On the Fourier Spectrum of Symmetric Boolean Functions*

Mihail N. Kolountzakis† Richard J. Lipton‡ Evangelos Markakis§
Aranyak Mehta¶ Nisheeth K. Vishnoi ‖

Abstract

We study the following question:

> What is the smallest t such that every symmetric boolean function on k variables (which is not a constant or a parity function), has a non-zero Fourier coefficient of order at least 1 and at most t?

We exclude the constant functions for which there is no such t and the parity functions for which t has to be k. Let $\tau(k)$ be the smallest such t. Our main result is that for large k, $\tau(k) \leq 4k/\log k$.

The motivation for our work is to understand the complexity of learning symmetric juntas. A k-junta is a boolean function of n variables that depends only on an unknown subset of k variables. A symmetric k-junta is a junta that is symmetric in the variables it depends on. Our result implies an algorithm to learn the class of symmetric k-juntas, in the uniform PAC learning model, in time $n^{o(k)}$. This improves on a result of Mossel, O'Donnell and Servedio in [14], who show that symmetric k-juntas can be learned in time $n^{\frac{2k}{3}}$.

1 Introduction

Problem statement

The study of the Fourier representation of boolean functions has proved to be extremely useful in computational complexity and learning theory. In this paper we focus on the Fourier spectrum of symmetric boolean functions and we study the following question:

> What is the smallest t such that every symmetric boolean function on k variables (which is not a constant or a parity function), has a non-zero Fourier coefficient of order at least 1 and at most t?

*This work was done when all authors were at the Georgia Institute of Technology and it is based on the preliminary versions [12] and [9].

†Department of Mathematics, Univ. of Crete, GR-71409 Iraklio, Greece. E-mail: kolount@gmail.com. Partially supported by European Commission IHP Network HARP (Harmonic Analysis and Related Problems), Contract Number: HPRN-CT-2001-00273 - HARP, and by grant INTAS 03-51-5070 (2004) (Analytical and Combinatorial Methods in Number Theory and Geometry).

‡Georgia Tech, College of Computing, Atlanta, GA 30332, USA, and Telcordia Research, Morristown, NJ 07960, USA, E-mail: rjl@cc.gatech.edu. Research supported by NSF grant CCF-0431023.

§*Corresponding author:* University of Toronto, Department of Computer Science, Toronto, ON M5S3G4, Canada, E-mail: vangelis@cs.toronto.edu

¶IBM Almaden Research Center, 650 Harry Rd, San Jose, CA 95120, USA, E-mail: mehtaa@us.ibm.com

‖College of Computing, Georgia Institute of Technology, Atlanta GA 30332, USA, and IBM India Research Lab, Block-1, IIT Delhi, New Delhi, 110016, India, E-mail: nkv@cc.gatech.edu

We exclude the two constant functions, for which there is no such t, and the two parity functions, for which t has to be k. Let $\tau(k)$ be the smallest such t. While the above question is interesting in its own right, there is also an important learning theory application behind it, which we outline next.

Motivation

The motivation to study $\tau(k)$ comes from the following fundamental problem in computational learning theory: learning in the presence of irrelevant information. One formalization of the problem is as follows: we want to learn an unknown boolean function of n variables, which depends only on $k \ll n$ variables. Typically, k is $O(\log n)$. Such a function is referred to as a k-junta. The input is a set of labeled examples $\langle \mathbf{x}, f(\mathbf{x}) \rangle$, where the $\mathbf{x}$'s are picked uniformly and independently at random from the domain $\{0,1\}^n$. The goal is to identify the k relevant variables and the truth table of the function.

The problem was first posed by Blum [1] and Blum and Langley [4], and it is considered [2, 14] to be one of the most important open problems in the theory of uniform distribution learning. It has connections with learning DNF formulas and decision trees of super-constant size, see [5, 8, 13, 17, 18] for more details. The general case is believed to be hard and has even been used in the construction of a cryptosystem [3]. A trivial algorithm runs in time roughly n^k by doing an exhaustive search over all possible sets of relevant variables. Two important classes of juntas are learnable in polynomial time: parity and monotone functions. Learning parity functions can be reduced to solving a system of linear equations over $\mathbb{F}_2$ [7]. Monotone functions have non-zero singleton Fourier coefficients (see [14]). For the general case, the first significant breakthrough was given in [14] - learning with confidence $1-\delta$ in time $n^{0.7k}\mathrm{poly}(2^{\mathrm{k}}, \mathrm{n}, \log 1/\delta)$. Note that we allow the running time to be polynomial in 2^k, since this is the size of the truth-table which is output. In the typical setting of $k = O(\log n)$, this becomes polynomial in n.

Fourier based techniques in learning were introduced in [11] and have proved to be very successful in several problems. Fourier coefficients are easy to compute in the uniform distribution learning model and furthermore, if a Fourier coefficient is non-zero then its entire support is contained in the set of relevant variables. Hence, it is interesting to ask: what are the sub-classes of juntas for which Fourier based techniques yield fast learning algorithms? An important and natural subclass is the class of symmetric juntas. While this subclass contains only 2^{k+1} functions, the problem is not known to be significantly easier than the general case. The bound before our work was $n^{2k/3}$ [14], which is not much better than the best bound for general juntas (also obtained in [14]). Our results imply an improved bound for learning symmetric juntas via the Fourier based algorithm.

We believe that the case of symmetric juntas constitutes a good "challenge problem" towards the goal of learning general juntas. One motivation for this is a consideration of the following well-known challenge problem [2] :
Let $f(x_1, \ldots, x_n) := \mathrm{MAJORITY}(x_1, \ldots, x_{2k/3}) \bigoplus \left(x_{2k/3+1} \oplus \cdots \oplus x_k\right)$, where $x_1, \ldots, x_k$ are some *unknown* variables among $x_1, \ldots, x_n$. This subclass has been identified as a candidate hard to learn class [2]. The current bound for learning this subclass of juntas is $n^{k/3}$, and it is asked in [2] if a faster algorithm exists. Note that f is invariant under permutations of $\{x_1, \ldots, x_{2k/3}\}$ and under permutations of $\{x_{2k/3}, \ldots, x_k\}$, i.e., it is invariant under a large group of symmetries. This suggests that it is interesting to begin with the case of symmetric juntas.

2 Our Results

There are two main results in this paper:

2.1 The Self-Similarity Theorem

Theorem 2.1. *Let $1 \leq s \leq l$ be fixed integers such that $\tau(l) \leq s$. Then there exists $k_0 := k_0(s,l)$, such that for every $k \geq k_0$, $\tau(k) \leq \left(\frac{s+1}{l+1}\right) k + o(k)$.*

It was observed in [12], via a computer search, that $\tau(30) = 2$. This implies that $\tau(k) \leq 3k/31$.

Proof Technique. Not surprisingly, the study of $\tau(k)$ is equivalent to the study of 0/1 solutions of a system of Diophantine equations involving binomial coefficients. As a first step, we simplify these Diophantine equations by moving to a representation which is equivalent to the Fourier representation, but seems much simpler for the application of number theoretic tools. Once this is done, we reduce these Diophantine equations modulo carefully chosen prime numbers to get a simpler system of equations which we can analyze. Finally, we combine the information about the equations over the finite fields in a combinatorial manner to deduce the nature of the 0/1 solutions. The following well-known self-similarity property of Pascal's Triangle (known as Lucas' Theorem) plays an important role: If $m = lp$ for some integer l, and some prime p, then the values obtained by reducing the m-th row of Pascal's Triangle modulo p, can be read off directly from the l-th row of Pascal's Triangle.

