Explicit Nonlinear Functions beyond the Fourier bound

Swastik Kopparty* Rishabh Kothary† Shanthanu S. Rai‡

Abstract

We study the problem of constructing highly nonlinear vectorial maps $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Concretely, we want an F and an $A = A(m, n) > 0$ as small as possible, so that for every affine map $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ (of the form $L(x) = Mx + b$) we have:

$$\mathsf{agree}(F, L) := |\{x \in \mathbb{F}_2^n \mid F(x) = L(x)\}| \leq A.$$

Such questions have been studied by Nyberg [Nyb91, Nyb93], Carlet and Ding [CD04, CD07], Liu, Mesnager and Chen [LMC17], Nagy [Nag25], and Biryukov, Turecek, and Udovenko [BTU26]. There is a classical method of constructing such functions from bent-functions and Fourier analytic ideas; the best bound achievable by this method is:

$$A(m, n) = \Theta(2^{n-m} + 2^{n/2}),$$

and in particular, is never smaller than $2^{n/2}$.

In this work, we show how to construct highly nonlinear functions beyond this Fourier bound. Concretely, we show how to construct for every $\gamma > 0$, a function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $m = \mathcal{O}_\gamma(n)$, achieving

$$A(m, n) \leq (1 + \gamma)^n.$$

Surprisingly, we even achieve the same quantitative behavior for the much harder question of having low agreement with m-tuples of degree d polynomials $Q : \mathbb{F}_2^n \to \mathbb{F}_2^m$, with $m = \mathcal{O}_{\gamma,d}(n)$. Here the previously best bounds were of the form $A(m, n) = \mathcal{O}(2^{-\frac{n}{2^{d+1}}} \cdot 2^n)$ of Ben-Sasson and Kopparty [Kop10], based on Gowers-norm-type arguments.

All our results generalize to all finite fields $\mathbb{F}_q$ in place of $\mathbb{F}_2$.

Our methods are based on a new connection to classical results on counting solutions to systems of polynomial equations via algebraic methods. This connection brings us to basic questions in combinatorics, about graphs and hypergraphs with simultaneously a small number of edges and independent sets.

1 Introduction

This paper is about *highly nonlinear* functions $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, and their degree d analogues. Our main result is the first explicit construction of functions whose nonlinearity goes beyond the classical Fourier bound, and corresponding results for the degree d analogues. Our results are obtained by a new connection of this problem to (1) the basic algebraic problem of bounding the number of solutions to systems of polynomial equations, and (2) the basic combinatorial problem of bounding the number of independent sets in graphs or hypergraphs with a given number of edges.

*Department of Mathematics and Department of Computer Science, University of Toronto. Email: swastik.kopparty@utoronto.ca. Research partially supported by an NSERC Discovery Grant.

†Department of Computer Science, University of Toronto. Email: rishabh.kothary@mail.utoronto.ca. Research partially supported by an NSERC Discovery Grant.

‡School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai. Email: shanthanu.rai@tifr.res.in. Research supported by the Department of Atomic Energy, Government of India, under project number RTI4014, TCS Research Fellowship and in part by research grants from Google Research, Premji Invest and SERB.

A vectorial Boolean function is a function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Linearity properties of vectorial Boolean functions have been extensively studied in differential cryptanalysis and symmetric-key cryptography (where vectorial Boolean functions are called *substitution boxes* or *S-boxes*).

An important parameter of a vectorial Boolean function F is its *nonlinearity*[1] $\mathsf{nl}(F)$. This is defined as the minimum Hamming distance of F from affine maps $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ (namely, maps where $L(x) = Mx + b$ for a matrix M and vector b):

$$\mathsf{nl}(F) = \min_{\substack{L:\mathbb{F}_2^n \to \mathbb{F}_2^m \\ \text{affine map}}} |\{x \in \mathbb{F}_2^n \mid F(x) \neq L(x)\}|.$$

For our discussion, it will be more natural to talk about the "linearity", denoted $\mathsf{maxagree}(F, \deg 1)$ defined to equal $2^n - \mathsf{nl}(F)$, which is equivalently the smallest $A \geq 0$ such that for all affine $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$:

$$|\{x \in \mathbb{F}_2^n \mid F(x) = L(x)\}| \leq A.$$

For cryptography and pseudorandomness purposes, it is interesting to find F which have provably small $\mathsf{maxagree}(F, \deg 1)$. This (and its higher degree analogue) is the main question that our paper focuses on.

Nonlinearity of vectorial Boolean functions has been studied by many authors [Nyb91, Nyb93, CD04, CD07, CF01, LC14, LMC17, Nag25, BTU26]. In the very recent work of Biryukov, Turecek, and Udovenko [BTU26], nonlinearity plays a central role in their study of linearization of S-boxes.

For the case $m = 1$ (namely the case of Boolean functions), it is well known that the smallest that $\mathsf{maxagree}(F, \deg 1)$ can be is $\frac{1}{2} \cdot 2^n + O(2^{n/2})$. That it cannot be smaller is proved by studying the Fourier transform of F and applying Parseval's inequality. Functions which achieve equality are known as (almost) bent functions, and have been very well studied with numerous applications.

For general $m > 1$, a simple application of the probabilistic method shows that there are $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with

$$\mathsf{maxagree}(F, \deg 1) \leq \begin{cases} (1 + o(1)) \cdot \frac{1}{2^m} \cdot 2^n & m \ll n \\ \mathrm{poly}(n) & m \geq n \end{cases}.$$

and this is also known to be best possible.

In particular, once $m \geq n$, there exist F with $\mathsf{maxagree}(F, \deg 1) \leq \mathrm{poly}(n)$.

However, constructions of F with this behaviour are not known. There is a very natural and elegant approach to constructing nonlinear vectorial Boolean functions based on bent functions and Fourier analysis. This dates back to Nyberg [Nyb91],

with more recent work by [CD04, CD07, CF01, LC14, LMC17, BTU26, Nag25]. Specifically, suppose $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is such that for every nonzero $y \in \mathbb{F}_2^m$, the one-bit-output function $F_y : \mathbb{F}_2^n \to \mathbb{F}_2$ given by:

$$F_y(x) = \langle F(x), y \rangle$$

is almost bent; namely $\mathsf{maxagree}(F_y, \deg 1) \leq \frac{1}{2} \cdot 2^n + O(2^{n/2})$. Then, via the results in these papers, we have:

$$\mathsf{maxagree}(F, \deg 1) \leq \frac{1}{2^m} \cdot 2^n + O(2^{n/2}).$$

We call this the *Fourier bound* on $\mathsf{maxagree}(F, \deg 1)$. It is the best possible bound on $\mathsf{maxagree}(F, \deg 1)$ based solely on the smallness of all the $\mathsf{maxagree}(F_y, \deg 1)$.

For $m \ll n$, this is quite close to best possible, but for $m \geq n$ it is significantly suboptimal. The best upper bound on $\mathsf{maxagree}(F, \deg 1)$ that can be proved via this method is $2^{n/2}$. There are no better upper bounds known for any explicit function, and it is of great interest to develop new techniques to prove strong upper bounds on $\mathsf{maxagree}(F, \deg 1)$.

Our main result for degree 1 (see Theorem 4.12) gives explicit functions which provably go beyond the Fourier bound. Specifically, for every fixed $\gamma > 0$, for large enough n and $m \geq \Omega_\gamma(n)$, we give explicit functions $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with

$$\mathsf{maxagree}(F, \deg 1) \leq (1 + \gamma)^n.$$

[1] This specific definition is more commonly called nonlinearity of the second type, or vector nonlinearity [Nag25].

One can ask similar questions for higher degree $d = O(1)$. Here we replace affine maps $L : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with m-tuples of degree d polynomials. Specifically, define $\mathsf{maxagree}(F, \deg\, d)$ to be the smallest A such that for all m-tuples $Q = (Q_1, \ldots, Q_m)$ of polynomials (where $Q_i \in \mathbb{F}_2[X_1, \ldots, X_n]$, with $\deg(Q_i) \leq d$), we have:

$$|\{x \in \mathbb{F}_2^n \mid F(x) = Q(x)\}| \leq A.$$

For $m \ll n$, it is again known that the smallest possible value of $\mathsf{maxagree}(F, \deg\, d)$ is about $\left(\frac{1}{2^m} + o(1)\right) \cdot 2^n$, and explicit constructions of such functions are known from multiple routes: (1) the Weil bounds and the Fourier bound, or (2) the direct product theorem of Viola and Wigderson [VW08]. However for $m \geq n$, the known results get much weaker. Using ideas related to the Gowers' norm (which is a higher-degree version of Fourier analysis), Ben-Sasson and Kopparty [Kop10, Theorem 7.2.1] gave constructions of explicit functions F satisfying:

$$\mathsf{maxagree}(F, \deg\, d) \leq 2^{-n/2^{d+1}} \cdot 2^n.$$

As far as we could find, this is the best upper bound known for any $m = O(n)$ and $d \geq 2$.

Our main result for degree d (see Theorem 4.15) gives explicit functions significantly improving this. In fact, our bounds for general (constant) d even bypass the Fourier bound from the $d = 1$ case! Concretely, for every fixed $\gamma > 0$ and $d \in \mathbb{N}$, for large enough n and $m \geq \Omega_{\gamma,d}(n)$, we give explicit functions $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with:

$$\mathsf{maxagree}(F, \deg\, d) \leq (1 + \gamma)^n.$$

Our results also generalize to $\mathbb{F}_q$ in place of $\mathbb{F}_2$, achieving the same qualitative behaviour of going below the Fourier bound (see Theorem 5.14).

1.1 Techniques and Related Work

Our results are based on a combination of algebraic and combinatorial tools. The explicit functions F we construct will be computed by carefully chosen polynomial maps. The problem of bounding the agreement of these explicit functions with m-tuples of degree-d polynomials $Q : \mathbb{F}_2^n \to \mathbb{F}_2^m$ then becomes a question of bounding the number of solutions to a system of polynomial equations $F - Q = 0$. We thus need (1) a method for bounding the number of solutions to a system of polynomial equations, and (2) a choice of F so that this method will work on $F - Q$ *for all* Q.

The method in (1) will be related to *Macaulay's theorem*, which is a beautiful linear algebraic technique: it bounds the size of the set S of solutions to a system of equations by examining the vector space W of polynomial functions on S, and finding a small basis B for W using the polynomial equations defining S. This gives us an upper bound $|S| = \dim(W) \leq |B|$, where the first equality is because every function on S is a polynomial function.

To make this work in our setting, we need to choose an F appropriately. We show how to choose an F based on any $(d+1)$-uniform hypergraph $H = (V, E)$, with $|V| = n$ and $|E| = m$. Each edge E specifies a monomial in the input variables $X_1, \ldots, X_n$. The bound on the number of solutions obtained by the Macaulay theorem argument turns out to be precisely $|\mathsf{Ind}(H)|$, the number of independent sets in H. This brings us to an extremely clean and concrete combinatorial question about hypergraphs: how do we design $(d+1)$-uniform hypergraphs on n vertices with $m = O(n)$ edges so that H has few independent sets.

Already simple instantiations of this framework, taking H to be a disjoint union of constant-size complete hypergraphs, yield the main qualitative aspects of our results. For example, when $d = 1$, for every integer k we get explicit functions $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $m = \frac{k-1}{2} n$ and

$$\mathsf{maxagree}(F, \deg\, 1) \leq (k+1)^{n/k},$$

which is at most $(1 + \gamma)^n$ for large enough $k = O_\gamma(1)$. In particular, this goes beyond the Fourier bound for any $k \geq 6$.

For the case of $\mathbb{F}_q$ in place of $\mathbb{F}_2$, our techniques generalize with some modifications. Here some additional complexities are introduced by the presence of non-multilinear monomials. The resulting combinatorial

problem that needs to be solved then becomes a bit more complicated, and one has to study *multiset* analogues of edges vs a *multiset* analogue of the independent sets question for graphs and hypergraphs.

Interestingly, the explicit functions that arise from this framework turn out to have a *perfect* direct product theorem (see Lemma 4.8) for their agreement with degree d polynomials; this is in contrast to the best known general direct-product theorem of Viola and Wigderson [VW08], which is lossy.

We also remark that the Macaulay theorem bound is also known as the *footprint bound*, and has a number of interesting applications in coding theory and theoretical computer science [GH00, GMT08, STV20, ABP+25].

Another strong motivation for this work comes from the beautiful matrix construction problem of Aggarwal, Dutta, Mukherjee, Nagargoje and Obremski [ADM+25], as a route to constructing 2-source extractors. It seems like highly-nonlinear functions are very relevant to that problem, which also has a Fourier bound barrier.

1.2 Refinements

We now discuss the finer quantitative questions arising from our framework.

- For $d = 1$ (namely for constructing highly nonlinear functions), our framework asks to find *graphs* with m edges minimizing the number of independent sets. This question was solved by a difficult theorem of Sah, Sawhney, Stoner and Zhao [SSSZ19], who showed that the disjoint union of constant sized cliques is in fact the optimal choice.

- For larger d, our framework asks to find $(d+1)$-uniform hypergraphs with m edges minimizing the number of independent sets. It turns out that the disjoint union of constant size complete hypergraphs is *not* the optimal choice. For sufficiently large fixed d, one can use the *Turán hypergraphs* (which are conjectured extremal examples for the hypergraph Turán problem) to beat those constructions.

- It turns out that the Turán hypergraphs are not always optimal either. We know about an ad hoc construction of an 8-vertex 3-uniform hypergraph[2] H^* such that disjoint unions of H^* beat every other construction that we know for $d = 2$ and $m = n$.

