery, 2008, pp. 557–562. ISBN: 9781605580470. DOI: 10.1145/1374376.1374455. URL: https://doi.org/10.1145/1374376.1374455.

[LS11] Shachar Lovett and Srikanth Srinivasan. "Correlation bounds for poly-size AC0 circuits with n1-o(1) symmetric gates". In: *Proceedings of the 14th International Workshop and 15th International Conference on Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques.* APPROX'11/RANDOM'11. Princeton, NJ: Springer-Verlag, 2011, pp. 640–651. ISBN: 9783642229343.

[LVW93] M. Luby, B. Velickovic, and A. Wigderson. "Deterministic approximate counting of depth-2 circuits". In: *[1993] The 2nd Israel Symposium on Theory and Computing Systems.* 1993, pp. 18–24. DOI: 10.1109/ISTCS.1993.253488.

[NN93] Joseph Naor and Moni Naor. "Small-Bias Probability Spaces: Efficient Constructions and Applications". In: *SIAM Journal on Computing* 22.4 (1993), pp. 838–856. DOI: 10.1137/0222053. eprint: https://doi.org/10.1137/0222053. URL: https://doi.org/10.1137/0222053.

[ODo21] Ryan O'Donnell. "Analysis of Boolean Functions". In: *CoRR* abs/2105.10386 (2021). arXiv: 2105.10386. URL: https://arxiv.org/abs/2105.10386.

[Sch76] Wolfgang M. Schmidt. "Equations over Finite Fields: An Elementary Approach". In: 1976. URL: https://api.semanticscholar.org/CorpusID:119017517.

[Sch95] René Schoof. "Counting points on elliptic curves over finite fields". en. In: *Journal de théorie des nombres de Bordeaux* 7.1 (1995), pp. 219–254. URL: http://www.numdam.org/item/JTNB_1995__7_1_219_0/.

[Smo93] R. Smolensky. "On representations by low-degree polynomials". In: *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science.* 1993, pp. 130–138. DOI: 10.1109/SFCS.1993.366874.

[ST18a] Rocco A. Servedio and Li-Yang Tan. "Improved pseudorandom generators from pseudorandom multi-switching lemmas". In: *CoRR* abs/1801.03590 (2018). arXiv: 1801.03590. URL: http://arxiv.org/abs/1801.03590.

[ST18b] Rocco A. Servedio and Li-Yang Tan. "Luby-Veličković-Wigderson revisited: Improved correlation bounds and pseudorandom generators for depth-two circuits". In: *CoRR* abs/1803.04553 (2018). arXiv: 1803.04553. URL: http://arxiv.org/abs/1803.04553.

[Vio07] Emanuele Viola. "Pseudorandom Bits for Constant-Depth Circuits with Few Arbitrary Symmetric Gates". In: *SIAM Journal on Computing* 36.5 (2007), pp. 1387–1403. DOI: 10.1137/050640941. eprint: https://doi.org/10.1137/050640941. URL: https://doi.org/10.1137/050640941.

[Vio08] Emanuele Viola. “The Sum of d Small-Bias Generators Fools Polynomials of Degree d”. In: *2008 23rd Annual IEEE Conference on Computational Complexity*. 2008, pp. 124–127. DOI: 10.1109/CCC.2008.16.

[Vio21] Emanuele Viola. “Fourier conjectures, correlation bounds, and Majority *”. In: *ICALP*. In Coll. on Automata, Languages and Programming (ICALP), 2021. Virtual conference, United Kingdom: Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Jan. 2021. DOI: 10.4230/LIPIcs.ICALP.2021.111. URL: https://hal.science/hal-04277563.

[Vio22] Emanuele Viola. “Correlation bounds against polynomials”. In: *Electron. Colloquium Comput. Complex.* TR22 (2022). URL: https://api.semanticscholar.org/CorpusID:263892365.

[Wan02] Daqing Wan. “A Chevalley-Warning Approach To P-adic Estimates Of Character Sums”. In: *Proceedings of the American Mathematical Society* 123 (May 2002). DOI: 10.2307/2160608.

[Wan08] Daqing Wan. “Modular Counting of Rational Points over Finite Fields”. In: *Foundations of Computational Mathematics* 8 (Sept. 2008), pp. 597–605. DOI: 10.1007/s10208-007-0245-y.