2.2 The $O(k/\log k)$ Theorem

Theorem 2.2. *There is an absolute constant $k_0 > 0$ such that for $k \geq k_0$, $\tau(k) \leq 4k/\log k$.*

Proof Technique.

- We start again by looking at the 0/1 solutions of the system of Diophantine equations, as in the proof of Theorem 2.1. We then take a departure from this approach by further reducing this to the problem of showing that a certain integer-valued polynomial P is constant over the set $\{0, 1, ..., k\}$. We manage to prove this in two steps:

- First, we show that P is constant over the union of two small intervals $\{0, ..., t\} \cup \{k-t, ..., k\}$. This is obtained by looking at P modulo carefully chosen prime numbers. One way to prove this (at least infinitely often) would be to assume the twin primes conjecture (that there are an infinite number of pairs of primes whose difference is 2). We manage to replace the use of the twin prime conjecture (and get a result which works for all large enough k) by choosing four different primes in a more involved manner. To choose these prime numbers we use the Siegel-Walfisz theorem on the density of primes in arithmetic progressions with modulus of moderate growth. This is a generalization of Dirichlet's Theorem, and is stated precisely in Section 6.

- In the second step, we extend the constant nature of P to the whole interval $\{0, ..., k\}$ by repeated applications of Lucas' Theorem. One additional interesting aspect of our proof is the use of an equivalence between (a) the vanishing of Fourier coefficients, and (b) the equality of moments of certain random variables under the uniform measure on the hypercube and under the measure defined by the function itself. This equivalence helps in the proof by eliminating the need for a large amount of case analysis.

Our results imply a bound of $n^{o(k)}$ for the Fourier based learning algorithm for the class of symmetric k-juntas. To our knowledge, this is the best known upper bound for learning symmetric juntas under the uniform distribution. Independent of the learning problem, the fact that symmetric

boolean functions have non-zero Fourier coefficients of relatively small order provides new insight into the structure of these functions.

2.3 Related Work

Previously, the idea of reducing binomial coefficients modulo a prime number has been used in [19] to prove lower bounds on the degree of polynomials representing symmetric boolean functions. In [19], their problem reduces to showing that a certain sum of binomial coefficients is non-zero, which is done by reducing the sum modulo a prime number. Our problem involves a collection of sums which we have to prove are unequal. For this we need to consider reductions modulo many different primes which have to be carefully chosen so as to satisfy certain properties. Combining the information obtained by these reductions is also more involved in our case.

The result of [19] has in fact been used in the proof of the previous best $n^{2k/3}$ bound for learning symmetric juntas [14]. Using [19], it is shown in [14] that if a symmetric function f is *balanced*, i.e., $Pr[f(x) = 1] = 1/2$, then it has a non-zero Fourier coefficient of order $o(k)$. The $2k/3$ bottleneck comes in the case of *unbalanced* symmetric functions, which are analyzed through a different argument. As noted in [14] and as we also note in Section 6, the result of [19] does not seem to be applicable to learning unbalanced functions.

3 Notation

We consider boolean functions from $\{0,1\}^k \to \{0,1\}$. For a set $S \subseteq [k]$, define $\chi_S : \{0,1\}^k \to \{1,-1\}$ to be the function $\chi_S(\mathbf{x}) := (-1)^{\sum_{i \in S} x_i}$. By convention, the boldface $\mathbf{x}$ denotes a vector, in this case $(x_1, \ldots, x_k)$. For a function $f : \{0,1\}^k \to \{0,1\}$, and $S \subseteq [k]$, define the *Fourier coefficient* corresponding to S as $\hat{f}(S) := \frac{1}{2^k} \sum_{\mathbf{x} \in \{0,1\}^k} f(\mathbf{x}) \chi_S(\mathbf{x})$. The *order* of a Fourier coefficient $\hat{f}(S)$ is $|S|$. The Fourier expansion of f is: $f(\mathbf{x}) = \sum_{S \subseteq [k]} \hat{f}(S) \chi_S(\mathbf{x})$.

If f is symmetric, f is completely determined by its value on any $k+1$ vectors of distinct *weights* where the weight of a boolean vector is the number of 1's in it. We will use the following vector representation of f: $\nu(f) := (f_0, f_1, \ldots, f_k)$. Here f_i is the value of f on a vector of weight i. Further f has precisely $k+1$ (non-equivalent) Fourier coefficients, $(\hat{f}_0, \ldots, \hat{f}_k)$. Here $\hat{f}_t$ is defined as $\hat{f}(S)$, for some $S \subseteq [k]$ with cardinality t. Since f is symmetric, this does not depend on the choice of S. The following four special symmetric functions on k variables will appear often: the two constant functions $\mathbf{0}$ and $\mathbf{1}$, the parity function $\oplus$, and its complement $\overline{\oplus}$.

4 An Equivalent Formulation as a Diophantine Problem

In this section we give an equivalent condition for the existence of a non-zero Fourier coefficient of a boolean function f. While we prove the equivalence for all boolean functions, we use it only for the special case of symmetric functions.

Let $f : \{0,1\}^k \mapsto \{0,1\}$ be a boolean function. For a vector $\mathbf{x} = (x_1, \ldots, x_k)$, and a set $S \subseteq [k]$, $\mathbf{x}_S$ is the projection of $\mathbf{x}$ on the indices of S. Let $\sigma \in \{0,1\}^{|S|}$. Define the following probabilities:

$$p_{S,\sigma}(f) := \Pr\left[f(\mathbf{x}) = 1 | \mathbf{x}_S = \sigma\right].$$

Unless mentioned, all probabilities are over the uniform distribution. For $t \geq 1$, call a boolean function f on k variables *t-null*, if for all sets $S \subseteq [k]$, with $|S| = t$, and for all $\sigma \in \{0,1\}^t$, the probabilities $p_{S,\sigma}(f)$ are all equal to each other. The following lemma reveals the connection with the Fourier coefficients of f.

Lemma 4.1. *Let f be a boolean function on k variables. f is t-null for some $1 \leq t \leq k$, if and only if, for all $\emptyset \neq S \subseteq [k]$ with cardinality at most t, $\hat{f}(S) = 0$.*

Proof. It can be easily verified that if f is t-null, then for all $\emptyset \neq S \subseteq [k]$ with cardinality at most t, $\hat{f}(S) = 0$. This follows from the fact that the Fourier coefficients of order at most t can be expressed as ± 1 combinations of $p_{S,\sigma}(f)$ with $\sigma \in \{0,1\}^t$, and $S \subseteq [k], |S| = t$. When f is t-null, the terms cancel out. The proof of the other direction is by induction and we omit it here. □

The following is an immediate corollary of this lemma.