To summarize the status of the finer quantitative questions: for $d = 1$ we understand the limits of our framework, but for $d \geq 2$ there are very interesting combinatorial questions that arise. Table 1 summarizes and compares the different explicit constructions of $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$.

Construction	**Degree** d	**Output ratio** m/n	**Agreement upper bound**
Fourier Bound	$d = 1$	≥ 1	$\mathcal{O}(2^{n/2})$
Ben-Sasson, Kopparty [Kop10]	Arbitrary d	≥ 1	$\mathcal{O}\left(2^{\left(1-\frac{1}{2^{d+1}}\right)\cdot n}\right)$
Disjoint Union of Cliques Theorem 4.12	$d = 1$	$\frac{k-1}{2}$ for arbitrary k	$(k+1)^{n/k}$
Disjoint Union of Hyper-cliques Theorem 4.15	Arbitrary $d \geq 2$	$\frac{1}{k} \cdot \binom{k}{d+1}$ for arbitrary k	$\mathcal{O}(k^{nd/k})$
Disjoint Union of Turán blocks Corollary 4.19	Arbitrary d	$\leq \frac{4}{k(d+3)} \cdot \binom{k}{d+1}$ for large enough k	$\mathcal{O}(k^{n(d+1)/k})$
Disjoint Union of H^* Lemma 4.23	$d = 2$	$=1$	$115^{n/8}$

Table 1: Comparison of different constructions of $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$

[2] See Definition 4.21 for the definition of H^*

AI use: Our framework for constructing explicit functions was developed without any AI involvement. The writing was done without any AI involvement. We used AI to search for fixed size hypergraphs with a given number of edges that optimized the number of independent sets. AI found for us the 3-uniform Turan hypergraph and gave the improved ad-hoc example with $d = 2$. We used AI for literature search, and for comments on our writeup.

Acknowledgements

We are very grateful to Divesh Aggarwal and Maciej Obremski for telling us about [ADM+25], for valuable discussions, and for encouraging us to consider the $m > n$ regime (which is the setting for our strongest results).

Organization of this paper: We begin in Section 2 by going over some useful preliminaries about polynomials, hypergraphs and Fourier analysis. Then, in Section 3, we formally define the notion of nonlinearity for vectorial boolean functions and show the existence of highly nonlinear functions. We go over the existing approach of constructing explicit nonlinear functions using Fourier analysis and comment on the limitations of the "basic Fourier bound". In Section 4, we introduce the *hypergraph framework* for constructing nonlinear and "far from degree d" functions over $\mathbb{F}_2$. In Section 5, we generalize this to do the same over $\mathbb{F}_q$. The presentation turns out to be cleaner for $\mathbb{F}_2$, wherein we show that the maximum agreement between f_H and tuples of polynomials $\bar{Q} = (Q_1, \ldots, Q_m)$ of degree at most d is equal to the number of independent sets in H. After describing the hypergraph framework, we instantiate the framework with explicit hypergraphs to construct explicit nonlinear and "far from degree d" functions which beat the Fourier bound.

2 Preliminaries

- For a prime power q, $\mathbb{F}_q$ denotes the finite field of size q.
- $[n]$ denotes the set $\{1, 2, \ldots, n\}$.
- We use $\bar{X}$ to denote the tuple $(X_1, X_2, \ldots, X_n)$, where the number of variables n will be clear from context or specified when necessary.

2.1 Polynomials and Functions

Definition 2.1. *For any $S \subset \mathbb{F}_q^n$, let $\mathbb{F}_q^S$ denote the $\mathbb{F}_q$-vector space of functions from S to $\mathbb{F}_q$.* ⌟

Note that $\dim_{\mathbb{F}_q}(\mathbb{F}_q^S) = |S|$.

Definition 2.2 (Support of Polynomials)**.** *Let $f \in \mathbb{F}_q[\bar{X}]$ such that*

$$f = \sum_{i \in \mathbb{N}^n} c_i \bar{X}^i$$

then we define

$$\mathsf{supp}(f) = \{\bar{X}^i \mid c_i \neq 0\}$$

which we call the support of f. ⌟

We denote the set of multilinear monomials of n variables by $\mathcal{M}_{mult,n}$. There is a natural correspondence between monomials $M \in \mathcal{M}_{mult,n}$ and the subsets of $[n]$. In particular for any $W \subset [n]$, the monomial corresponding to it is $\bar{X}^{\mathbf{1}_W}$ where $\mathbf{1}_W \in \{0,1\}^n$ is the characteristic vector of W. Whenever n is clear from context, we will call $\mathcal{M}_{mult,n}$ by $\mathcal{M}_{mult}$ for ease of notation.

We will require the following simple form of the Combinatorial Nullstellensatz [Alo99].

Lemma 2.3. *[Alo99, Lemma 2.1] Let $P \in \mathbb{F}[X_1, ..., X_n]$ be a non-zero polynomial such that for $i \in [n]$, $\deg_{X_i}(P) \leq t_i$. Let $S_i \subseteq \mathbb{F}$, such that $|S_i| \geq t_i + 1$ for $i \in [n]$, then there exists $(\alpha_1, ..., \alpha_n) \in S_1 \times \cdots \times S_n$ such that $P(\alpha_1, ..., \alpha_n) \neq 0$.*

2.2 Hypergraphs

An r-uniform hypergraph $H = (V, E)$ is a set of vertices V, along with a set E of r-element subsets of V. The elements of E are called the (hyper)edges of H. For the hypergraph H, we use the notation $V(H)$ to denote its vertex set, and $E(H)$ to denote its (hyper)edge set.

We next define the k-shadow of a r-uniform hypergraph, where $k \leq r$ as follows.

Definition 2.4 (k-shadow of r-uniform hypergraph)**.** *Let $H = (V, E)$ be an r-uniform hypergraph. Then for any $1 \leq k \leq r$, we define the k-shadow of H as the following k-graph $\partial_k H = (V, E_k)$ where*

$$E_k = \{\{i_1, .., i_k\} \subset V \mid \exists i_{k+1}, ..., i_r \in V \text{ such that } \{i_1, ..., i_r\} \in E\}.$$

⌟

We next define independent sets of an r-uniform hypergraph H.

Definition 2.5. *Let $H = (V, E)$ be an r-uniform hypergraph. A subset $V' \subseteq V$ is an* independent set *in H if there is no edge $e \in E$ such that $e \subseteq V'$.*

For $k \in \mathbb{N}$, we define the set of all independent sets of H of size k as

$$\mathsf{Ind}(H, k) = \{V' \subset V \mid V' \text{ is an independent set in } H \text{ and } |V'| = k\}.$$

Moreover, we define the set of all independent sets of H as

$$\mathsf{Ind}(H) = \{V' \subset V \mid V' \text{ is an independent set in } H\}.$$

⌟

Definition 2.6 (Independence Number)**.** *Let $H = (V, E)$ be an r-uniform hypergraph. Then the independence number of H, denoted as $\alpha(H)$, is the size of largest independent set in H.* ⌟

We denote the complete r-uniform hypergraph on n vertices as $K_n^{(r)}$.

2.3 Hypergraph Turán Problem

In this subsection, we will talk about the Turán(n, s, r) systems.

Definition 2.7 (Turán(n, s, r) system)**.** *Let $H = (V, E)$ be an r-uniform hypergraph on n vertices, then H is called Turán(n, s, r) system if $\alpha(H) = s - 1$. Moreover, define*

$$T(n, s, r) = \min\{|E(H)| \mid H \text{ is a Turán}(n, s, r) \text{ system over } n \text{ vertices}\},$$

which is the size of the smallest hypergraph on n vertices with independence number $s - 1$. ⌟

In the combinatorics literature it is more common to study $ex(n, s, r) = \binom{n}{r} - T(n, s, r)$, which is the problem of studying the largest r-uniform hypergraphs that avoids the complete r-uniform hypergraph on s vertices, $K_s^{(r)}$.

Let $t(s, r) = \lim_{n\to\infty} \dfrac{T(n, s, r)}{\binom{n}{r}}$. It is known that $t(s, r)$ exists and $T(n, s, r) \leq \binom{n}{r} \cdot t(s, r)$ for $n \geq s$.

For the case $r = 2$ (i.e., graphs), it follows from Turán's Theorem [Tur41] that $t(s, 2) = \frac{1}{s-1}$. For $s > r \geq 3$, determining $t(s, r)$ is a major open problem. The problem has a very rich history and we refer the readers to [Kee11] for a more comprehensive survey of the problem.

We will require the following theorem proved by Gao, Kuang, Pikhurko and Wang ([GKPW26]).

Theorem 2.8 ([GKPW26]). *Let $r \geq 2$ be fixed. For sufficiently large n, there is an explicit construction of a Turán$(n, r+1, r)$ system H with:*

$$|E(H)| \leq \left(\frac{2}{1+\lfloor\frac{r+1}{2}\rfloor} + o(1)\right) \cdot \binom{n}{r}.$$

2.4 Fourier Analysis

We introduce basic notation and recall facts from Fourier analysis in this subsection and refer to [Kop19] for a more comprehensive treatment of the topic. Let G be an abelian group. A group homomorphism $\psi : G \to \mathbb{C}^\times$ is called the character of the group G. For a finite abelian group G, all the characters themselves form a group under pointwise multiplication and this group is isomorphic to G.

For example, the characters of the underlying additive group of $\mathbb{F}_q$ are $\{\psi_a : \mathbb{F}_q \to \mathbb{C}^\times\}_{a \in \mathbb{F}_q}$ where,

$$\psi_a(x) = \omega_p^{\mathsf{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(ax)},$$

where the characteristic of $\mathbb{F}_q$ equals p, ω_p is a primitive p'th root of 1 in $\mathbb{C}$, and Tr is the field trace. (In the case of q being prime, the Tr map is simply identity). Below, we recall the Fourier Transform with respect to the additive group of $\mathbb{F}_q$ (it can be similarly defined more generally over finite abelian groups, but we won't need it).

These characters ψ_a can be viewed as vectors in $\mathbb{C}^{\mathbb{F}_q}$. We define a notion of inner product on this space,

Definition 2.9. *For two vectors $u, v \in \mathbb{C}^{\mathbb{F}_q}$, we define their inner product $\langle u, v\rangle$ as,*

$$\langle u, v\rangle = \mathbb{E}_{x \in \mathbb{F}_q}[u(x)\overline{v(x)}]$$

⌟

The characters ψ_a form an orthonormal basis for the vector space $\mathbb{C}^{\mathbb{F}_q}$.

Lemma 2.10. *Let $\{\psi_a : a \in \mathbb{F}_q\}$ be the characters of the additive group $\mathbb{F}_q$. Then,*

$$\langle \psi_a, \psi_b\rangle = \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases}$$

So, a function $F : \mathbb{F}_q \to \mathbb{C}$ can be written as a linear combination of the characters ψ_a,

$$F = \sum_{a \in \mathbb{F}_q} \widehat{F}(a)\psi_a.$$

The scalars $\widehat{F}(a) \in \mathbb{C}$ are known as the Fourier coefficients of F. Since the characters ψ_a are orthonormal, we can obtain the Fourier coefficients $\widehat{F} : \mathbb{F}_q \to \mathbb{C}$ from the following inner product,

$$\widehat{F}(a) = \langle F, \psi_a\rangle = \mathbb{E}_{x \in \mathbb{F}_q}[F(x)\overline{\psi_a(x)}].$$

Since the "change of basis" linear operation to obtain $\widehat{F}$ from F is a unitary operation, the norm of $\widehat{F}$ is equal to the norm of F. In literature, this fact is stated as the Parseval identity.

Lemma 2.11 (Parseval identity). *Let $F : \mathbb{F}_q \to \mathbb{C}$. Then,*

$$\langle F, F\rangle = \mathbb{E}_{x \in \mathbb{F}_q}[|F(x)|^2] = \sum_{a \in \mathbb{F}_q} |\widehat{F}(a)|^2$$

3 Nonlinearity, $\mathsf{maxagree}(F, \deg d)$, and the Fourier bound

In this section, we will formally define the nonlinearity of a function and its extension to larger degrees. First we define affine functions below.

Definition 3.1 (Affine Functions)**.** *We call $L : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, an $\mathbb{F}_q$ affine function if there exists an $m \times n$ matrix $M \in \mathbb{F}_q^{m \times n}$ and $b \in \mathbb{F}_q^m$ such that for any $x \in \mathbb{F}_q^n$, $L(x) = Mx + b$. We call M the associated matrix of L and b the associated vector of L.* ⌟

Next we define the nonlinearity of any function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$. The notion of nonlinearity we will use denoted by nl is known as vector nonlinearity or nonlinearity of the second type in the literature ([CD04],[CD07],[CF01],[LC14],[LMC17], [BTU26], [Nag25]).

Definition 3.2 (Nonlinearity)**.** *Let $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ be a function, and $L : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ be a $\mathbb{F}_q$ affine function, then we define the agreement of F with respect to L as*

$$\mathsf{agree}(F, L) = |\{x \in \mathbb{F}_q^n \mid F(x) = L(x)\}|.$$

We define the nonlinearity of F as

$$\mathsf{nl}(F) = q^n - \max_{\substack{L:\mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m \\ L \text{ is an } \mathbb{F}_q \text{ affine function}}} \mathsf{agree}(F, L).$$

⌟

Computing nonlinearity is computationally very difficult. [Fin14] showed that computing $\mathsf{nl}(F)$ is $\#P$-hard, when F is given as a circuit and $m = 1$.

For our purposes, it will be more convenient to talk about the "linearity", namely $q^n - \mathsf{nl}(F)$. We give this definition, along with its generalization for larger degrees below.

Definition 3.3 (Max Agreement)**.** *Let $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ be a function and $Q_1, ..., Q_m \in \mathbb{F}_q[\bar{X}]$ be polynomials of degree at most d. Then we define the agreement of F with respect to $\bar{Q} = (Q_1, ..., Q_m)$ as follows*

$$\mathsf{agree}(F, \bar{Q}) = |\{x \in \mathbb{F}_q^n \mid F(x) = (Q_1(x), ..., Q_m(x))\}|.$$

We define the max agreement of F with respect to degree d polynomials as

$$\mathsf{maxagree}(F, \deg d) = \max_{\substack{\bar{Q}=(Q_1,...Q_m) \\ Q_i \in \mathbb{F}_q[\bar{X}] \text{ and } \deg(Q_i) \leq d}} \mathsf{agree}(F, \bar{Q}).$$

⌟

Remark 3.4. *For a function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, $\mathsf{nl}(F) = q^n - \mathsf{maxagree}(F, \deg 1)$.* ⌟

Next, we state a lower bound on $\mathsf{maxagree}(F, \deg d)$ for any function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ and give its proof in Section A.

Lemma 3.5 (Interpolation Bound)**.** *For any function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$*

$$\mathsf{maxagree}(F, \deg d) \geq M(q, n, d).$$

where $M(q, n, d)$ is the number of n-variate monomials of degree at most d and individual degree at most $q - 1$.

Remark 3.6. *If $q > d$, then $M(q, n, d) = \binom{n+d}{d}$, otherwise by the principle of inclusion and exclusion:*

$$M(q, n, d) = \sum_{i=0}^{\min\{n, \lfloor \frac{d}{q} \rfloor\}} (-1)^i \cdot \binom{n}{i} \cdot \binom{n + d - i \cdot q}{n}.$$

For q, d fixed and $n \to \infty$,

$$M(q, n, d) = \Theta(n^d).$$

⌟

3.1 Existence of Hard Functions

We now state the lemma that there exists function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ that are extremely far from any tuple of degree d functions.

The proof uses the probabilistic method (indeed, a random F has this property), and can be found in Appendix B.

Lemma 3.7 (Existence of very hard function)**.** *Assume $m \geq n$, then there exists a function $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, such that*

$$\mathsf{maxagree}(F, \deg d) < \min\left(\frac{m}{m-n} \cdot M(q,n,d), \quad \mathcal{O}_q(M(q,n,d) \cdot m \cdot \log\left(m \cdot M(q,n,d)\right)\right).$$

where $M(q,n,d)$ is the number of n-variate monomials of degree at most d and individual degree at most $q-1$

Remark 3.8. *For any $\varepsilon > 0$, by setting $m = \mathcal{O}_\varepsilon(n)$, Lemma 3.7, implies that there exists $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ such that $\mathsf{maxagree}(F, \deg d) \leq (1+\varepsilon)M(q,n,d)$ which is close to the optimal (Lemma 3.5).*

For $m = n$, and $q, d, = O(1)$, we get that there exists F with $\mathsf{maxagree}(F, \deg d) \leq O(M(q,n,d) \cdot n \log(M(q,n,d)) \cdot n) = O_{q,d}(n^{d+1} \log n)$. ⌟

3.2 The Basic Fourier Bound

We now present the basic Fourier bound on nonlinearity – which is the best bound on nonlinearity that can be deduced from "smallness of Fourier coefficients of projections" considerations. The proofs here are over an arbitrary finite field $\mathbb{F}_q$.

Let $\psi : \mathbb{F}_q \to \mathbb{C}$ be a nontrivial additive character.

Lemma 3.9. *Let $F : \mathbb{F}_q^n \to \mathbb{F}_q^m$ be an arbitrary function. For $y \in \mathbb{F}_q^m$, define $F_y : \mathbb{F}_q^n \to \mathbb{C}$ by:*

$$F_y(x) = \psi\big(\langle y, F(x)\rangle\big).$$

Then for all affine functions $L : \mathbb{F}_q^n \to \mathbb{F}_q^m$:

$$\mathsf{agree}(F, L) \leq \frac{q^n}{q^m} + q^n \cdot \max_{(y,\alpha)\neq(0,0)} |\widehat{F_y}(\alpha)|,$$

where $\alpha \in \mathbb{F}_q^n$.

The proof of this lemma appears in Appendix C.

By the Parseval identity, it is well known that some non-trivial Fourier coefficient has magnitude at least $q^{-n/2}$.

Lemma 3.10. *For all $y \in \mathbb{F}_q^m \setminus \{0\}$, there exists some $\alpha \in \mathbb{F}_q^n$ such that,*

$$|\widehat{F_y}(\alpha)| \geq q^{-n/2}.$$

As a consequence of Lemma 3.10, there is no function F for which the basic Fourier bound, Lemma 3.9, can let us conclude that $\mathsf{maxagree}(F, \deg 1) \leq o(q^{n/2})$. This is the limitation of the Fourier bound that motivates this paper.

Remark 3.11. *We now give a quick sketch of why the Fourier bound (Lemma 3.9) is essentially the best possible upper bound on $\mathsf{maxagree}(F, \deg 1)$ in terms of smallness of Fourier coefficients of the F_y. Concretely, we show that there exists an F which has quite large $\mathsf{maxagree}(F, \deg 1)$, while having all Fourier coefficients of all the F_y essentially as small as possible.*

Suppose we choose $F : \mathbb{F}_q^n \to \mathbb{F}_q^m$ as follows. Pick a uniformly random subset $S \subseteq \mathbb{F}_q^n$ of size $n^2 \cdot q^{n/2}$. Define $F|_S = 0$. For all $x \in \mathbb{F}_q^n \setminus S$, choose $F(x)$ uniformly at random from $\mathbb{F}_q^m$.

Then the following hold with high probability (as $n \to \infty$):

- $\mathsf{maxagree}(F, \deg 1) \geq q^{n-m} + \Omega(n^2 \cdot q^{n/2})$.
- $\max_{(y,\alpha)\neq(0,0)} |\widehat{F}_y(\alpha)| \leq O(n^2 \cdot q^{-n/2})$.

Thus, up to $\mathrm{poly}(n)$ *factors, the Fourier bound is the best possible bound on* $\mathsf{maxagree}(F, \deg 1)$ *in terms of the smallness of Fourier coefficients of the* F_y. ⌟

To demonstrate the utility of the Fourier bound, below we give a construction of a function F for which the Fourier bound yields a pretty good bound on the linearity:

$$\mathsf{maxagree}(F, \deg 1) \leq q^{n-m} + \mathcal{O}\left(\frac{m}{n} + 1\right) \cdot q^{n/2}.$$

Example 3.12. *There is a known construction of* $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$*, where* $m \geq n/2$*, that achieves*

$$\max_{(y,\alpha)\neq(0,0)} |\widehat{F}_y(\alpha)| \leq O\left(\frac{m}{n} + 1\right) \cdot q^{-n/2}.$$

which is essentially the smallest possible.

This is achieved by the following function. Let $\beta_1, \ldots, \beta_n$ *be a basis for* $\mathbb{F}_{q^n}$ *over* $\mathbb{F}_q$*. View* $\mathbb{F}_q^n$ *as* $\mathbb{F}_{q^n}$ *via this basis.*

Let $S \subseteq \mathbb{N} \times [n]$ *be a set of size* m *such that for each* $(i, j) \in S$*, the following holds true*

1. $i \geq 2$
2. $\mathrm{char}(\mathbb{F}_q)$ *does not divide* i
3. $i \leq 2 \cdot \lceil \frac{m}{n} \rceil + 2$

Then for $x \in \mathbb{F}_{q^n}$*, define*

$$F(x) = \left(\mathsf{Tr}(\beta_j \cdot x^i)\right)_{(i,j)\in S}$$

It is a consequence of the Weil bounds for character sums that this function $F : \mathbb{F}_q^n \to \mathbb{F}_q^m$ *satisfies the above claimed bound:*

$$\max_{(y,\alpha)\neq(0,0)} |\widehat{F}_y(\alpha)| \leq O\left(\frac{m}{n} + 1\right) \cdot q^{-n/2}.$$

⌟

4 Hard Functions over $\mathbb{F}_2$ using Hypergraphs

In this section, we will define hard functions over $\mathbb{F}_2$ using hypergraphs. To bound the number of agreements of the hard function with a low degree function, we will use Macaulay's theorem (also known as the footprint bound), although our proofs won't be written in the language of Hilbert functions. This is a classical technique and can be found in [GH00],[Mac27],[BD18],[GM17],[HP98], etc.

4.1 Hypergraph Framework

We will first introduce the hypergraph framework for constructing explicit functions over $\mathbb{F}_2$ which are far from tuples of polynomials of degree at most $r-1$. Let H be an r-uniform hypergraph. Using H, we define an $\mathbb{F}_2$ function, denoted by F_H and use it to study $\mathsf{maxagree}(F_H, \deg d)$, where $d < r$. In particular, we will show that $\mathsf{maxagree}(F_H, \deg d) = |\mathsf{Ind}(H)|$.

Definition 4.1. *Let* $H = (V, E)$ *be an* r*-uniform hypergraph, where* $V = [n]$*. Let* $m = |E(H)|$ *and* $[m]$ *be indexed by the elements of* E*. Define function* $F_H : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$*, where for any* $x \in \mathbb{F}_2^n$ *and* $e = (i_1, ..., i_r) \in E(H)$*, define* $F_H(x)_e = \prod_{j\in[r]} x_{i_j}$. ⌟

Our main theorem for this section shows that the hardness of F_H is controlled by the number of independent sets in H.

Theorem 4.2. *Let $H = (V, E)$ be a r-uniform hypergraph where $V = [n]$. Then for any $1 \leq d < r$,*

$$\mathsf{maxagree}(F_H, \deg\, d) = |\mathsf{Ind}(H)|.$$

We now make some definitions to help us reason about the hardness of F_H.

Definition 4.3. *Let $H = (V, E)$ be an r-uniform hypergraph where $V = [n]$, then define the set of indestructible monomials with respect to H as*

$$\mathcal{M}_H = \{\bar{X}^{\mathbf{1}_W} \mid W \in \mathsf{Ind}(H)\}.$$

⌟

Next we define our main object of interest, the set of agreements between F_H and tuples of polynomials of degree at most $r - 1$.

Definition 4.4. *Let $H = (V, E)$ be an r-uniform hypergraph where $V = [n]$. Moreover, let $m = |E(H)|$ and $\bar{Q} = (Q_1, ..., Q_m)$ be a tuple of n-variate polynomials of degree less than r. Define*

$$S_{H,\bar{Q}} = \{x \in \mathbb{F}_2^n \mid F_H(x) = \bar{Q}(x)\},$$

to be the set of agreement between F_H and $\bar{Q}$ in $\mathbb{F}_2^n$. ⌟

Observe that $S_{H,\bar{Q}}$ is the set of common solutions to the following set of equations over $\mathbb{F}_2$:

$$\begin{aligned} &\forall e = (i_1, \ldots, i_r) \in E(H) : F_H(\bar{X})_e = Q_e(\bar{X}) \\ \iff &\forall e = (i_1, \ldots, i_r) \in E(H) : \prod_{j \in [r]} X_{i_j} = Q_e(\bar{X}) \end{aligned} \tag{4.5}$$

We give a brief overview of the main ideas for proving $\mathsf{maxagree}(F_H, \deg\, d) = |\mathsf{Ind}(H)|$ (where $d < r$), before going into the technical details.

Upper bound for $\mathsf{maxagree}(F_H, \deg\, d)$: We will upper bound the size of $S_{H,\bar{Q}}$. For this, we will study the space $W = \mathbb{F}_2^{S_{H,\bar{Q}}}$, which is the $\mathbb{F}_2$-linear vector space of all functions from $S_{H,\bar{Q}}$ to $\mathbb{F}_2$.

We will pick $|\mathsf{Ind}(H)|$ functions in W and show that they span W. This will imply that $|S_{H,\bar{Q}}| = \dim(W) \leq |\mathsf{Ind}(H)|$, as desired.

Which $|\mathsf{Ind}(H)|$ functions in W do we pick? The indestructible monomials. Why do they span W? We show this by an algorithmic procedure. First notice that any function $g \in W$ is computed by *some* polynomial $g'(\bar{X}) \in \mathbb{F}_2[\bar{X}]$. We now repeatedly "reduce" $g'(\bar{X})$ while preserving the property that $g'(\bar{X})$ computes g. The reduction step looks for a monomial in $g'(\bar{X})$ divisible by $F_H(\bar{X})_e$, and replaces $F_H(\bar{X})_e$ in that monomial by $Q_e(\bar{X})$. In Lemma 4.6, we show that the above process terminates and in the end, we obtain a polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ which is supported on the indestructible monomials $\mathcal{M}_H$, such that $\forall x \in S_{H,\bar{Q}}$, $g(x) = g'(x) = \hat{g}(\bar{x})$ (this follows from Equation 4.5). Thus, every function $g \in W$ is spanned by the indestructible monomials $\mathcal{M}_H$, as claimed.

Lower bound for $\mathsf{maxagree}(F_H, \deg\, d)$: We will show an explicit set of tuples $\bar{Q} = (Q_1, ..., Q_m)$ such that $\mathsf{agree}(f_H, \bar{Q}) \geq |\mathsf{Ind}(H)|$. The trivial choice $\bar{Q} = (0, \ldots, 0)$ works for this, since for each $W \in \mathsf{Ind}(H)$, $\mathbf{1}_W \in S_{H,\bar{Q}}$.