Corollary 4.2. *Let f be a boolean function on k variables. If f is t-null for some $1 \leq t \leq n$ then f is s-null for $1 \leq s \leq t$.*

When we consider the case of symmetric functions, $p_{S,\sigma}(f)$ just depends on $s := |S|$ and the weight w of σ. We denote this by $p_{s,w}(f)$. It is clear that

$$p_{s,w}(f) = \frac{1}{2^{k-s}} \sum_{i=0}^{k} f_i \binom{k-s}{i-w},$$

where $\binom{l}{m}$ is 0 if $m < 0$ or $m > l$, and $\binom{0}{0}$ is 1. By definition, f is s-null if for $0 \leq w \leq s$, $p_{s,w}(f)$ are all equal. Hence, f is s-null iff there exists $c := c(f,s,k)$ such that

$$\sum_{i=0}^{k} \binom{k-s}{i-w} f_i = c, \quad \forall\ 0 \leq w \leq s. \tag{1}$$

Thus, we have

Lemma 4.3. *For $1 \leq s \leq k$, let $A_{k,s}$ be the $(s+1) \times (k+1)$ matrix:*

$$A_{k,s}(i,j) := \binom{k-s}{j-i}.$$

A symmetric function f is s-null if and only if there exists a positive integer $c := c(f,s,k)$ such that:

$$A_{k,s} \cdot \nu(f) = c\mathbf{1}.$$

It is easy to see that the constant boolean functions $\{\mathbf{0}, \mathbf{1}\}$ satisfy this system of equations for all s, i.e., they are s-null for all s, s.t. $1 \leq s \leq k$. One can also see that the boolean functions $\{\oplus, \overline{\oplus}\}$ are s-null for all s s.t. $1 \leq s < k$. From Lemma 4.1 and Lemma 4.3 we get:

Corollary 4.4. *All symmetric boolean functions $f \notin \{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$ have a non-zero Fourier coefficient of order at most s_0 (and at least 1) iff there exists s, $1 \leq s \leq s_0$ s.t. $\{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$ are the only 0/1 solutions to:*

$$\sum_{i=0}^{k-s} f_i \binom{k-s}{i} = \sum_{i=1}^{k-s+1} f_i \binom{k-s}{i-1} = \cdots = \sum_{i=s}^{k} f_i \binom{k-s}{i-s}. \tag{2}$$

5 The Self-Similarity Theorem

In this section we prove Theorem 2.1. First we recall a few results from number theory that we will use repeatedly. The following result is a special case of Lucas' Theorem [6, Ch. 3] and illustrates the *self-similar* nature of the Pascal's Triangle modulo primes.

Lemma 5.1. *For a prime p, an integer $m \geq 0$ and $0 \leq i \leq mp$, $\binom{mp}{i} \equiv \binom{m}{j} \mod p$ if $i = jp$ for some $0 \leq j \leq m$, and 0 otherwise.*

On numerous occasions, we will use the following result about the density of primes. This follows from the Prime Number Theorem.

Lemma 5.2. *For large enough n, there is a prime $p \leq n$, such that $p = n - o(n)$.*

5.1 A Simple Bound of $k/2$

In this section we give a self-contained proof of the following (weaker) result. The aim is to illustrate the key ideas behind the proof of Theorem 2.1.

Theorem 5.3. *For any symmetric boolean function f on k variables ($f \notin \{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$), there exists $1 \leq t \leq \frac{k}{2} + o(k)$ such that $\hat{f}_t \neq 0$.*

We need the following combinatorial lemma. For positive integers k, p, q, s.t. $p \neq q$, let $G_{k,p,q}$ be the graph with vertex set $\{0, 1, 2, \ldots, k\}$, and the edge set $\{(i, j) : |i - j| = p \text{ or } q\}$.

Lemma 5.4. *For positive integers k, p, q such that $(p, q) = 1$ and $p + q \leq k$, $G_{k,p,q}$ is connected.*

Proof. We proceed by induction on $p + q$. Without loss of generality, let $p > q$. Clearly, the lemma holds for the base case. Let i, j be s.t. $0 \leq i < j \leq k$ and $j - i = p - q$. Since $p + q \leq k$, either $i + p \leq k$ or $i - q \geq 0$. In either case, there is a path of length 2 between i and j. Hence, replacing the edges $\{(u, v) : |u - v| = p\}$ by the new edges $\{(u', v') : |u' - v'| = p - q\}$ does not increase the connectivity of the graph. It suffices to show that $G_{k,p-q,q}$ is connected, which follows by the induction hypothesis. □

Proof of Theorem 5.3 : Let f be a symmetric function such that for every $1 \leq t \leq \frac{k}{2} + o(k)$, $\hat{f}_t = 0$. We will show that $f \in \{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$.

By Lemma 5.2, we can pick primes p, q, s.t. $\frac{k}{2} - o(k) = p < q \leq \frac{k}{2}$. Since $k - p$ and $k - q$ are both at most $\frac{k}{2} + o(k)$, we get from Lemma 4.1 that f is $(k - p)$-null and $(k - q)$-null. Hence, by Lemma 4.3, there are constants c_1, c_2 such that

$$A_{k,k-p}\nu(f) = c_1\mathbf{1} \quad \text{and} \quad A_{k,k-q}\nu(f) = c_2\mathbf{1}.$$

Consider these two systems of equations modulo p and q respectively. Let $0 \leq c_p < p$ and $0 \leq c_q < q$ be s.t. $c_p \equiv c_1 \mod p$, and $c_q \equiv c_2 \mod q$. We will use $\equiv_p$ to denote congruences $\mod p$ (and similarly for q). The systems become:

$$A_{k,k-p}\nu(f) \equiv_p c_p\mathbf{1} \quad \text{and} \quad A_{k,k-q}\nu(f) \equiv_q c_q\mathbf{1}.$$

Now, from Lemma 5.1, we see that $\binom{p}{i} \equiv_p 1$ if $i = 0$ or $i = p$, and $\binom{p}{i} \equiv_p 0$ otherwise (and similarly for q). Hence, we see that the equations are of the form

$$f_i + f_{i+p} \equiv_p c_p \quad \text{for } 0 \leq i \leq k - p$$

and

$$f_i + f_{i+q} \equiv_q c_q \text{ for } 0 \leq i \leq k-q.$$

Since $f_i \in \{0,1\}$ and $p > 2$, these modular equations are in fact exact equalities and $c_p, c_q \in \{0,1,2\}$. If $c_p = 0$, then it follows that $c_q = 0$ and $f = \mathbf{0}$. If $c_p = 2$, then $c_q = 2$ and $f = \mathbf{1}$. The only remaining case is $c_p = c_q = 1$. This gives

$$f_i = 1 - f_{i+p} \text{ for } 0 \leq i \leq k-p \quad \text{and} \quad f_i = 1 - f_{i+q} \text{ for } 0 \leq i \leq k-q.$$

In other words, $|i-j| = p$ or q implies that $f_i = 1 - f_j$. Since $G_{k,p,q}$ is connected (Lemma 5.4) it follows that fixing the value of any one f_i uniquely determines f, and hence, there are at most 2 possible choices for f. We can see that $\{\oplus, \overline{\oplus}\}$ are solutions to these equations, and hence, they are the only solutions in this case.

□

5.2 Proof of Theorem 2.1

Recall that the hypothesis of the Theorem is that $\tau(l) \leq s$. Let f be a symmetric boolean function on k variables. Suppose that f is t-null, for all $t \leq \left(\frac{s+1}{l+1}\right) k + o(k)$. We will show that $f \in \{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$.

Let $m = l - s$. As of now, assume that there is a prime p such that $k = (m+s+1)p - 1$. We handle the case when there is no such prime p later. Set $t := k - mp = (s+1)p - 1$. Since $p = \frac{k+1}{l+1}$,

$$t = \left(\frac{s+1}{l+1}\right) k + \frac{s+1}{l+1} - 1 < \left(\frac{s+1}{l+1}\right) k.$$

Hence, f being t-null implies that there is an integer c such that

$$A_{k,t}\nu(f) = c\mathbf{1}. \tag{3}$$

We remark that the role of $o(k)$ term is redundant in this case. It will play a role when we cannot choose p such that $k - t = mp$.