We formally prove below that the repeated "replacement" of monomials using Equation 4.5 in $g(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ indeed terminates, and in the end we get a polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ that is supported on the indestructible monomials $\mathcal{M}_H$.

Lemma 4.6. *Let $g(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ be a multilinear polynomial and $H = (V, E)$ be an r-uniform hypergraph where $V = [n]$. Moreover, let $m = |E(H)|$ and $\bar{Q} = (Q_1, ..., Q_m)$ be a tuple of n-variate polynomials of degree less than r. Then, there exists a multilinear polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ such that $\hat{g}(\bar{X})$ lies in the $\mathbb{F}_2$ span of $\mathcal{M}_H$ and for any $x \in S_{H,\bar{Q}}$, $g(x) = \hat{g}(x)$.*

Proof. We will prove the lemma using induction on the degree of g. For the base case, if $\deg(g) = 0$, then g is a constant polynomial and thus lies in the span of $\mathcal{M}_H$. Setting $\hat{g}(\bar{X}) = g(\bar{X})$ finishes the base case.

Without loss of generality, we can assume g is a monomial M of degree d, since the general case follows by linearity. For the inductive step, if $M \in \mathcal{M}_H$ then after setting $\hat{g}(\bar{X}) = M$, we are done. Suppose $M \notin \mathcal{M}_H$, then there exists $e = (i_1, .., i_r) \in E(H)$ such that $M = \left(\prod_{j=1}^{r} X_{i_j}\right) \cdot M'$, where M' is a monomial of degree $d - r$. Let $\widetilde{g}(\bar{X}) = Q_e(X) \cdot M'(X) \mod \langle X_i^2 - X_i \mid i \in [n]\rangle$ be the multilinear polynomial obtained by replacing $\prod_{j=1}^{r} X_{i_j}$ in M by $Q_e(X)$, and multilinearizing (which never increases degree). Then $\deg(\widetilde{g}) \leq (r-1) + (d-r) < d$. By induction, we can find $\hat{g}_1(\bar{X})$ such that $\hat{g}_1(\bar{X})$ lies in the $\mathbb{F}_2$ span of $\mathcal{M}_H$ and for any $x \in S_{H,\bar{Q}}$, $\hat{g}_1(x) = \widetilde{g}(x)$.

Since for any $x \in S_{H,\bar{Q}}$, $\prod_j x_{i_j} = Q_e(x)$, it follows by definition of $\widetilde{g}(x)$, that for any $x \in S_{H,\bar{Q}}$, $\widetilde{g}(x) = g(x)$. Hence, setting $\hat{g}(\bar{X}) = \hat{g}_1(\bar{X})$ finishes the proof. □

We now prove the main result of this section, Theorem 4.2, using Lemma 4.6 above.

Proof. Let $m = |E(H)|$ and $\bar{Q} = (Q_1, ..., Q_m)$ be a tuple of n-variate polynomials of degree at most d. We will first show that $\mathsf{maxagree}(F_H, \deg d) \leq |\mathsf{Ind}(H)|$ by proving that $|S_{H,\bar{Q}}| \leq |\mathsf{Ind}(H)|$. In order to upper bound $|S_{H,\bar{Q}}|$, we will bound the dimension of $\mathbb{F}_2^{S_{H,\bar{Q}}}$.

Take any function $g \in \mathbb{F}_2^{S_{H,\bar{Q}}}$. Extend it to a multilinear polynomial $g'(\bar{X}) \in \mathbb{F}_2[\bar{X}]$, such that for all $x \in S_{H,\bar{Q}}$, $g(x) = g'(x)$. (This can be done because of the elementary fact that every function from $\mathbb{F}_2^n \to \mathbb{F}_2$ can be expressed as a multilinear polynomial.)

By Lemma 4.6, there exists a multilinear polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_2[\bar{X}]$ such that for all $x \in S_{H,\bar{Q}}$, $g(x) = g'(x) = \hat{g}(x)$, and $\hat{g}(\bar{X})$ lies in the $\mathbb{F}_2$ span of $\mathcal{M}_H$. Since $\hat{g}(\bar{X})$ is supported on the monomials $\mathcal{M}_H$, we have shown that every function of $\mathbb{F}_2^{S_{H,\bar{Q}}}$ is spanned by monomials $\mathcal{M}_H$. This implies $\dim(\mathbb{F}_2^{S_{H,\bar{Q}}}) \leq |\mathsf{Ind}(H)|$ and hence $|S_{H,\bar{Q}}| \leq |\mathsf{Ind}(H)|$.

To show $\mathsf{maxagree}(F_H, \deg d) \geq |\mathsf{Ind}(H)|$, we need to show there exists $\bar{Q} = (Q_1, ..., Q_m)$ a tuple of n-variate polynomials of degree at most d such that $\mathsf{agree}(F_H, \bar{Q}) \geq |\mathsf{Ind}(H)|$. We will set $\hat{Q}$ to be the zero map, then we claim the following

Claim 4.7. *Let $\hat{Q}$ be the zero map, then $\mathsf{agree}(F_H, \hat{Q}) \geq |\mathsf{Ind}(H)|$.*

Proof of Claim. To prove the claim, it is enough to show that $\{\mathbf{1}_W \in \mathbb{F}_2^n \mid W \in \mathsf{Ind}(H)\} \subseteq S_{H,\hat{Q}}$ (since $\mathsf{agree}(f_h, \hat{Q}) = |S_{H,\hat{Q}}|$). Recall that every $x \in S_{H,\hat{Q}}$ satisfies the following set of equations,

$$\forall e = (i_1, \ldots, i_r) \in E(H) : \prod_{j \in [r]} x_{i_j} = \hat{Q}_e(x) = 0$$

Let $W \in \mathsf{Ind}(H)$. Then, for each edge $e = (i_1, \ldots, i_r) \in E(H)$, there exists some $j \in [r]$ such that i_j is not in W. Thus, for every edge $e = (i_1, \ldots, i_r)$, we have $\prod_{j \in [r]} x_{i_j} = 0$ for $x = \mathbf{1}_W$. Hence, $\mathbf{1}_W \in S_{H,\hat{Q}}$ and this proves the claim. □

Thus, we have proved that $\mathsf{maxagree}(F_H, \deg d) = |\mathsf{Ind}(H)|$.

□

Theorem 4.2 gives us a *Direct Product Lemma* in our hypergraph framework, which will play a crucial role in constructing explicit hard functions in Section 4.2.

Lemma 4.8 (Direct Product Lemma)**.** *Let $H = (V_1 \sqcup V_2 \sqcup \cdots \sqcup V_t, E_1 \sqcup E_2 \sqcup \cdots \sqcup E_t)$ be a disjoint union of r-uniform hypergraphs $H_1 = (V_1, E_1), H_2 = (V_2, E_2), \ldots, H_t = (V_t, E_t)$. Then,*

$$\mathsf{maxagree}(F_H, \deg\, d) = \prod_{i=1}^{t} \mathsf{maxagree}(F_{H_i}, \deg\, d)$$

where $F_{H_i} : \mathbb{F}_2^{|V_i|} \longrightarrow \mathbb{F}_2^{|E_i|}$ and $d < r$.

Proof. Since $|\mathsf{Ind}(H)| = |\mathsf{Ind}(H_1)| \cdot |\mathsf{Ind}(H_2)| \cdots |\mathsf{Ind}(H_t)|$, the lemma follows by Theorem 4.2. □

4.2 Instantiating the Hypergraphs

To construct explicit hard functions $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$, *Theorem* 4.2 motivates the following question.

Question 4.9. *What r-uniform hypergraphs on n vertices and at most m edges have the minimum number of independent sets?*

For the case of graphs, the question was solved for the case $m \leq n$, by Cutler and Radcliffe ([CR11]). Later, Sah, Sawhney, Stoner and Zhao ([SSSZ19]) in their breakthrough work, solved the problem when the degree distribution of the vertices is given. In particular, they proved the following.

Theorem 4.10 ([SSSZ19])**.** *Let $G = (V, E)$ be a graph and d_v be the degree of vertex v in G. Then*

$$|\mathsf{Ind}(G)| \geq \prod_{v \in V(G)} (d_v + 2)^{1/(d_v+1)}$$

and the equality holds if and only if G is a disjoint union of cliques.

Motivated by the above theorem, we instantiate the hypergraphs with the disjoint union of equal-sized cliques and get the following result.

Lemma 4.11 (Disjoint Union of Cliques)**.** *Let $k|n$ and $r \leq k$. Let $H_{clique,k,r}$ be the disjoint union of n/k $K_k^{(r)}$ then $|E(H_{clique,k,r})| = \binom{k}{r} \cdot \frac{n}{k}$ and for any $d < r$,*

$$\mathsf{maxagree}(F_{H_{clique}k,r}, \deg\, d) = \left(\sum_{i=0}^{r-1} \binom{k}{i} \right)^{n/k},$$

where $F_{H_{clique,k,r}} : \mathbb{F}_2^n \to \mathbb{F}_2^{|E(H_{clique,k,r})|}$.

Proof. Note that $|E(K_k^{(r)})| = \binom{k}{r}$ and thus $|E(H_{clique,k,r})| = \binom{k}{r} \cdot \frac{n}{k}$. Furthermore observe that

$$|\mathsf{Ind}(K_k^{(r)})| = \sum_{i=0}^{r-1} \binom{k}{i},$$

and thus by Lemma 4.8, for any $d < r$,

$$\mathsf{maxagree}(F_{H_{clique}k,r}, \deg\, d) = \left(\sum_{i=0}^{r-1} \binom{k}{i} \right)^{n/k}$$

which finishes the proof. □

We can construct explicit non-linear functions over $\mathbb{F}_2$ by instantiating Lemma 4.11 with the appropriate set of parameters.

Theorem 4.12 (Explicit Non-Linear Functions beyond the Fourier Bound)**.** *Let $k|n$. Suppose $G = (V, E)$ is a graph such that $|V| = n$ and G is a disjoint union of n/k cliques K_k. Then,*

$$\mathsf{maxagree}(F_G, \deg 1) = (1+k)^{n/k},$$

where $F_G : \mathbb{F}_2^n \to \mathbb{F}_2^{n(k-1)/2}$.

Proof. Instantiate Lemma 4.11 for $r = 2$ and $d = 1$. □

Remark 4.13. *In particular, when $k = n$ (i.e., $G = K_n$),*

$$\mathsf{maxagree}(F_{K_n}, \deg 1) = n + 1,$$

where $F_{K_n} : \mathbb{F}_2^n \to \mathbb{F}_2^{n(n-1)/2}$. By Lemma 3.5, any function $g : \mathbb{F}_2^n \to \mathbb{F}_2^m$ has agreement at least $(n+1)$ with affine functions. Thus, F_{K_n} achieves this optimal agreement. ⌟

Remark 4.14. *Setting $k = 6$, we obtain function $F_G : \mathbb{F}_2^n \to \mathbb{F}_2^{5n/2}$ with $\mathsf{maxagree}(F_G, \deg 1) = 7^{n/6} \approx 1.383^n$. This beats the agreement of $2^{n/2} \approx 1.414^n$ which one gets by the Fourier method (refer Section 3.2).*

In fact, for every constant c, we can choose a large enough constant k such that $\mathsf{maxagree}(F_G, \deg 1) \leq 2^{n/c}$, where $F_G : \mathbb{F}_2^n \to \mathbb{F}_2^{\mathcal{O}(kn)}$. A limitation of our (hyper)graph framework is that we cannot get a sub-exponential agreement for non-linear functions in the regime $m = O(n)$ (due to the optimality of our construction for graphs from Theorem 4.10). ⌟

We can also construct explicit functions over $\mathbb{F}_2$ which are far from tuples of degree d polynomials.

Theorem 4.15. *Let $k|n$. Suppose $H = (V, E)$ is a $d+1$-uniform hypergraph such that $|V| = n$ and H is a disjoint union of n/k cliques $K_k^{(d+1)}$. Then,*

$$\mathsf{maxagree}(F_H, \deg d) = \left(\sum_{i=0}^{d} \binom{k}{i}\right)^{n/k},$$

where $F_H : \mathbb{F}_2^n \to \mathbb{F}_2^{\frac{n}{k} \cdot \binom{k}{d+1}}$.

Proof. Instantiate Lemma 4.11 for $r = d + 1$. □

Corollary 4.16 (Functions with polynomial agreement with low degree functions)**.** *Let d be a large constant. Suppose $K_n^{(d+1)}$ be a $(d+1)$-uniform clique. Then,*

$$\mathsf{maxagree}(F_{K_n^{(d+1)}}, \deg d) = \mathcal{O}(n^d),$$

where $F_{K_n^{(d+1)}} : \mathbb{F}_2^n \to \mathbb{F}_2^{\mathcal{O}(n^{d+1})}$.

Proof. Instantiate Theorem 4.15 with $k = n$. □

Remark 4.17. *By Lemma 3.5, any function $g : \mathbb{F}_2^n \to \mathbb{F}_2^m$ has agreement at least $\Omega(n^d)$ with tuples of degree d polynomials. Thus, $F_{K_n^{(d+1)}}$ achieves this optimal agreement.* ⌟

We will next show that for a large enough constant r, the disjoint union of cliques is not the best extremal solution to Question 4.9. In particular, consider the following.

Lemma 4.18. *Let $r \geq 2$ be a fixed number. For a sufficiently large k, n where $k|n$, there exists an explicit r-uniform hypergraph $H_{Turán,k,r}$ over n vertices where*

$$|E(H_{Turán,k,r})| = \lambda \cdot \binom{k}{r} \cdot \frac{n}{k}$$

and

$$|\mathsf{Ind}(H_{Turán,k,r})| \leq \left((1-\lambda+o(1))\cdot\binom{k}{r}\right)^{n/k}.$$

Moreover,

$$\lambda \leq \left(\frac{2}{1+\lfloor\frac{r+1}{2}\rfloor}+o(1)\right).$$

Proof. Let $\hat{H}$ be the r-uniform graph over k vertices from Theorem 2.8. Let $\lambda = \frac{|\hat{H}|}{\binom{k}{r}}$, then since $\hat{H}$ is Turán$(k, r+1, r)$ system, the following holds true

$$|\mathsf{Ind}(\hat{H})| = \sum_{i=0}^{r-1}\binom{k}{i} + (1-\lambda)\cdot\binom{k}{r}$$

Theorem 2.8 also implies that

$$\lambda \leq \left(\frac{2}{1+\lfloor\frac{r+1}{2}\rfloor}+o(1)\right).$$

Let H be the disjoint union of n/k copies of $\hat{H}$, then

$$\begin{aligned}|\mathsf{Ind}(H)| &= \left(\sum_{i=0}^{r-1}\binom{k}{i} + (1-\lambda)\binom{k}{r}\right)^{n/k}\\ &\leq \left((1-\lambda-o(1))\cdot\binom{k}{r}\right)^{n/k}.\end{aligned}$$

The explicitness guarantees of the lemma follows from Theorem 2.8.

□

As a corollary of the above lemma, we can prove the following.

Corollary 4.19 (Disjoint Union of Turán$(n, r+1, r)$ systems)**.** *Let $d \geq 1$ be fixed. For large enough k, n such that $k \mid n$, then $F_{H_{Turán,k,d+1}} : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$, where $m = \lambda \cdot \binom{k}{d+1} \cdot \frac{n}{k}$ for*

$$\lambda \leq \left(\frac{2}{1+\lfloor\frac{d+2}{2}\rfloor}+o(1)\right)$$

and

$$\mathsf{maxagree}(F_{H_{Turán}k,d+1}, \deg\, d) \leq \left((1-\lambda-o(1))\cdot\binom{k}{d+1}\right)^{n/k}.$$

Proof. Instantiate Lemma 4.18 with $r = d+1$. □

We will now show that we can beat the disjoint union of clique construction for large enough constant r. In particular, let k be a large enough number and

$$\lambda \leq \left(\frac{2}{1+\lfloor\frac{r+1}{2}\rfloor}+o(1)\right)$$

be the constant from Lemma 4.18. If we set $a = \lceil k \cdot \lambda^{1/(r-1)} \rceil + r - 1$, and n be a large enough number such that $k \mid n$ and $a \mid n$, then by Lemma 4.18 and Lemma 4.11,

$$\begin{aligned}\frac{|E(H_{\text{Turán},k,r})|}{|E(H_{clique,a,r})|} &\leq \frac{\lambda \cdot \binom{k}{r} \cdot \frac{n}{k}}{\binom{a}{r} \cdot \frac{n}{a}} = \frac{\lambda \cdot \binom{k-1}{r-1}}{\binom{a-1}{r-1}} \\ &\leq \lambda \cdot \prod_{i=1}^{r-1} \frac{k-i}{a-i} \\ &< \lambda \cdot \frac{k^{r-1}}{\lambda \cdot k^{r-1}} = 1.\end{aligned}$$

In the following lemma we will compare $|\mathsf{Ind}(H_{clique,a,r})|$ and $|\mathsf{Ind}(H_{turan,k,r})|$ and show that $|\mathsf{Ind}(H_{\text{Turán},k,r})| << |\mathsf{Ind}(H_{clique,a,r})|$ for fixed r.

Lemma 4.20. *Let $r \geq 2$ be a fixed number. Let k be a large enough number that is growing and $\lambda \leq \left(\frac{2}{1+\lfloor \frac{r+1}{2} \rfloor} + o(1)\right)$ is the constant from Lemma 4.18. Let $a = \lceil k \cdot \lambda^{1/(r-1)} \rceil + (r-1)$ and n be a large enough number such that $k \mid n$ and $a \mid n$. Then*

$$|\mathsf{Ind}(H_{Turán,k,r})| \leq |\mathsf{Ind}(H_{clique,a,r})|^{\frac{r}{r-1}\lambda^{1/(r-1)}(1+o(1))}$$

Proof. For ease of notation let $H_1 = H_{\text{Turán},k,r}$ and $H_2 = H_{clique,a,r}$, then by Lemma 4.18 and Lemma 4.11, we know that

$$\begin{aligned}\frac{\log(|\mathsf{Ind}(H_1)|)}{\log(|\mathsf{Ind}(H_2)|)} &\leq \frac{a}{k} \cdot \frac{\log\left((1-\lambda+o(1)) \cdot \binom{k}{r}\right)}{\log\left(\sum_{i=0}^{r-1} \binom{a}{i}\right)} \\ &\leq \frac{a}{k} \cdot \frac{\log \binom{k}{r}}{\log \binom{a}{r-1}} \\ &\leq \frac{r}{r-1} \cdot \lambda^{1/(r-1)} \cdot (1+o(1))\end{aligned}$$

The above inequality implies

$$|\mathsf{Ind}(H_1)| \leq |\mathsf{Ind}(H_2)|^{\frac{r}{r-1} \cdot \lambda^{1/(r-1)} \cdot (1+o(1))}.$$

and hence we are done with the proof. □

It is not clear to us what extremal family of hypergraphs will answer Question 4.9.

We will now give an explicit construction of a family of sparse 3-uniform hypergraph H^* over 8 vertices with 115 independent sets.

Definition 4.21. *Let $H^* = (V^*, E^*)$ be a 3-uniform hypergraph where $V^* = \mathbb{F}_3^2 \setminus \{(0,0)\}$. E^* consists of affine lines not passing through the origin. In particular*

$$E^* = \{\{x,y,z\} \in \left(\mathbb{F}_3^2 \setminus \{(0,0)\}\right)^3 \mid \ x,y,z \text{ are distinct and } x+y+z=0\}$$

where $|E^*| = 8$. ⌟

Now we will count the number of independent sets in H^*, in the following proposition.

Proposition 4.22. $|\mathsf{Ind}(H^*)| = 115$.

Proof. Notice that any subset of size at most 2, is an independent set. Thus $|\mathsf{Ind}(H^*, 0)| = 1$, $|\mathsf{Ind}(H^*, 1)| = 8$ and $|\mathsf{Ind}(H^*, 2)| = \binom{8}{2} = 28$. The independent sets of size 3 are the non edges and thus $|\mathsf{Ind}(H^*, 3)| =$

$\binom{8}{3} - 8 = 48$.

To find $|\mathsf{Ind}(H^*, 4)|$, we count the number of sets of size 4 that does not contain an affine line. Note that any set of size 4 will contain at most 1 affine line. Thus, the number of sets of size 4 with an affine line is $8 \cdot 5 = 40$ which implies $|\mathsf{Ind}(H^*, 4)| = \binom{8}{4} - 40 = 30$.

Now we will show that any set of size 5 will contain an affine line. For the sake of contradiction, if there exists a set of 5 points with no affine line, then any unordered pair of points will give a distinct line which gives us $\binom{5}{2} = 10$ distinct lines.. Since there are only 4 possible directions for lines, then by the pigeonhole principle, there exists 3 parallel lines such that S contains two points of each line. However, this gives a contradiction as then $|S|$ needs to be at least 6. Hence $|\mathsf{Ind}(H^*, t)| = 0$ for $t \geq 5$, and thus

$$|\mathsf{Ind}(H^*)| = \sum_{i=0}^{8} |\mathsf{Ind}(H^*, i)| = 115$$

which finishes the proof. □

From this sparse uniform hypergraph with very few independent sets, we can construct hard explicit functions $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n$ using the direct product lemma (Lemma 4.8). In particular, we prove the following lemma

Lemma 4.23. *Suppose* $8|n$*, then there exists an explicit function* $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n$ *such that* $\mathsf{maxagree}(F, \deg 2) = 115^{n/8}$

Proof. Let H be the 3-uniform hypergraph consisting of $n/8$ disjoint copies of H^*. Then by Lemma 4.8, it follows that

$$\mathsf{maxagree}(f_H, \deg 2) = (\mathsf{maxagree}(F_{H^*}, \deg 2))^{n/8}.$$

Set $F = F_H$, then by Theorem 4.2 and Proposition 4.22, $\mathsf{maxagree}(F, \deg 2) = 115^{n/8}$, which finishes the proof. □

Remark 4.24. *Note that if we take the disjoint union of* $n/4$ *copies* $K_4^{(3)}$*, then the corresponding hypergraph will have* n *edges and* $11^{n/4}$ *independent sets. However if we take* $n/8$ *copies of* H^**, then the corresponding hypergraph will have* n *edges and* $115^{n/8}$ $(< 11^{n/4})$ *independent sets. This shows that disjoint union of hypercliques are not the solution to Question 4.9 in the sparse regime.* ⌟

We discuss the limitations of the hypergraph framework in Appendix D.

5 Hard Functions over $\mathbb{F}_q$ using Hypergraphs

As in Section 4, we will use the hypergraph framework to build hard functions over $\mathbb{F}_q$. But unlike the $\mathbb{F}_2$ case, the agreements will not be characterized neatly by the number of independent sets in the hypergraphs. We will define a new quantity associated to the hypergraph H, called $N_H(q)$, that will characterize the agreements over $\mathbb{F}_q$.

5.1 Hypergraph Framework

Let H be an r-uniform hypergraph. We define the $\mathbb{F}_q$ function corresponding to H, denoted by $F_{H,q}$, as follows. We first define a set of tuples $M_{H,q}$ that will be helpful in defining each output component of $F_{H,q}$.

Definition 5.1 (Components of Hypergraph Function)**.** *Let* $H = (V, E)$ *be an* r*-uniform hypergraph, where* $V = [n]$*. Let* q *be any prime power and* $e \in E(H)$*, where* $e = \{i_1, ..., i_r\}$*. Define* $M_{e,q}$ *as follows*

$$M_{e,q} = \left\{ T \in \{0, 1, ..., q-1\}^n \mid T_{|_{[n]\setminus\{i_1,...,i_r\}}} = 0, \text{ and } \sum_{j\in[r]} T_{i_j} = r \right\}$$

Further, define the set $M_{H,q} = \bigcup_{e \in E} M_{e,q}$. ⌟

Remark 5.2. *In the above definition, two distinct edges* e_1 *and* e_2 *might contain common elements in their corresponding sets* $M_{e_1,q}$ *and* $M_{e_2,q}$*:* $M_{e_1,q} \cap M_{e_2,q} \neq \phi$*. This redundancy is taken care of when we define* $M_{H,q} = \bigcup_{e \in E} M_{e,q}$. ⌟

Definition 5.3 (Hypergraph Function over $\mathbb{F}_q$)**.** *Let* $H = (V, E)$ *be an* r*-uniform hypergraph, where* $V = [n]$*. Let* q *be any prime power, and* $m_{H,q} = |M_{H,q}|$*, where we index* $[m_{H,q}]$ *by the elements of* $M_{H,q}$*. We define the function* $F_{H,q} : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^{m_{H,q}}$ *where for any* $x \in \mathbb{F}_q^n$ *and* $T \in M_{H,q}$*,*

$$F_{H,q}(x)_T = \prod_{i=1}^{n} x_i^{T_i}.$$

⌟

In Section 4, the hardness of the hypergraph function was characterized by $\mathsf{Ind}(H)$. Over $\mathbb{F}_q$, this role will be played by $N_H(q)$, which we define below.

Definition 5.4. *Let* $H = (V, E)$ *be an* r*-uniform hypergraph, where* $V = [n]$*. Let* q *be a prime power, then for any* $e \in E$*, where* $e = \{i_1, ..., i_r\}$ *define*

$$N_H(q, e) = \left\{ T \in \{0, 1, ..., q-1\}^n \mid \sum_{j \in [r]} T_{i_j} < r \right\}.$$

Also define the set $N_H(q) = \bigcap_{e \in E} N_H(q, e)$. ⌟

Remark 5.5. *Observe that when* $q = 2$*,* $N_H(2) = \mathsf{Ind}(H)$*. Hence,* $N_H(q)$ *generalizes the notion of independent sets.* ⌟

Next, we define our main object of interest, the set of agreements between $F_{H,q}$ and tuples of polynomials of degree at most $r - 1$.

Definition 5.6. *Let* $H = (V, E)$ *be an* r*-uniform hypergraph where* $V = [n]$*. Let* $\bar{Q} = (Q_1, ..., Q_{m_{H,q}})$ *be* n*-variate polynomials of degree less than* r*, where* $m_{H,q} = |M_{H,q}|$*. Then define*

$$S_{H,\bar{Q},q} = \{x \in \mathbb{F}_q^n \mid F_{H,q}(x) = \bar{Q}(x)\},$$

the set of agreement between $F_{H,q}$ *and* $\bar{Q}$ *in* $\mathbb{F}_q^n$. ⌟

Observe that $S_{H,\bar{Q},q}$ is the set of common solutions to the following set of equations over $\mathbb{F}_q$:

$$\begin{aligned} &\forall T \in M_{H,q} : F_{H,q}(\bar{X})_T = Q_T(\bar{X}) \\ \iff &\forall T \in M_{H,q} : \prod_{i \in [n]} X_i^{T_i} = Q_T(\bar{X}) \end{aligned} \tag{5.7}$$

As in Section 4, we will bound the size of $S_{H,\bar{Q},q}$ to analyze $\mathsf{maxagree}(F_{H,q}, \deg d)$ (where $d < r$). For this, we define the set of indestructible monomials with respect to the hypergraph H.