Reducing to a smaller problem

Note that, by definition of t, $k - t = mp$. For $0 \leq i \leq p-1$, let $\mathbf{F}_i := (f_i, f_{i+p}, f_{i+2p}, \ldots, f_{i+lp})$. Hence, reducing Equations (3) modulo p, and using Lemma 5.1, one obtains the following systems of equations.

$$\begin{aligned} A_{l,s}\mathbf{F}_0 &\equiv c'\mathbf{1} \mod p \\ A_{l,s}\mathbf{F}_1 &\equiv c'\mathbf{1} \mod p \\ &\vdots \\ A_{l,s}\mathbf{F}_{p-1} &\equiv c'\mathbf{1} \mod p. \end{aligned}$$

Here $c' \equiv c \mod p$. If k is greater than $(l+1)2^{l-s}$, then it follows that $p > 2^{l-s}$. Therefore, for such a k, these modular equations are in fact exact. That is, there is a positive integer $d \geq 0$, such that the following set of equations hold.

$$\begin{aligned} A_{l,s}\mathbf{F}_0 &= d\mathbf{1} \\ A_{l,s}\mathbf{F}_1 &= d\mathbf{1} \\ &\vdots \\ A_{l,s}\mathbf{F}_{p-1} &= d\mathbf{1}. \end{aligned} \tag{4}$$

Using the fact that $\tau(l) \leq s$, we deduce that for any i, the system of equations $A_{l,s}\mathbf{F}_i = d\mathbf{1}$ has at most 4 solutions. Hence, fixing any two variables in $\mathbf{F}_i$ fixes all its variables. This implies that there are at most 4^p choices for f. Now we show how to narrow down these choices to 4.

Combining the smaller instances

Let $\frac{k}{2} < mp \leq q \leq (m+s)p$ be a prime. Since f is t-null, and $t = k - mp \geq k - q$, by Corollary 4.2, f is $(k-q)$-null. Now, consider the system of equations $A_{k,k-q}\nu(f) = c\mathbf{1}$ modulo the prime q. Since $q > 2$, we get, for some $e \geq 0$, exact equations of the following form:

$$\begin{aligned} f_0 + f_q &= e \\ f_1 + f_{q+1} &= e \\ &\vdots \\ f_{k-q} + f_k &= e. \end{aligned} \tag{5}$$

The idea is that these equations, along with Equations (4), are sufficient to restrict f to one of the four functions, as desired. First, we need a simple fact. For an integer $r \geq 0$, let $(r)_p := r \mod p$. Also, for $0 \leq i \leq p-1$, let $[iq]_p := \{(iq)_p, (iq)_p + p, \ldots, (iq)_p + (m+s)p\}$.

Fact 5.5. *Let p, q be distinct primes. Then, for $0 \leq i < j \leq p-1$, $[iq]_p \cap [jq]_p = \emptyset$, and $[i+q]_p \cap [j+q]_p = \emptyset$.*

Now, fix $f_0, f_p \in \mathbf{F}_0$. As noticed before, this fixes all the variables in $\mathbf{F}_0$. Using Equations (5), in particular, we get that f_q and f_{q+p} are fixed. Notice that $f_q, f_{q+p} \in \mathbf{F}_{(q)_p}$. Now Equations (4) imply that all the indices in $\mathbf{F}_{(q)_p}$ get fixed. Note that for any $0 \leq i' < p$, we have that $i' + q \leq k$ by the choice of q. Now applying this argument to $f_{(q)_p}$ and $f_{(q)_p+p}$ (which are in $\mathbf{F}_{(q)_p}$), we get that $f_{(q)_p+q}$ and $f_{(q)_p+p+q}$ are fixed. Note that these variables are in $\mathbf{F}_{(q+1)_p}$. By Fact 5.5, $\mathbf{F}_{(q+1)_p}$ is disjoint from $\mathbf{F}_{(q)_p}$.

Iterating the alternate use of these two systems of equations, along with Fact 5.5, one obtains that all the variables in $\mathbf{F}_i$, for every i, are fixed, once f_0 and f_p are fixed. Hence, f has at most four choices: $\{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$, one for every possible fixing of $\{f_0, f_p\}$. Thus, since $p > 2^{l-s}$ and $k = (l+1)p - 1$, we can choose $k_0 := k_0(l)$ such that for all $k \geq k_0$, $\tau(k) \leq t = \left(\frac{s+1}{l+1}\right)k + \frac{s+1}{l+1} - 1 \leq \left(\frac{s+1}{l+1}\right)k$.

Handling the residual class of variables

Now we consider the case when there is no prime p such that $k = (m+s+1)p - 1$. In this case, we pick a prime p in the interval $\left[\frac{k}{m+s+1} - o(k), \frac{k}{m+s+1}\right]$. We are guaranteed the existence of such a prime by Lemma 5.2. Let $t = k - mp$. Hence, $(s+1)p + o(p) \geq t \geq (s+1)p$. Since we think of m as a constant, $p = \Omega(k)$. Hence, there is a small number ($o(k)$) of variables, say $\mathbf{R}$, which remain to be dealt with in the previous argument. In particular, these are the variables starting from position $(m+s+1)p$ all the way to k and $\{f_0, ..., f_k\} = \left(\cup_{i=0}^{p-1}\mathbf{F}_i\right) \cup \mathbf{R}$. By the argument

in the previous case, fixing f_0 and f_p fixes all the variables in $\cup_{i=0}^{p-1}\mathbf{F}_i$. Further, since $|\mathbf{R}| = o(k)$, and $q > k/2$, every variable in $\mathbf{R}$ will appear in one of the Equations (5) along with a variable in $\cup_{i=0}^{p-1}\mathbf{F}_i$, and hence, get fixed.

Thus, since $p > 2^{l-s}$ and $k = (l+1)p - 1$, we can choose $k_0 := k_0(l,s)$ such that for all $k \geq k_0$, $\tau(k) \leq \left(\frac{s+1}{l+1}\right) k + o(k)$. This completes the proof of Theorem 2.1.

6 A bound of $O(k/\log k)$

This section is devoted to the proof of Theorem 2.2. We start with some general discussion about the proof. The preliminary setup is the following. Suppose f is a boolean function on $G = \mathbb{Z}_2^k$, such that all its non-constant Fourier coefficients of order up to $\epsilon k = k - N$ are 0. Then the values f_j of f satisfy (2) with $s = k - N$, which, changing indices, can be rewritten as:

$$\sum_j \binom{N}{j} f_{\nu+j} = c_N, \quad \text{for all } \nu = 0, \ldots, k-N. \tag{6}$$

It is easy to show by induction on N, starting with $N = k$ and going down, that

$$c_N = 2^N \mathrm{Avg}\, f = 2^{N-k} \sum_{x \in \{0,1\}^k} f(x). \tag{7}$$

We want to show that if $k - N = \epsilon k = 4k/\log k$, then f_j is either constant or alternates between 0 and 1. We prove this for all k sufficiently large.

Define $D_j = f_{j+1} - f_j$, for $j = 0, \ldots, k-1$, and observe that the sequence D_j satisfies the homogeneous version of (6):

$$\sum_j \binom{N}{j} D_{\nu+j} = 0, \quad \text{for all } \nu = 0, \ldots, k-N-1. \tag{8}$$

Recall that in (8) the number N can be replaced by any other integer N_1 in the interval $[N, k]$ by Corollary 4.2 and Lemma 4.3.