Definition 5.8 (Indestructible monomials with respect to H)**.** *The set of indestructible monomials with respect to* H *is defined as*

$$\mathcal{M}_{H,q} = \{X_1^{T_1} \cdots X_n^{T_n} \mid T = (T_1, ..., T_n) \in N_H(q)\}.$$

In particular, note that $|\mathcal{M}_{H,q}| = |N_H(q)|$. ⌟

We will now bound $\mathsf{maxagree}(F_{H,q}, \deg d)$ for $d < r$. Similar to Lemma 4.6, we first formally prove below that the repeated "replacement" of monomials using Equation 5.7 in $g(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ indeed terminates, and in the end we get a polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ that is supported on the indestructible monomials $\mathcal{M}_{H,q}$.

Lemma 5.9. *Let $g(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ be a polynomial of individual degree less than q and $H = (V, E)$ be an r-uniform hypergraph where $V = [n]$. Moreover, let $m_{H,q} = |M_{H,q}|$ and $\bar{Q} = (Q_1, ..., Q_{m_{H,q}})$ be a tuple of n-variate polynomials of degree less than r. Then, there exists a polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ of individual degree less than q such that, $\hat{g}(\bar{X})$ lies in the $\mathbb{F}_q$ span of $\mathcal{M}_{H,q}$ and for any $x \in S_{H,\bar{Q},q}$, $g(x) = \hat{g}(x)$.*

Proof. We prove the lemma by induction on the degree of g. For the base case, if $\deg(g) = 0$, then g is a constant polynomial which implies $g(\bar{X})$ lies in the $\mathbb{F}_q$ span of $\mathcal{M}_{H,q}$. Setting $\hat{g}(\bar{X}) = g(\bar{X})$ finishes the proof.

For the inductive case, assume without loss of generality, that g is a monomial M of degree d. The general case will follow by linearity. If $M \in \mathcal{M}_{H,q}$, then setting $\hat{g}(\bar{X}) = g(\bar{X})$ will finish the proof. Suppose $M \notin \mathcal{M}_{H,q}$, then there exists $T \in M_{H,q}$ such that $M = \bar{X}^T \cdot M'$ where M' is a monomial of degree $d - r$. Let $\widetilde{g}(\bar{X}) = Q_T(\bar{X}) \cdot M' \mod \langle X_i^q - X_i \mid i \in [n] \rangle$ be a polynomial of individual degree at most $q - 1$, then $\deg(\widetilde{g}) \leq r - 1 + (d - r) < d$. By induction, there exists $\hat{g}_1(\bar{X})$ such that $\hat{g}_1(\bar{X})$ lies in the $\mathbb{F}_q$ span of $\mathcal{M}_{H,q}$ and for any $x \in S_{H,\bar{Q},q}$, $\hat{g}_1(x) = \widetilde{g}(x)$.

Since for any $x \in S_{H,Q,q}$, $x^T = Q_T(x)$, it follows by the definition of $\widetilde{g}$, that $\widetilde{g}(x) = g(x)$. Hence setting $\hat{g}(\bar{X}) = \hat{g}_1(\bar{X})$ finishes the proof. □

We now prove the main result that $\mathsf{maxagree}(F_{H,q}, \deg d) \leq |N_H(q)|$ using the above Lemma 5.9.

Theorem 5.10. *Let $H = (V, E)$ be a r-uniform hypergraph where $V = [n]$ and let $1 \leq d < r$. Then,*

$$\mathsf{maxagree}(F_{H,q}, \deg d) \leq |N_H(q)|.$$

Moreover, $\mathsf{maxagree}(F_{H,q}, \deg r - 1) = |N_H(q)|$.

Proof. Let $m_{H,q} = |M_{H,q}|$ and $\bar{Q} = (Q_1, ..., Q_{m_{H,q}})$ be a tuple of n-variate polynomials of degree at most d, where $m_{H,q} = |M_{H,q}|$. We will show $|S_{H,\bar{Q},q}| \leq |N_H(q)|$ to prove $\mathsf{maxagree}(F_{H,q}, \deg d) \leq |N_H(q)|$. In order to upper bound $|S_{H,\bar{Q},q}|$, we will bound the dimension of $\mathbb{F}_q^{S_{H,\bar{Q},q}}$.

Take any function $g \in \mathbb{F}_q^{S_{H,\bar{Q},q}}$, then we can extend g to a polynomial $g'(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ of individual degree less than q, such that for any $x \in S_{H,\bar{Q},q}$, $g(x) = g'(x)$. By Lemma 5.9, there exists a polynomial $\hat{g}(\bar{X}) \in \mathbb{F}_q[\bar{X}]$ of individual degree less than q, such that for any $x \in S_{H,\bar{Q},q}$, $g(x) = g'(x) = \hat{g}(x)$, and $\hat{g}^(\bar{X})$ lies in the $\mathbb{F}_q$ span $\mathcal{M}_{H,q}$. Hence we have shown that any function in $\mathbb{F}_q^{S_{H,\bar{Q},q}}$ can be spanned by elements of $\mathcal{M}_{H,q}$, which implies $\dim(\mathbb{F}_q^{S_{H,\bar{Q},q}}) \leq |N_H(q)|$.

In order to prove that $\mathsf{maxagree}(F_{H,q}, \deg r - 1) = |N_H(q)|$, we will construct $\hat{Q} = (\hat{Q}_T)_{T \in M_{H,q}}$, where each $\hat{Q}_T$ is a $n-$variate polynomial of degree at most $r - 1$. Let $\mathbb{F}_q = \{\alpha_0, ..., \alpha_{q-1}\}$ be an enumeration of the field elements. For any $T \in M_{H,q}$, define the polynomial

$$P_T(\bar{X}) = \prod_{i=1}^{n} \prod_{j=0}^{T_i - 1} (X_i - \alpha_j)$$

and $\hat{Q}_T(\bar{X}) = \bar{X}^T - P_T(\bar{X})$. By design of $P_T(\bar{X})$, we have $\deg(\hat{Q}_T(\bar{X})) \leq r - 1$.

For any $\beta \in N_H(q)$, define $a(\beta) = (\alpha_{\beta_1}, ..., \alpha_{\beta_n}) \in \mathbb{F}_q^n$. To finish the proof, it suffices to show that for any $T \in M_{H,q}$ and $\beta \in N_H(q)$, $P_T(a(\beta)) = 0$ since it would imply that $\mathsf{agree}(F_{H,q}, \hat{Q}) \geq |N_H(q)|$. Take any $\beta \in N_H(q)$, $T \in M_{H,q}$ and take edge $e \in E$ such that $T \in M_{e,q}$. Let $e = (i_1, ...i_r)$, then there exists $j \in [r]$ such that $\beta_{i_j} < T_{i_j}$ by definition of $N_H(q)$. Hence $a(\beta)_{i_j} \in \{\alpha_0, ..., \alpha_{T_{i_j} - 1}\}$, and thus $P_T(a(\beta)) = 0$. □

Remark 5.11. *The polynomial $P_T(\bar{X})$ in the proof of Theorem 5.10 also appears in [BD18],[GM17],[HP98].* ⌟

As in Section 4, Theorem 5.10 gives us a *Direct Product Lemma* in our hypergraph framework, which will play a crucial role in constructing explicit hard functions in Section 5.2.

Lemma 5.12 (Direct Product Lemma)**.** *Let $H = (V_1 \sqcup V_2 \sqcup \cdots \sqcup V_t, E_1 \sqcup E_2 \sqcup \cdots \sqcup E_t)$ be a disjoint union of r-uniform hypergraphs $H_1 = (V_1, E_1), H_2 = (V_2, E_2), \ldots, H_t = (V_t, E_t)$. Then,*

$$\mathsf{maxagree}(F_{H,q}, \deg\, d) \leq \prod_{i=1}^{t} |N_{H_i}(q)|,$$

where $F_{H_i,q} : \mathbb{F}_q^{|V_i|} \longrightarrow \mathbb{F}_q^{|M_{H_i,q}|}$ and $d < r$. In particular, when $d = r-1$, we have,

$$\mathsf{maxagree}(F_{H,q}, \deg\, r-1) = \prod_{i=1}^{t} \mathsf{maxagree}(F_{H_i,q}, \deg\, r-1).$$

Proof. By definition $N_H(q) = N_{H_1}(q) \times \cdots \times N_{H_t}(q)$, which implies

$$|N_H(q)| = |N_{H_1}(q)| \cdots |N_{H_t}(q)|$$

and thus the lemma holds true by Theorem 5.10. □

5.2 Instantiating the Hypergraph

To construct explicit hard functions $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, Theorem 5.10 , motivates the following question.

Question 5.13. *For a fixed integer $q \geq 2$, what r-uniform hypergraph H on n vertices and $|M_{H,q}| \leq m$, has the smallest value of $|N_H(q)|$?*

If we restrict the question to graphs without isolated vertices, then this question is the same Question 4.9. Just like the case with hypergraph functions over $\mathbb{F}_2$, we will instantiate the hypergraph with disjoint union of equal sized cliques to get explicit functions that go beyond the Fourier Bound

Theorem 5.14 (Explicit Functions beyond the fourier bound)**.** *Let $k|n$ and $r \leq k$. Let $H_{clique}(k,r)$ be the disjoint union of n/k $K_k^{(r)}$, then*

$$|M_{H_{clique}(k,r),q}| = \frac{n}{k} \cdot \left(\sum_{i=0}^{\min\{k, \lfloor \frac{r}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{r+k-1-iq}{k-1} \right)$$

and for any $d < r$,

$$\mathsf{maxagree}(F_{H_{clique}(k,r),q}, \deg\, d) \leq \left(\sum_{i=0}^{\min\{k, \lfloor \frac{r-1}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{k+r-1-iq}{k} \right)^{n/k}$$

where $F_{H_{clique}(k,r),q} : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^{|M_{H_{clique}(k,r),q}|}$. Moreover, we have

$$\mathsf{maxagree}(F_{H_{clique}(k,r),q}, \deg\, r-1) = \left(\sum_{i=0}^{\min\{k, \lfloor \frac{r-1}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{k+r-1-iq}{k} \right)^{n/k}.$$

Proof. Let $\hat{H} = K_k^{(r)}$, then $|M_{\hat{H},q}|$ is the number of integral solutions of the form $X_1 + \cdots + X_k = r$, where $0 \leq X_i \leq q-1$ for $i \in [k]$. By the inclusion and exclusion principle, we get

$$|M_{\hat{H},q}| = \sum_{i=0}^{\min\{k, \lfloor \frac{r}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{r+k-1-iq}{k-1}$$

Since $H_{clique}(k,r)$ is the disjoint union of n/k copies of $\hat{H}$, thus

$$|M_{H_{clique}(k,r),q}| = \frac{n}{k} \cdot \left(\sum_{i=0}^{\min\{k, \lfloor \frac{r}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{r+k-1-iq}{k-1} \right).$$

Next note that $|N_{\hat{H}}(q)|$ is the number of solutions of the form $X_1 + \cdots + X_k < r$, where $0 \leq X_i \leq q-1$ for $i \in [k]$. By the inclusion and exclusion principle, we get

$$|N_{\hat{H}}(q)| = \left(\sum_{i=0}^{\min\{k, \lfloor \frac{r-1}{q} \rfloor\}} (-1)^i \cdot \binom{k}{i} \cdot \binom{k+r-1-iq}{k} \right).$$

Hence, the proof follows by Lemma 5.12. □

As a corollary, for $q > r$, we get the following result.

Corollary 5.15. *Let $k|n$ and $r \leq k$ and $r < q$. Let $H_{clique,k,r}$ be the disjoint union of n/k $K_k^{(r)}$, then*

$$|M_{H_{clique}(k,r),q}| = \binom{k+r-1}{r} \cdot \frac{n}{k}$$

and for any $d < r$,

$$\mathsf{maxagree}(F_{H_{clique}(k,r),q}, \deg\, d) \leq \binom{k+r-1}{r-1}^{n/k}.$$

Moreover, we have

$$\mathsf{maxagree}(F_{H_{clique}(k,r),q}, \deg\, r-1) = \binom{k+r-1}{r-1}^{n/k}.$$

For the case $q \geq r$, we will show that disjoint union of cliques are not the best extremal solution to Question 5.13.

Lemma 5.16. *Let $r \geq 2$ be a fixed number and $r < q$. For sufficiently large k, n where $k \mid n$, there exists an explicit r-uniform hypergraph $H_{Turán}(k,r)$ over n vertices where*

$$|M_{H_{Turán}(k,r),q}| = \lambda \cdot \frac{n}{k} \cdot \binom{k+r-1}{r}$$

and

$$|N_{H_{Turán}(k,r)}(q)| \leq \binom{k}{r}^{n/k}$$

Moreover,

$$\lambda \leq \left(\frac{2}{1 + \lfloor \frac{r+1}{2} \rfloor} + o(1) \right).$$

The proof of the above lemma can be found in Appendix E. We get the following corollary from the above lemma.

Corollary 5.17. *Let $r \geq 2$ be a fixed number and $r < q$. For sufficiently large k, n where $k|n$, $F_{H_{Turán}(k,r),q} : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, where*

$$m \leq \left(\frac{2}{1 + \lfloor \frac{r+1}{2} \rfloor} + o(1) \right) \cdot \binom{k+r-1}{r} \cdot \frac{n}{k}$$

and for any $d \leq r-1$,

$$\mathsf{maxagree}(F_{H_{Turán}(k,r),q}, \deg\, d) \leq \binom{k}{r}^{n/k}$$

Let k be a large enough number and let λ be the constant from Lemma 5.16. Let $r \geq 2$ be a fixed number such that $q > r$. If we set $a = \lceil (2\lambda)^{1/(r-1)} \cdot (k+r-1) \rceil$, and n be a large enough number such that $k \mid n$ and $a \mid n$, then by Theorem 5.14 and Lemma 5.16,

$$\begin{aligned}\frac{|M_{H_{\text{Turán}}(k,r),q}|}{|M_{H_{clique}(a,r),q}|} &= \lambda \cdot \frac{a}{k} \cdot \frac{\binom{k+r-1}{r}}{\binom{a+r-1}{r}} \\ &\leq \lambda \cdot \frac{a}{k} \cdot \frac{(k+r-1)^r}{a^r} \\ &\leq \lambda \cdot \frac{k+r-1}{k} \cdot \frac{1}{2\lambda} < 1.\end{aligned}$$

In the following lemma, we compare $|N_{H_{\text{Turán}}(k,r)}(q)|$ and $|N_{H_{clique}(a,r)}(q)|$ and show that $|N_{H_{\text{Turán}}(k,r)}(q)| << |N_{H_{clique}(a,r)}(q)|$ for a large enough constant r.

Lemma 5.18. *Let $r \geq 2$ be a fixed number and $q > r$. Let k be a large enough number that is growing and $\lambda \leq \left(\frac{2}{1+\lfloor \frac{r+1}{2} \rfloor} + o(1) \right)$ be the constant from Lemma 5.16. Let $a = \lceil (2\lambda)^{1/(r-1)} \cdot (k+r-1) \rceil$ and n be a large enough number such that $k \mid n$ and $a \mid n$. Then*

$$|N_{H_{Turán}(k,r)}(q)| \leq |N_{H_{clique}(a,r)}(q)|^{\frac{r}{r-1} \cdot (2\lambda)^{1/(r-1)} \cdot (1+o(1))}$$

Proof. For ease of notation let $H_1 = H_{\text{Turán}}(k,r)$ and $H_2 = H_{clique}(a,r)$. Then

$$\begin{aligned}\frac{\log |N_{H_1}(q)|}{\log |N_{H_2}(q)|} &\leq \frac{a}{k} \cdot \frac{\log \binom{k}{r}}{\log \binom{a+r-1}{r-1}} \\ &\leq \frac{r}{r-1} \cdot (2\lambda)^{1/(r-1)} \cdot (1+o(1))\end{aligned}$$

which implies

$$|N_{H_1}(q)| \leq |N_{H_2}(q)|^{\frac{r}{r-1} \cdot (2\lambda)^{1/(r-1)} \cdot (1+o(1))}$$

which finishes the proof. □

It is not clear to us what is the optimal extremal family of r-uniform hypergraphs for Question 5.13. We discuss the limitations of the hypergraph framework in Appendix D.

References

[ABP+25] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan. Low degree local correction over the boolean cube. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5504–5511. SIAM, 2025. 4

[ADM+25] Divesh Aggarwal, Pranjal Dutta, Saswata Mukherjee, Satyajeet Nagargoje, and Maciej Obremski. Efficient randomized strong 2-source non-malleable extractor for any linear min-entropy. In *Advances in Cryptology – CRYPTO 2025*, pages 679–713, 2025. 4, 5

[Alo99] Noga Alon. Combinatorial nullstellensatz. *Combinatorics, Probability and Computing*, 8(1–2):7–29, 1999. 5, 6

[ASE08] Noga Alon, Joel H. Spencer, and Paul Erdős. Correlation inequalities. In *The Probabilistic Method*, chapter 6, pages 85–96. John Wiley & Sons, Ltd, 2008. 26

[BD18] Peter Beelen and Mrinmoy Datta. Generalized hamming weights of affine cartesian codes. *Finite Fields and Their Applications*, 51:130–145, 2018. 10, 19

[BTU26] Alex Biryukov, Philip Tureček, and Aleksei Udovenko. Algorithmic toolkit for linearization of s-boxes. In Joan Daemen and Emmanuel Thomé, editors, *Advances in Cryptology – EUROCRYPT 2026*, pages 243–272, Cham, 2026. Springer Nature Switzerland. 1, 2, 8

[CD04] Claude Carlet and Cunsheng Ding. Highly nonlinear mappings. *Journal of Complexity*, 20(2):205–244, 2004. Festschrift for Harald Niederreiter, Special Issue on Coding and Cryptography. 1, 2, 8

[CD07] Claude Carlet and Cunsheng Ding. Nonlinearities of s-boxes. *Finite Fields and Their Applications*, 13(1):121–135, 2007. 1, 2, 8

[CF01] L Chen and F Fu. On the nonlinearity of multi-output boolean functions. *Acta Sci. Nat. Univ. Nankai*, 34(4):28–33, 2001. 2, 8

[CR11] Jonathan Cutler and A. J. Radcliffe. Extremal problems for independent set enumeration. *The Electronic Journal of Combinatorics*, 18(1):#P169, Aug. 2011. 13

[Fin14] Magnus Gausdal Find. On the complexity of computing two nonlinearity measures. In Edward A. Hirsch, Sergei O. Kuznetsov, Jean-Éric Pin, and Nikolay K. Vereshchagin, editors, *Computer Science - Theory and Applications*, pages 167–175, Cham, 2014. Springer International Publishing. 8

[GH00] O. Geil and T. Hoholdt. Footprints or generalized bezout's theorem. *IEEE Transactions on Information Theory*, 46(2):635–641, 2000. 4, 10

[GKPW26] Jun Gao, Peiru Kuang, Oleg Pikhurko, and Yan Wang. An improved bound on the minimum size of turán $(r+1, r)$-systems. *arXiv preprint arXiv:2608.23967*, 2026. 6, 7

[GM17] Olav Geil and Stefano Martin. Relative generalized hamming weights of q-ary reed-muller codes. *Advances in Mathematics of Communications*, 11(3):503–531, 2017. 10, 19

[GMT08] Olav Geil, Ryutaroh Matsumoto, and Casper Thomsen. On field size and success probability in network coding. In Joachim von zur Gathen, José Luis Imaña, and Çetin Kaya Koç, editors, *Arithmetic of Finite Fields: Second International Workshop, WAIFI 2008*, volume 5130 of *Lecture Notes in Computer Science*, pages 157–173. Springer, 2008. 4

[HP98] P. Heijnen and R. Pellikaan. Generalized hamming weights of q-ary reed-muller codes. *IEEE Transactions on Information Theory*, 44(1):181–196, 1998. 10, 19

[Kee11] Peter Keevash. *Hypergraph Turán problems*, page 83–140. London Mathematical Society Lecture Note Series. Cambridge University Press, 2011. 6

[Kop10] Swastik Kopparty. *Algebraic Methods in Randomness and Pseudorandomness*. Ph.d. thesis, Massachusetts Institute of Technology, 2010. 1, 3, 4

[Kop19] Swastik Kopparty. Fourier analysis, 2019. Lecture notes for Topics in Finite Fields, Rutgers University. 7

[LC14] J Liu and L Chen. On nonlinearity of the second type of multi-output boolean functions. *Chinese Journal of Engineering Mathematics*, 31(1):9–22, 2014. 2, 8

[LMC17] Jian Liu, Sihem Mesnager, and Lusheng Chen. On the nonlinearity of s-boxes and linear codes. *Cryptography and Communications*, 9(3):345–361, 2017. 1, 2, 8

[Mac27] F. S. MacAulay. Some properties of enumeration in the theory of modular systems. *Proceedings of the London Mathematical Society*, s2-26(1):531–555, 1927. 10

[Nag25] Gábor P. Nagy. On the minimum hamming distance between vectorial boolean and affine functions. *Cryptography and Communications*, 17(6):1703–1720, 2025. 1, 2, 8

[Nyb91] Kaisa Nyberg. Perfect nonlinear s-boxes. In Donald W. Davies, editor, *Advances in Cryptology — EUROCRYPT '91*, pages 378–386, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg. 1, 2

[Nyb93] Kaisa Nyberg. On the construction of highly nonlinear permutations. In Rainer A. Rueppel, editor, *Advances in Cryptology — EUROCRYPT' 92*, pages 92–98, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg. 1, 2

[SSSZ19] Ashwin Sah, Mehtaab Sawhney, David Stoner, and Yufei Zhao. The number of independent sets in an irregular graph. *Journal of Combinatorial Theory, Series B*, 138:172–195, 2019. 4, 13

[STV20] Srikanth Srinivasan, Utkarsh Tripathi, and S. Venkitesh. Decoding variants of reed–muller codes over finite grids. *ACM Transactions on Computation Theory*, 12(4):1–11, 2020. 4

[Tur41] Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48:436–452, 1941. In Hungarian, with German summary. 6

[VW08] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(7):137–168, 2008. 3, 4

A Proof of the Interpolation Bound

Proof of Lemma 3.5. To prove the lemma we will construct $\bar{Q} = (Q_1, ..., Q_m)$, a tuple of n-variate polynomials of degree at most d, that agrees on at least $M(q, n, d)$ points. We will define the formal matrix A,

$$A \in \mathbb{F}_q[X_{1,1}, \ldots, X_{1,n}, \ldots, X_{M(q,n,d),1} \ldots X_{M(q,n,d),n}]^{M(q,n,d)\times M(q,n,d)},$$

whose rows are indexed by n-variate monomials of degree at most d and individual degree at most $q-1$. For any n-variate monomial R of degree at most d and individual degree at most $q-1$, and $i \in [M(q,n,d)]$ we define,

$$A_{R,i} = R(X_{i,1}, ..., X_{i,n}).$$

A is the formal interpolation matrix in the following sense: The evaluation of

$$Q(\bar{X}) = \sum_{\bar{e}\in\{0,..,q-1\}^n \text{and } |\bar{e}|\leq d} q_{\bar{e}}\bar{X}^{\bar{e}},$$

at the points $\{(\alpha_{i,1}, \ldots \alpha_{i,n}) : i \in [M(q,n,d)]\}$ can be expressed as by the matrix vector product

$$\vec{Q} \cdot A(\alpha_{1,1}, \ldots, \alpha_{1,n}, \ldots, \alpha_{M(q,n,d),1}, \ldots, \alpha_{M(q,n,d),n}),$$

where $\vec{Q}$ is the coefficient vector of Q.

By the formula of the determinant,

$$\det(A) = \sum_{\sigma\in S(M(q,n,d))} (-1)^{sgn(\sigma)} \prod_{\substack{R \text{ is a monomial} \\ \text{of } \textit{degree} \text{ at most d and} \\ \text{individual } \textit{degree} \text{ at most } q-1}} R(X_{\sigma(R),1}, \ldots, X_{\sigma(R),n}),$$

where $S(M(q,n,d))$ is the symmetric group on $M(q,n,d)$ elements. Since the variables are disjoint and $R's$ are distinct, none of the terms of the determinant cancels out, which implies that $\det(A) \neq 0$.

Moreover, $\det(A)$ has individual degree at most $q-1$. Thus by Lemma 2.3, there exists $S \in \mathbb{F}_q^{n \cdot M(q,n,d)}$ of the form $S = (x_{1,1}, \ldots, x_{1,n}, \ldots, x_{M(q,n,d),1} \ldots x_{M(q,n,d),n})$ such that,

$$\det(A)(x_{1,1}, \ldots, x_{1,n}, \ldots, x_{M(q,n,d),1} \ldots x_{M(q,n,d),n}) \neq 0.$$

Thus, $\hat{S} = \{(x_{i,1}, \ldots, x_{i,n}) : i \in [M(q,n,d)]\}$ is a set of interpolation points for n-variate polynomials with total degree at most d and individual degree at most $q-1$. This implies that we can interpolate $\bar{Q}$ such that for any $\alpha \in \hat{S}$, $\bar{Q}(\alpha) = F(\alpha)$. Hence, we are done. □

B Proof of Existence of Hard Functions

Proof of Lemma 3.7. To prove this lemma we will use the probabilistic method. Let $\mathcal{Q}$ be the set of tuples of size m of n-variate polynomial functions of degree d over $\mathbb{F}_q$. Observe that $|\mathcal{Q}| = q^{M(q,n,d) \cdot m}$, since for each of the m coordinates we need to choose one of $q^{M(q,n,d)}$ many polynomials of degree at most d. Let $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ be a random function where the function evaluations are chosen uniformly at random. Note that $\mathsf{maxagree}(F, \deg d)$ would be at least t when there exists an $\bar{Q} \in \mathcal{Q}$ such that $\mathsf{agree}(F, \bar{Q}) \geq t$. So,

$$\begin{aligned}
\mathbb{P}\left[\mathsf{maxagree}(F, \deg d) \geq t\right] &= \mathbb{P}\left[\exists \bar{Q} \in \mathcal{Q} : \mathsf{agree}(F, \bar{Q}) \geq t\right] \\
&\leq \sum_{\bar{Q} \in \mathcal{Q}} \mathbb{P}\left[\mathsf{agree}(F, \bar{Q}) \geq t\right] && \text{(By Union Bound)} \\
&\leq q^{M(q,n,d) \cdot m} \cdot \binom{q^n}{t} \cdot q^{-mt} \\
&\leq e^{M(q,n,d) \cdot m \ln q} \cdot e^{t(n \ln q + 1) - t \ln t} \cdot e^{-mt \ln q} && \left(\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b\right) \\
&\leq e^{M(q,n,d) \cdot m \ln q} \cdot e^{-t(m-n) \ln q} \cdot e^{-t(\ln t - 1)}.
\end{aligned}$$

Thus if either $t \cdot (m-n) \cdot \ln q$ or $t(\ln t - 1)$ is greater than $M(q,n,d) \cdot m \cdot \ln q$, the above probability is less than 1 (this uses the fact that $m \geq n$). Thus by setting

$$t = \min\left(\frac{m}{m-n} \cdot M(q,n,d), \quad M(q,n,d) \cdot m \cdot \ln q \cdot \ln\left(M(q,n,d) \cdot m \cdot \ln q\right)\right),$$

we ensure that $\mathbb{P}[\mathsf{maxagree}(F, \deg d) \geq t] < 1$, and thus by the probabilistic method there exists an $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$, such that

$$\mathsf{maxagree}(F, \deg d) < t$$

which finishes the proof.