From (8) the sequence D_j may be defined for all $j \in \mathbb{Z}$ and $D_j \in \mathbb{Z}$ for all j. From the theory of recurrence relations we know then that the sequence D_j may be written as a linear combination of the following sequences:

$$(-1)^j, (-1)^j j, (-1)^j j^2, \ldots, (-1)^j j^{N-1}.$$

The reason for this is that -1 is the only root of the characteristic polynomial of the recurrence, $\phi(z) = \sum_j \binom{N}{j} z^j = (1+z)^N$. Therefore there is a polynomial $P(x)$, of degree at most $N-1$, such that

$$D_j = (-1)^j P(j), \quad \text{for all } j \in \mathbb{Z}.$$

Clearly $P(x)$ takes integer values on integers and in particular $P(j) \in \{-1, 0, 1\}$ for $j = 0, \ldots, k-1$. From the well known characterization of integer-valued polynomials [15, p. 129, Problem 85] it follows that we may write

$$P(x) = \sum_{j=0}^{N-1} a_j \binom{x}{j}, \quad \text{with } a_j \in \mathbb{Z}. \tag{9}$$

At this point it is instructive to give a proof, in this framework, of a result of [14]. This proof will also serve to clarify the relation of our method to that of [19]. A boolean function is called *balanced* if it takes the value 1 as often as it takes the value 0.

Theorem 6.1. (Mossel, O'Donnel and Servedio, 2003) *If $f : \{0,1\}^k \to \{0,1\}$ is a balanced symmetric function which is not constant or a parity function then some of its Fourier coefficients of order at most $O(k^{0.548})$ are non-zero.*

Proof. Subtracting c_N from both sides of (6) and using (7) we obtain that the sequence $f_n - \frac{c_N}{2^N} = f_n - \operatorname{Avg} f = f_n - \frac{1}{2}$ satisfies the homogeneous recurrence relation (8) in place of D_n. By the same reasoning as above $(-1)^n(f_n - \frac{1}{2})$ is then a polynomial of degree at most $N-1$. But it only takes the values $\pm\frac{1}{2}$ for $n = 0, 1, \ldots, N, \ldots, k-1$. Von zur Gathen and Roche [19] have shown that any polynomial $Q(n)$ which takes only two values for $n = 0, 1, \ldots, k$ must have degree $d \geq k - O(k^{0.548})$, hence $k - N = O(k^{0.548})$, which is what we wanted to prove. □

Remark. The method of [19] says nothing about polynomials which may take 3 or 4 values. If one ommits the assumption that f is balanced then the sequence $(-1)^n(f_n - \operatorname{Avg} f)$ may take up to 4 possible values.

Plan of proof. We assume that f has all non-constant Fourier coefficients of order up to $k - N$ equal to 0 and we want to show that $f \in \{\mathbf{0}, \mathbf{1}, \oplus, \overline{\oplus}\}$. Since $D_j = f_{j+1} - f_j$ it is enough to show that either D_j is identically 0 or that $D_j = (-1)^j$ or $D_j = (-1)^{j+1}$. This is equivalent to showing that $P(j) = (-1)^j D_j$ is a constant polynomial, constantly equal to $-1, 0$ or 1.

We will first show that the polynomial P is constant in two "small" intervals at the endpoints of the interval $[0, k]$ (Lemma 6.3). To achieve this we will first show that P has period 2 in each of these intervals (Lemma 6.2). For this we use some elaborate number-theoretic results (Theorem A) on the distribution of primes. Many of the technicalities in that part would not be needed if one knew that there are plenty of twin primes, that is integers p such that p and $p+2$ are both primes.

Once we have that P is constant in these two intervals near the endpoints of $[0, k]$ we show using the modular approach that P is also constant on a similar interval around the midpoint of $[0, k]$ (Lemma 6.4). At this point a significant element of our method is to eliminate the possibility that P is 0 (we are assuming of course that f is not constant). To show this we interpret f as a probability measure on the discrete cube and the vanishing of Fourier coefficients up to order r becomes equivalent with r-wise independence of the marginals of that measure (Theorem 6.5). It follows that if P vanishes in the middle interval in question then the second moment of a certain random variable would be larger than we know it is (Corollary 6.6). This elimination of 0 as a possible value is what makes the method work. We repeatedly obtain that P is constant in more and more intervals of the same length, each in the middle of the existing gaps, until the whole interval $[0, k]$ is covered (Lemma 6.8).

Notation. In what follows we repeatedly use the letter C to denote a positive constant which depends on no parameter (unless we say otherwise). As is customary, this constant C need not be the same in all its occurences.

Lemma 6.2. *The polynomial P satisfies the 2-periodicity condition*

$$P(j) = P(j+2),$$

whenever $j, j+2 \in \mathcal{A} = [0, k-N-\Gamma] \cup [N+\Gamma, k-1]$.

Proof. If $p \geq N$ is a prime, and since all the factors that appear in denominators in (9) are strictly less than p (hence invertible mod p), it follows that the sequence $P(j) \bmod p$, $j \in \mathbb{Z}$, may be viewed as a polynomial with coefficients in $\mathbb{Z}_p$ and therefore is a p-periodic sequence mod p, i.e.

$$P(j+p) = P(j) \bmod p, \quad \text{for all } j \in \mathbb{Z} \text{ and } p \geq N. \tag{10}$$

If, in addition, $0 \le j < j+p < k$, when all P-values that appear in (10) are in $\{-1, 0, 1\}$, it follows that we have the non-modular equality

$$P(j+p) = P(j), \quad (N \le p \le p+j < k). \tag{11}$$

We shall need various primes in intervals from now on. The version of the prime number theorem that we will be using is the Siegel-Walfisz theorem (see [10, Theorem 2]). Define the logarithmic integral

$$\operatorname{Li} x = \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}, \quad (x \to \infty).$$

The Euler function $\varphi(q)$ below denotes the number of moduli mod q which are coprime to q.

Theorem A (Siegel-Walfisz) Let $\pi(x; M, a)$ be the number of primes $\le x$ which are equal to $a \bmod M$ and assume that $(M, a) = 1$. Then if $M \le (\log x)^A$, A a constant, we have

$$\pi(x; M, a) = \frac{\operatorname{Li} x}{\varphi(M)} + O(x \exp(-c\sqrt{\log x})), \quad (\text{as } x \to \infty). \tag{12}$$

where c depends on A only (the constant in the $O(\cdot)$ term is absolute).

For $\pi(x)$, the number of primes up to x without any restriction, we thus have $\pi(x) = \operatorname{Li}(x) + O(x \exp(-c\sqrt{\log x})$, for some absolute constant c.

These theorems guarantee that, for $x \to \infty$, the interval $[x, x+\Delta]$ has the "expected" number of primes whenever $\Delta \ge Cx/(\log x)^A$, whatever the constant A, even if we impose the condition that these primes are equal to $a \bmod M$, as long as $M \le (\log x)^B$, for any constant B.

We use the above theorems along with the p-periodicity of P to deduce that P is in fact 2-periodic on the union of 2 small sub-intervals of $[0, k-1]$.

Definition 6.1. Γ *denotes the maximum difference between succesive primes in the interval* $[0, k]$.

From Theorem A it follows, for instance, that $\Gamma = O(k/\log^{10} k)$ which is $o(k-N)$.

Assume $q < r$ are two primes in $[N, N+h]$, where $h = (k-N)/3 = \frac{\epsilon}{3}k$. (The length of the interval $[N, N+h]$ is large enough to guarantee the existence of many primes in it.) From (11) it follows that the finite sequences

$$P(0), \ldots, P(k-q) \quad \text{and} \quad P(q), \ldots, P(k)$$

are identical. Applying (11) again with r we get that the finite sequences

$$P(0), \ldots, P(k-r) \quad \text{and} \quad P(r), \ldots, P(k)$$

are identical. It follows that

$$P(j+r-q) = P(j), \quad \text{for all } j \text{ with } N+h \le j \le N+2h \text{ and } r > q \text{ primes in } [N, N+h]. \tag{13}$$

We now assume, as we may, that the difference $M = r - q$ is the smallest difference between two primes in $[N, N+h]$. By the prime number theorem $M \le C \log k$. Hence, we can apply Theorem A with modulus M. Since $\varphi(M) \le M \le C \log k$ in that case Theorem A guarantees that the number of primes equal to $a \bmod M$ in $[N, N+h]$ is at least

$$C\frac{h}{\log^2 k} \sim C\frac{k}{\log^3 k},$$

whenever $(M, a) = 1$. All that matters here is that this number is positive for large k.