□

C Proof of the Fourier Bound

Proof of Lemma 3.9. We will apply the Fourier Transform to F_y with respect to the characters of the additive group $G = \mathbb{F}_q^n$. For $\alpha \in \mathbb{F}_q^n$, $\chi_\alpha : \mathbb{F}_q^n \to \mathbb{C}$ defined as $\chi_\alpha(x) = \psi(\langle \alpha, x \rangle)$ are all the characters of G. Applying the Fourier Transform on F_y gives us,

$$\begin{aligned}
\widehat{F_y}(\alpha) &= \mathbb{E}_{x \in \mathbb{F}_q^n}[F_y(x)\chi_\alpha(x)] \\
&= \mathbb{E}_{x \in F_q^n}[\psi(\langle y, F(x) \rangle)\overline{\psi(\langle \alpha, x \rangle)}]. \qquad \text{(C.1)}
\end{aligned}$$

We will upper bound the agreement between F and $L = Ax + b$ by relating it to the Fourier transform of F_y. For this, first observe the following orthogonality criterion for $z \in \mathbb{F}_q^m$,

$$\mathbf{1}[z = 0] = \mathbb{E}_{y \in \mathbb{F}_q^m}[\psi(\langle y, z \rangle)]. \qquad \text{(C.2)}$$

We use the above criterion to capture the information about the agreement between F and L in terms of the Fourier coefficients of F_y,

$$\begin{aligned}
\mathsf{agree}(F,L) &= \sum_{x\in\mathbb{F}_q^n} \mathbf{1}[F(x)-L(x)=0] \\
&= \sum_{x\in\mathbb{F}_q^n} \mathbb{E}_{y\in\mathbb{F}_q^m}\Big[\psi(\langle y, F(x)-L(x)\rangle)\Big] && \text{(From Equation C.2)} \\
&= q^n \mathbb{E}_{x\in\mathbb{F}_q^n}\mathbb{E}_{y\in\mathbb{F}_q^m}\Big[\psi(\langle y, F(x)-L(x)\rangle)\Big] \\
&= q^n \mathbb{E}_{x\in\mathbb{F}_q^n}\mathbb{E}_{y\in\mathbb{F}_q^m}\Big[\psi(\langle y, F(x)\rangle - \langle y, Ax+b\rangle)\Big] \\
&= q^n \mathbb{E}_{x\in\mathbb{F}_q^n}\mathbb{E}_{y\in\mathbb{F}_q^m}\Big[\psi(\langle y, F(x)\rangle)\overline{\psi(\langle y, Ax+b\rangle)}\Big] \\
&= q^n \mathbb{E}_{x\in\mathbb{F}_q^n}\mathbb{E}_{y\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\Big[\psi(\langle y, F(x)\rangle)\overline{\psi(\langle y, Ax\rangle)}\Big] \\
&= q^n \mathbb{E}_{y\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\mathbb{E}_{x\in\mathbb{F}_q^n}\Big[\psi(\langle y, F(x)\rangle)\overline{\psi(\langle y, Ax\rangle)}\Big] && \text{(Interchanging expectations)} \\
&= q^n \mathbb{E}_{y\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\mathbb{E}_{x\in\mathbb{F}_q^n}\Big[\psi(\langle y, F(x)\rangle)\overline{\psi(\langle A^Ty, x\rangle)}\Big] && \text{(Since } \langle y, Ax\rangle = \langle A^Ty, x\rangle \text{)} \\
&= q^n \mathbb{E}_{y\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\mathbb{E}_{x\in\mathbb{F}_q^n}\Big[F_y(x)\overline{\chi_{A^Ty}(x)}\Big] && \text{(By definition of } F_y \text{ and } \chi_{A^Ty} \text{)} \\
&= q^n \mathbb{E}_{y\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\widehat{F}_y(A^Ty) && \text{(From Equation C.1)} \\
&= \frac{q^n}{q^m} + \frac{q^n}{q^m}\sum_{y\neq 0\in\mathbb{F}_q^m}\overline{\psi(\langle y,b\rangle)}\widehat{F}_y(A^Ty) \\
&\leq \frac{q^n}{q^m} + q^n\cdot\frac{q^m-1}{q^m}\max_{(\alpha,y)\neq(0,0)}\left|\widehat{F}_y(\alpha)\right| \\
&\leq \frac{q^n}{q^m} + q^n\max_{(\alpha,y)\neq(0,0)}\left|\widehat{F}_y(\alpha)\right|.
\end{aligned}$$

□

D Limitations of the Hypergraph Framework

D.1 Over $\mathbb{F}_2$

Now we will give a lower bound on the number of independent sets for a r-uniform hypergraph with m edges. First, we will show the existence of a large independent set in a sparse hypergraph in the following lemma. The proof of the lemma is standard and can be found in [ASE08]. For the sake of completeness, we will provide the proof here.

Lemma D.1. *Let $H=(V,E)$ be a r-uniform hypergraph, where $V=[n]$ and $|E|=m$, then there exists an independent set $T\subset[n]$, such that*

$$|T| \geq \frac{n}{1+(r-1)\cdot r\cdot\left(\frac{m}{n}\right)}.$$

Proof. Let the degree of any vertex $v\in V$ be d_v, then we claim that it is enough to show that there exists an independent $T\subset[n]$ such that

$$|T| \geq \sum_{v\in V}\frac{1}{1+(r-1)d_v}$$

since by Jensen's inequality it will follow that

$$|T| \geq \frac{n}{1+(r-1)\sum_v \left(\frac{d_v}{n}\right)} = \frac{n}{1+(r-1)r\cdot\left(\frac{m}{n}\right)}$$

which would finish the proof.

Now we will show the existence of such a T using the probabilistic method. We first pick a uniform random ordering of all the vertices. Next, we construct the set $I \subset V$ as follows

$$I = \{v \in V \mid \text{ For any } u \in N(v), v < u\}$$

where $N(v)$ is the neighborhood of v in H. In particular, note that I is an independent set. Next, we will compute the expected size of $|I|$ by linearity of expectation. Define indicator random variables X_v for each $v \in V$ where

$$X_v = \begin{cases} 1 & v \in I \\ 0 & v \notin I \end{cases}$$

Then note that $|I| = \sum_v X_v$. Furthermore,

$$\mathbb{E}[X_v] = \frac{1}{1+|N(v)|} \geq \frac{1}{1+(r-1)d_v}$$

and thus by linearity of expectation,

$$\mathbb{E}[|I|] \geq \sum_{v\in V} \frac{1}{1+(r-1)d_v}$$

which implies there exists an independent set T such that

$$|T| \geq \sum_{v\in V} \frac{1}{1+(r-1)d_v}$$

which concludes the proof. □

As a corollary of the above lemma, we get the following lower bound for $|\mathsf{Ind}(H)|$.

Corollary D.2. *Let $H = (V, E)$ be a r-uniform hypergraph where $V = [n]$ and $|E| = m$, then*

$$|\mathsf{Ind}(H)| \geq 2^{n\cdot\left(\frac{1}{1+r^2\cdot(m/n)}\right)}$$

Proof. By Lemma D.1, there exists an independent set $T \subset [n]$, such that

$$|T| \geq \frac{n}{1+(r-1)\cdot r\cdot\left(\frac{m}{n}\right)} \geq n\cdot\left(\frac{1}{1+r^2\cdot(m/n)}\right)$$

. Since any subset of T will also be an independent set, thus

$$|\mathsf{Ind}(H)| \geq 2^{|T|} \geq 2^{n\cdot\left(\frac{1}{1+r^2\cdot(m/n)}\right)}$$

which finishes the proof. □

Remark D.3. *The above corollary means that for any d, we cannot create functions $F : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$ with $\mathsf{maxagree}(F, \deg d) < o\left(2^{n\cdot\left(\frac{1}{1+(d+1)^2\cdot(m/n)}\right)}\right)$ using our hypergraph framework.* ⌟

D.2 Over $\mathbb{F}_q$

Just like the $\mathbb{F}_2$ case, for an r-uniform hypergraph H such that $|M_{H,q}| \leq m$, we will show a lower bound on $|N_H(q)|$ in the following lemma.

Lemma D.4. *Let $H = (V, E)$ be an r-uniform hypergraph, such that $|M_{H,q}| = m$, then*

$$|N_H(q)| \geq 2^{n \cdot \left(\frac{1}{1+r^2 \cdot (m/n)}\right)}$$

Proof. Note that for any $W \in \mathsf{Ind}(H)$, $\mathbf{1}_W \in N_H(q)$, thus $|N_H(q)| \geq |\mathsf{Ind}(H)|$. Thus, to lower bound $|N_H(q)|$, it suffices to lower bound the number of independent sets of H. Since $|E(H)| \leq |N_H(q)|$, thus $|E(H)| \leq m$. By Lemma D.1, there exists an independent set $T \subset [n]$ such that

$$|T| \geq \frac{n}{1 + (r-1) \cdot r \cdot \left(\frac{m}{n}\right)} \geq n \cdot \left(\frac{1}{1 + r^2 \cdot (m/n)}\right).$$

Since any subset of T is also an independent set, thus

$$|N_H(q)| \geq 2^{|T|} \geq 2^{n \cdot \left(\frac{1}{1+r^2 \cdot (m/n)}\right)}$$

and we are done. □

Remark D.5. *The above corollary means that for any d, q, we cannot create functions $F : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^m$ with $\mathsf{maxagree}(F, \deg\, d) < o\left(2^{n \cdot \left(\frac{1}{1+(d+1)^2 \cdot (m/n)}\right)}\right)$ using our $(d+1)$-uniform hypergraph framework.* ⌟

E Proof of Lemma 5.16

Proof of Lemma 5.16. Let $\hat{H}$ be the r-uniform hypergraph over k vertices from Theorem 2.8. Furthermore add at most $\mathcal{O}(k^{r-1})$ edges to ensure $\partial_{r-1}\hat{H} = K_k^{(r-1)}$. Let $\lambda = \frac{|M_{\hat{H},q}|}{\binom{k+r-1}{r}}$. Let $H_{\text{Turán},k,r}$ be n/k disjoint copies of $\hat{H}$, then we get

$$|M_{H_{\text{Turán}}(k,r),q}| = \lambda \cdot \frac{n}{k} \cdot \binom{k+r-1}{r}.$$

Let $\hat{\lambda} = \frac{|\hat{H}|}{\binom{k}{r}}$ be the edge density, then by Theorem 2.8, we know that

$$\hat{\lambda} \leq \left(\frac{2}{1 + \lfloor \frac{r+1}{2} \rfloor} + o(1)\right)$$

Next we will show that $\lambda \leq \hat{\lambda} + o(1)$. Note that

$$\begin{aligned} |M_{\hat{H},q}| &\leq \binom{k+r-1}{r} - \binom{k}{r} + |\hat{H}| \\ &\leq \hat{\lambda} \cdot \binom{k}{r} \cdot (1 + o(1)) \end{aligned}$$

which implies

$$\begin{aligned} \lambda &\leq \hat{\lambda} \cdot \frac{\binom{k}{r}}{\binom{k+r-1}{r}} \cdot (1 + o(1)) \\ &\leq \hat{\lambda} \cdot (1 + o(1)). \end{aligned}$$

Now we will find $|N_{\hat{H}}(q)|$. For any $\alpha \in N_{\hat{H}}(q)$ such that $\sum_i \alpha_i \geq r$, since $\partial_{r-1} H = K_k^{(r-1)}$, thus it is necessary that $\alpha \in \{0,1\}^k$, which in turn implies that $\alpha = \mathbf{1}_W$ where $W \in \mathsf{Ind}(\hat{H})$. Using this observation, the following is true

$$|N_{\hat{H}}(q)| = \binom{k+r-1}{r-1} + \sum_{i \geq r} |\mathsf{Ind}(\hat{H}, i)|$$

where the first term is the number of number of non negative integer solutions to $X_1 + \cdots + X_k \leq r-1$. Since adding edges to an Turán(n, r, s) system won't increase the independence number which implies that for every $r+1$ size subset of $[k]$ is covered by an edge of $\hat{H}$. In particular, this gives the following lower bound on the number of edges of $\hat{H}$

$$|E(\hat{H})| \geq \frac{1}{r+1}\binom{k}{r},$$

which implies $|\mathsf{Ind}(\hat{H}, r)| \leq (1 - 1/(r+1)) \cdot \binom{k}{r}$, and thus $|N_{\hat{H}}(q)| \leq \binom{k}{r}$. Moreover, since $H_{\text{Turán},k,r}$ is n/k disjoint copies of $\hat{H}$, thus

$$|N_{H_{\text{Turán}}(k,r)}(q)| \leq \binom{k}{r}^{n/k}$$

This concludes the proof.

□