Let $t \in [N, N + h]$ be the smallest prime which is equal to $-1 \bmod M$. By Theorem A, applied to modulus M and residue -1, its existence is guaranteed and furthermore that $t \sim N$. The same theorem guarantees that we can find a prime $s \in (t, N + h]$ such that $s = 1 \bmod M$. Then $s-t = 2 \bmod M$ or $s-t = \ell M+2$, for some nonnegative integer ℓ. Therefore, for $N+h \le j \le N+2h$ we have

$$\begin{aligned} P(j) &= P(j + s - t) \ \text{(applying (13) for the primes } s, t) \\ &= P(j + \ell M + 2) \\ &= P(j + (\ell - 1)M + 2) \ \text{(applying (13) for the primes } r, q) \\ &\cdots \\ &= P(j + 2). \end{aligned}$$

This 2-periodicity

$$P(j) = P(j + 2) \tag{14}$$

is now transferred to all $j, j + 2 \in \mathcal{A}$ by using (11) repeatedly for appropriate primes p.

We use the following observation: if $P(j)$ is 2-periodic in an interval $[a, b] \subseteq [0, k]$ and $j \in [0, k]$ is such that there exists a prime $p \ge N$ for which $j + p, j + 2 + p \in [a, b]$ or $j - p, j + 2 - p \in [a, b]$ then $P(j) = P(j + 2)$.

Since we know that P is 2-periodic in the interval $[N+h, N+2h]$, we first apply the observation to obtain the 2-periodicity in the interval $[0, 2h - \Gamma]$, since for any j in that interval we can find an appropriate prime to apply the observation.

Using this new interval we now get the 2-periodicity in the interval $[N + \Gamma, k]$. Next we deduce the 2-periodicity in the interval $[0, k - N - \Gamma]$.

□

Notice that in the sequence D_j, if one erases the 0's, one sees an alternation of -1 and 1 (this follows from the fact that $f_j \in \{0, 1\}$). This property greatly reduces the number of allowed patterns in D_j and in fact it implies that P is constant in $\mathcal{A}$.

Lemma 6.3. *The polynomial P is constant in $\mathcal{A}$ (defined in Lemma 6.2).*

Proof. From Lemma 6.2 the values of P in $[N + \Gamma, k - 1]$ must be a 2-periodic sequence. The only essentially different non-constant 2-periodic patterns for the values of P in $[N + \Gamma, k - 1]$ are $010101\ldots$ and $(-1)1(-1)1\ldots$ and they both violate the property that $D_j = (-1)^j P(j)$ must satisfy, namely that if one erases the 0's then one must see an alternation of 1 and -1. Therefore P is constant in each of the two intervals of $\mathcal{A}$. From the p-periodicity (11), applied, say, for some $p \sim (k + N)/2$ it follows that the constant is the same in both intervals. □

We now extend the set on which P is constant to a superset of $\mathcal{A}$ that contains a small interval around $k/2$.

Lemma 6.4. *Let $a = \frac{N}{2} + \frac{3\Gamma}{2}$ and $b = \frac{N}{2} + (k - N) - \frac{5\Gamma}{2}$. Then $P(l) = P(0)$ for $a \le l \le b$.*

Proof. We shall apply Lemma 5.1 with $m = 2$ and with a prime r such that $2r$ is the least possible such number larger than $N + \Gamma$. It follows that $2r \le (N + \Gamma) + 2\Gamma = N + 3\Gamma$. And it follows from the remark after (8) that

$$\sum_j (-1)^j \binom{2r}{j} P(j + \nu) = 0, \quad (\nu \in \mathbb{Z}). \tag{15}$$

Taking residues mod r and using Lemma 5.1 for $m = 2$ we obtain

$$P(\nu) - 2P(\nu + r) + P(\nu + 2r) = 0 \bmod r, \quad (\nu \in \mathbb{Z}).$$

By our particular choice of r we have $P(\nu) = P(\nu + 2r) = P(0)$ whenever $\nu \in [0, k - N - 3\Gamma]$. It follows that $P(\nu + r) = P(0)$ for all such ν so we get $P(l) = P(0)$ for all l in the interval

$$\left[\frac{N}{2} + \frac{3\Gamma}{2}, \frac{N}{2} + (k - N) - \frac{5\Gamma}{2}\right].$$

□

So far we have proved $P(l) = P(0)$ on the set (a, b are defined in Lemma 6.4)

$$\mathcal{A}_2 = [0, k - N - \Gamma] \cup [a, b] \cup [N + \Gamma, k - 1],$$

which consists of three asymptotically equispaced intervals of asymptotic size ϵk. We consider two cases for P. The first is when P is 0 on $\mathcal{A}_2$ and the second is when P is 1 or -1.

To eliminate the case that P is 0 on $\mathcal{A}_2$, we shall need the following theorem, which already gives a lot of significant information about the function f. It should be thought of as analogous to the fact that the moments of a vector random variable can be read off the Fourier Transform of its distribution (the *characteristic function*) by looking at partial derivatives at 0.

Theorem 6.5. *Suppose $f : G = \mathbb{Z}_2^k = \{0, 1\}^k \to \mathbb{R}$ is nonnegative and not identically 0 and has all its Fourier coefficients of order at most r (and at least 1) equal to 0. Let μ denote the uniform probability measure on the cube G and ν denote the probability measure on G defined by*

$$\nu(A) = \sum_{x \in A} f(x) \Big/ \sum_{x \in G} f(x), \quad (A \subseteq G).$$

Let also $X_1, \ldots, X_k$ denote the coordinate functions on G, which we view as random variables. Then for all $i_1 < i_2 < \cdots < i_s$, $0 \le s \le r$, we have

$$\mathbf{E}_\nu(X_{i_1} \cdots X_{i_s}) = \mathbf{E}_\mu(X_{i_1} \cdots X_{i_s}).$$

Proof. Let $F = \sum_{x \in G} f(x)$. We assume for simplicity that $i_1 = 1, \ldots, i_s = s$. Then, writing $x = (x_1, x_2, \ldots, x_k)$ and $[s] = \{1, \ldots, s\}$, we have

$$\begin{aligned}
\mathbf{E}_\nu(X_1 \cdots X_s) &= \frac{1}{F} \sum_{x \in G} f(x) x_1 \cdots x_s \\
&= \frac{1}{F} \sum_{x \in G} f(x) \frac{1 + (-1)^{x_1 + 1}}{2} \cdots \frac{1 + (-1)^{x_s + 1}}{2} \\
&= \frac{1}{2^s F} \sum_{x \in G} f(x) \sum_{S \subseteq [s]} (-1)^{|S| + \sum_{i \in S} x_i} \\
&= \frac{|G|}{2^s F} \sum_{S \subseteq [s]} (-1)^{|S|} \frac{1}{|G|} \sum_{x \in G} f(x) (-1)^{\sum_{i \in S} x_i} \\
&= \frac{|G|}{2^s F} \sum_{S \subseteq [s]} (-1)^{|S|} \widehat{f}(S) \\
&= \frac{|G|}{2^s F} \widehat{f}(0) \quad \text{(by the vanishing of } \widehat{f}(S) \text{ for } \emptyset \ne S \subseteq [s]) \\
&= 2^{-s} \\
&= \mathbf{E}_\mu(X_1 \cdots X_s)
\end{aligned}$$

□

Remarks.
1. For functions $f : \{0,1\}^k \to \{0,1\}$, which is all we shall need here, the above theorem also follows directly from the definition of t-nullity in Section 4.
2. If the nonnegative function f is symmetric then the identity of moments up to order r with those of the uniform distribution (r-wise independence) and the vanishing of the non-constant Fourier coefficients of weight up to r are equivalent. This can be proved by induction on r. We do not use this here.

Corollary 6.6. *Under the assumptions and definitions of Theorem 6.5 the random variable $S = X_1 + \cdots + X_k$ has the same power moments $\mathbf{E}(S^s)$ under the probability measures μ and ν, up to order $s \leq r$.*

Proof. The power S^s, $s \leq r$, can be written as a sum of terms of the type $X_{i_1} \cdots X_{i_t}$, for $t \leq s$. One uses the fact that $X_j^2 = X_j$. □

Lemma 6.7. *If P is 0 on $\mathcal{A}_2$, then f is constant.*

Proof. Suppose the polynomial P is constantly equal to 0 on the set $\mathcal{A}_2$ and that f is not constant. The sequence f_j is then constant in each of the three intervals of $\mathcal{A}_2$. By possibly considering $1 - f$ (whose Fourier coefficients vanish exactly where those of f do, if f is not a constant function), we may assume that $f_j = 0$ on the middle interval (a, b). Let τ be the distribution of the random variable $S = X_1 + \cdots + X_k$ under the measure induced by f on G (each vertex $x \in G$ has probability proportional to $f(x)$), where $X_1, \ldots, X_k$ are the coordinate functions on G. Note that this is a well defined probability distribution since we assumed that f is not the $\mathbf{0}$ function.

The s-th moment with respect to the measure τ of the variable S in Corollary 6.6 is the expression

$$M(\tau, s) = \frac{1}{F} \sum_j f_j \binom{k}{j} j^s,$$

where again $F = \sum_j f_j \binom{k}{j}$. By Corollary 6.6, if $s \leq k - N$ this moment must equal the s-th moment with respect to the binomial measure μ, which is the quantity

$$M(\mu, s) = 2^{-k} \sum_j \binom{k}{j} j^s.$$

But the variance of S under μ is

$$M(\mu, 2) - M(\mu, 1)^2 = k, \tag{16}$$

since under μ the random variables $X_1, \ldots, X_k$ are independent, while the variance of S under τ is

$$\mathbf{E}_\tau (S - \mathbf{E}_\tau S)^2 = \mathbf{E}_\tau (S - \mathbf{E}_\mu S)^2 = \mathbf{E}_\tau (S - k/2)^2 \geq C\epsilon^2 k^2 \tag{17}$$

as the mass of τ sits to the left of $a \sim k/2 - \epsilon k/2$ and to the right of $b \sim k/2 + \epsilon k/2$. The orders of magnitude in (16) and (17) are different whenever $\epsilon \geq C/\sqrt{k}$, which is true in our case as $\epsilon = 4/\log k$. This contradiction proves that P cannot equal 0 on $\mathcal{A}_2$. □

Extending $\mathcal{A}_2$ to $[0, k-1]$.

For $2^l = m = 2, 4, \ldots$, we define the sets

$$B_m = \bigcup_{j=0}^{m} \left[\frac{j}{m} N + \Delta(m), \frac{j}{m} N + \epsilon k - \Delta(m) \right],$$

where $\Delta(m) = \Delta(m/2) + m\Gamma$, for $m \geq 4$, and $\Delta(2) = 3\Gamma$. (These intervals will be overlapping when m is large.)

Lemma 6.8. *There is a constant $k_0 > 0$ such that if $k \geq k_0$ and $\epsilon = 4/\log k$ then*
(a) the polynomial P is equal to 1 on $B_m \cap [0, k-1]$, for $m = 2, 4, 8, \ldots$ with $m \leq \frac{1}{2} \log k$, and
(b) if m takes the highest value allowed in (a) then B_m covers $[0, k-1]$, hence $P = 1$ on $[0, k-1]$.

Proof. To prove (a) we work by induction on $m = 2, 4, \ldots$. The base case $m = 2$ is settled since we have $B_2 \subseteq \mathcal{A}_2$ (that's why we chose $\Delta(2)$ large enough).

Assume now that we have proved $P = 1$ on $B_{m/2} \cap [0, k-1]$. We apply Theorem 5.1 for m and we choose a prime r such that mr is the least possible larger than N. Thus

$$N/m \leq r \leq N/m + \Gamma. \tag{18}$$

Lemma 5.1 together with relation (15) gives for all $\nu \in \mathbb{Z}$

$$P(\nu) - mP(\nu + r) + \binom{m}{2} P(\nu + 2r) - \cdots + (-1)^m P(\nu + mr) = 0 \bmod r. \tag{19}$$

We would like, for j even, the number $\nu + jr$ to belong to $B_{m/2}$, for most values of ν in the interval $[0, \epsilon k]$. That is we want

$$\frac{j}{m} N + \Delta(m/2) \leq \nu + jr \leq \frac{j}{m} N + \epsilon k - \Delta(m/2),$$

for $0 \leq j \leq m$, j even. Given (18) this follows from

$$\Delta(m/2) \leq \nu \leq \epsilon k - \Delta(m/2) - m\Gamma. \tag{20}$$

For ν satisfying (20) the range of the expression $\nu + jr$ (j fixed) contains the interval

$$[jr + \Delta(m/2), jr + \epsilon k - \Delta(m/2) - m\Gamma],$$

which, using (18) again, contains the interval

$$\left[\frac{j}{m} N + m\Gamma + \Delta(m/2), \frac{j}{m} N + \epsilon k - \Delta(m/2) - m\Gamma \right].$$

From the relation $\Delta(m) = \Delta(m/2) + m\Gamma$ it follows that this last interval is the j-th interval of B_m.

We have shown that whenever ν satisfies (20) the numbers $\nu + jr$, $0 \leq j \leq m$, j even, are all in $B_{m/2}$ so, by the induction hypothesis, the polynomial P takes the value 1 on them.

In the left hand side of (19) the sum of the absolute values of the coefficients is at most 2^m and as long as $2^m < r$ it follows that (mod r) can be dropped from (19). If (20) is satisfied it is clear that the sum of the terms of (19) corresponding to even j is 2^{m-1}, since these P terms are all 1. If, in addition $2^m < r$, we obtain that the terms corresponding to odd j must all have their P term

equal to 1. The reason for this is that the sum of absolute values of the odd terms is at most 2^{m-1} and is equal to that only in case all P's are equal to 1.

Letting ν run through all terms allowed by (20) we obtain that P has the value of 1 on all intervals of B_m corresponding to odd j. Since the intervals corresponding to even j are already contained in $B_{m/2}$ we obtain the desired conclusion, that P is equal to 1 on B_m, as long as $2^m < r$, which is clearly satisfied if $2^m < N/m$ or

$$m \leq \frac{1}{2} \log k. \tag{21}$$

This concludes the proof of (a).

To prove (b) observe that $\Delta(m) \leq 2m\Gamma$. Letting $\epsilon = 4/\log k$, we observe that if we let m be as large as part (a) allows then each of the intervals of B_m overlaps with the next one thus covering all of the interval $[0, k-1]$, which proves (b) and that P is constantly equal to 1, as we had to prove. □

7 Learning symmetric juntas

In this section we apply Theorem 2.2 to obtain faster learning algorithms for the class of symmetric k-juntas on n variables. First we need some preliminaries and well known tools from computational learning theory.

7.1 Preliminaries

We consider the PAC learning model [16]. The learning problem at hand is a *Concept Class* $\mathcal{C} = \bigcup_n \mathcal{C}_n$, where each $\mathcal{C}_n$ is a collection of boolean functions from $\{0,1\}^n \to \{0,1\}$. Let ϵ be an *accuracy parameter* and δ a *confidence parameter*. A learning algorithm $\mathcal{A}$ for $\mathcal{C}$ has access to an *oracle* $\mathcal{I}(f)$ for $f \in \mathcal{C}_n$. A query to $\mathcal{I}(f)$ outputs a labeled example $\langle \mathbf{x}, f(\mathbf{x}) \rangle$, where $\mathbf{x}$ is drawn from $\{0,1\}^n$ according to some probability distribution. $\mathcal{A}$ is said to be a learning algorithm for the class $\mathcal{C}$ if for all $f \in \mathcal{C}$, when $\mathcal{A}$ is run with oracle $\mathcal{I}(f)$, it outputs, with probability at least $1-\delta$, a hypothesis h such that $\Pr_{\mathbf{x}}[h(\mathbf{x}) = f(\mathbf{x})] \geq 1-\epsilon$. Although Valiant's PAC model is defined for general distributions, in this paper we will be concerned only with the uniform distribution.

We recall the definition of a k-junta. Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. We say that f *depends* on the variable i, if there are vectors $\mathbf{x}$ and $\mathbf{y}$ that differ only on the i'th coordinate and $f(\mathbf{x}) \neq f(\mathbf{y})$. A function that depends only on an (unknown) subset of $k \ll n$ variables is called a k-junta. The variables on which f depends are called the *relevant* variables of f. Typically $k = O(\log n)$. Hence, a running time that is polynomial in $2^k, n$ and $\log(1/\delta)$ is considered efficient. A symmetric k-junta is a boolean function which is symmetric in the variables it depends on. The class of all such functions defined on n variables is the class of symmetric k-juntas. In this section, we present an algorithm for learning this class in the uniform PAC model.

7.2 Analysis of the Fourier based algorithm

We will use the following facts about learning in the PAC model which are well known.

(i) We can exactly calculate the Fourier coefficients of the target function with confidence $1-\delta$ in time $poly(\log 1/\delta,\, 2^k, n)$ using standard Chernoff-Hoeffding bounds (see [11, 14]).

(ii) We can decide whether the target function f is constant or not in time $poly(\log 1/\delta, 2^k)$.

(iii) We can learn a parity function in time $n^{\omega} poly(\log 1/\delta, 2^k)$ [7]. Here ω is the exponent for matrix multiplication, $\omega < 2.376$.

We state the standard Fourier based algorithm below:
Throughout the algorithm, we maintain a set of relevant variables, R.

- Check if the function is constant or parity.
- If not, set $R := \emptyset$, $t := 1$.
  1. For every subset of t variables, say $S = \{x_{i_1}, ..., x_{i_t}\}$ do:
     (a) Compute $\hat{f}(S)$.
     (b) If $\hat{f}(S) \neq 0$, then $R := R \cup S$.
  2. If for all sets S of size t, $\hat{f}(S) = 0$ then $t := t + 1$ and go to step 1.
  3. Else, R now contains all the relevant variables. Draw enough samples to build f's truth table and halt.

If x_i is an irrelevant variable for f, then it is easy to see that for any S containing x_i, $\hat{f}(S) = 0$. Hence, if $\hat{f}(S) \neq 0$, for some S, then S contains only relevant variables. Since the function is symmetric, for any two sets S, T of relevant variables such that $|S| = |T|$, we have $\hat{f}(S) = \hat{f}(T)$. Hence, the first time that we will identify some relevant variables in the algorithm ($\hat{f}(S) \neq 0$ for some S, $|S| = s$), we will actually be able to identify all the relevant variables, and the running time will be roughly n^s. Hence, as a direct consequence of Theorem 2.2, we obtain a bound of $n^{o(k)}$ for learning symmetric juntas.

Theorem 7.1. *The class of symmetric k-juntas can be learned exactly under the uniform distribution with confidence $1 - \delta$ in time $n^{O(k/\log k)} \cdot \mathrm{poly}(2^k, n, \log(1/\delta))$.*

8 Discussion

The main open question is to obtain tight upper and lower bounds on the running time of the Fourier-based algorithm for symmetric juntas. It may even be that for large k, every symmetric function has a non-zero Fourier coefficient of constant order.

It should also be noted that in the case of balanced symmetric functions, i.e., symmetric functions with $Pr[f(x) = 1] = 1/2$, a bound of $O(k^{0.548})$ follows from [19] (see [14]). Hence, to improve our result, one may focus on finding new techniques for unbalanced functions.

References

[1] A. Blum. Relevant examples and relevant features: Thoughts from computational learning theory. In *AAAI Symposium on Relevance*, 1994.

[2] A. Blum. Open problems. COLT, 2003.

[3] A. Blum, M. Furst, M. Kearns, and R. J. Lipton. Cryptographic primitives based on hard learning problems. In *CRYPTO*, pages 278–291, 1993.

[4] A. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97:245–271, 1997.

[5] N. Bshouty, J. Jackson, and C. Tamon. More efficient PAC learning of DNF with membership queries under the uniform distribution. In *Annual Conference on Computational Learning Theory*, pages 286–295, 1999.

[6] P. Cameron. *Combinatorics: topics, techniques, algorithms*. Cambridge Univ. Press, 1994.

[7] D. Helmbold, R. Sloan, and M. Warmuth. Learning integer lattices. *SIAM Journal of Computing*, 21(2):240–266, 1992.

[8] J. Jackson. An efficient membership-query algorithm for learning dnf with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55:414–440, 1997.

[9] M. Kolountzakis, E. Markakis, and A. Mehta. Learning symmetric juntas in time $n^{o(k)}$. In *Proceedings of the conference Interface entre l'analyse harmonique et la theorie des nombres, CIRM, Luminy*, 2005.

[10] A. Kumchev. *The distribution of prime numbers*. manuscript, 2005.

[11] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[12] R. Lipton, E. Markakis, A. Mehta, and N. Vishnoi. On the fourier spectrum of symmetric boolean functions with applications to learning symmetric juntas. In *IEEE Conference on Computational Complexity*, pages 112–119, 2005.

[13] Y. Mansour. An $o(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50:543–550, 1995.

[14] E. Mossel, R. O'Donnel, and R. Servedio. Learning juntas. In *STOC*, pages 206–212, 2003.

[15] G. Pólya and G. Szego. *Problems and theorems in Analysis, II*. Springer, 1976.

[16] L. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[17] K. Verbeurgt. Learning DNF under the uniform distribution in quasi-polynomial time. In *Annual Workshop on Computational Learning Theory*, pages 314–326, 1990.

[18] K. Verbeurgt. Learning sub-classes of monotone DNF on the uniform distribution. In *Michael M. Richter, Carl H. Smith, Rolf Wiehagen, and Thomas Zeugmann, editors, Algorithmic Learning Theory, 9th International Conference*, pages 385–399, 1998.

[19] J. von zur Gathen and J. Roche. Polynomials with two values. *Combinatorica*, 17(3):345–362, 1997.