On the Minimal Fourier Degree of Symmetric Boolean Functions

Amir Shpilka[1] Avishay Tal[2]

[1]Technion & Microsoft Research

[2]Technion

IEEE Conference on Computational Complexity 2011

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Basic Definitions
Symmetric Boolean Functions

- **Boolean**:
$$f : \{0,1\}^k \to \{0,1\}$$
- **Symmetric**:
$$\forall \sigma \in S_k \ : \ f(x_1, x_2, \ldots, x_k) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(k)})$$
- $f(\vec{x})$ depends only on the weight of $\vec{x}$
- Equivalently:
$$F : \{0,1,\ldots,k\} \to \{0,1\}$$
- $f(\vec{x}) = F(|\vec{x}|)$
Different representations of the same function
- **Examples**: PARITY, MAJORITY, AND, OR

Boolean Functions Fourier Transform

- Mapping $b \mapsto (-1)^b \ : \ 0 \mapsto 1 \ , \ 1 \mapsto -1$ yields:

$$f : \{0,1\}^k \to \{0,1\} \rightsquigarrow f : \{-1,1\}^k \to \{-1,1\}$$

- For $S \subseteq \{1, \ldots, k\}$ denote the S-**character** as

$$\chi_S(\vec{x}) \triangleq \prod_{i \in S} x_i$$

- The set of characters is an orthonormal basis for the space of k-variables boolean functions with respect to this inner product:

$$< f_1, f_2 >= \frac{1}{2^k} \cdot \sum_{\vec{x} \in \{-1,1\}^k} f_1(\vec{x}) \cdot f_2(\vec{x})$$

Boolean Functions Fourier Transform

- **Fourier Transform (FT)**: One can represent f as a k-variate polynomial over $\mathbb{R}$:

$$f(x_1, x_2, \ldots, x_k) = \sum_{S \subseteq \{1,\ldots,k\}} \hat{f}(S) \cdot \prod_{i \in S} x_i$$

- $\hat{f}(S)$ is called the **Fourier coefficient** of the subset S
- Since the characters are orthonormal we can calculate the Fourier coefficient as follows:

$$\hat{f}(S) = < f, \chi_S > = \frac{1}{2^k} \sum_{\vec{x} \in \{-1,1\}^k} f(\vec{x}) \cdot \prod_{i \in S} x_i$$

Symmetric Boolean Functions Fourier Transform

- If f is **symmetric**, then $\hat{f}(S)$ depends only on $|S|$

Examples:

- $PARITY^k = \chi_{\{1,\ldots,k\}} = x_1 \cdot x_2 \cdot \ldots \cdot x_k$
- $MAJORITY^3 = \frac{1}{2} \cdot (x_1 + x_2 + x_3 - x_1 \cdot x_2 \cdot x_3)$

Research Questions

- What can we say about:
  - Degree of Fourier representation of f?
  - Minimal size $S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?
- Rule out the degenerated cases where f is constant, PARITY or $\neg$PARITY ("linear" functions)

Research Questions

- What can we say about:
  - Degree of Fourier representation of f?
  - Minimal size $S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?
- Rule out the degenerated cases where f is constant, PARITY or $\neg$PARITY ("linear" functions)
- Questions are closely related:
  - $\hat{f}(S) = (f \oplus \widehat{PARITY})(\overline{S})$
  - Upper bound on the minimal nonzero term in the FT $\Leftrightarrow$ Lower bound on the FT degree

Research Questions

- What can we say about:
  - Degree of Fourier representation of f?
  - Minimal size $S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?
- Rule out the degenerated cases where f is constant, PARITY or $\neg$PARITY ("linear" functions)
- Questions are closely related:
  - $\hat{f}(S) = \widehat{(f \oplus PARITY)}(\overline{S})$
  - Upper bound on the minimal nonzero term in the FT $\Leftrightarrow$ Lower bound on the FT degree

Main Question

What is the minimal size $\emptyset \neq S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?

Research Questions

- What can we say about:
  - Degree of Fourier representation of f?
  - Minimal size $S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?
- Rule out the degenerated cases where f is constant, PARITY or $\neg$PARITY ("linear" functions)
- Questions are closely related:
  - $\hat{f}(S) = \widehat{(f \oplus PARITY)}(\overline{S})$
  - Upper bound on the minimal nonzero term in the FT $\Leftrightarrow$ Lower bound on the FT degree

Main Question

What is the minimal size $\emptyset \neq S \subseteq \{1, \ldots, k\}$ such that $\hat{f}(S) \neq 0$?

- If we could guaranty that $\hat{f}(\emptyset) = 0$ (i.e. that f is balanced), then previous upper bounds would still hold

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Learning Juntas

Learning in the presence of irrelevant features

Real life example:

Understanding how a genetic feature depends on the DNA sequence

- **Definition**: a k-**junta** is an unknown $g(x_1, x_2, \ldots, x_n)$ that actually depends only on $k \ll n$ variables: $g(x_1, x_2, \ldots, x_n) = f(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$
- **Goal**: Given uniformly random labeled examples $< \vec{x}, g(\vec{x}) >$ find, with high probability:
  - the k relevant variables
  - the underlying function f
- Naive solution: Check consistency for all possible subsets - $O(\binom{n}{k} \cdot \mathrm{poly}(2^k, n))$

Fourier Based Learning Algorithm

Learning Algorithm:

- for all non empty subsets $S \subseteq \{1, 2, \ldots, n\}$, ordered by size:
  - estimate $\hat{g}(S)$
  - if $\hat{g}(S) \neq 0$ then S is relevant! halt!

- $\hat{g}(S) = \begin{cases} \hat{f}(S) & S \subseteq \{i_1, i_2, \ldots, i_k\} \\ 0 & \text{otherwise} \end{cases}$
- One can estimate $\hat{g}(S)$ with error rate δ using $poly(2^k, \log(1/\delta))$ samples by:

$$\hat{g}_{estimate}(S) = \frac{1}{|Samples|} \cdot \sum_{\vec{x} \in Samples} g(\vec{x}) \cdot \chi_S(\vec{x})$$

- Previous work: Finding 1 relevant variable is good enough

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Previous Works

- Degree of Fourier representation of f:
  - **[vzGathen Roche]** $\deg(f) \geq k - k^{0.525}$
  - Conjecture **[vzGathen Roche]** $\deg(f) \geq k - O(1)$
- Degree of minimal term in Fourier spectrum of f:
  - **[Mossel O'Donnell Servedio]** $|S| \leq 2k/3$ (alternative proof by O.Regev)
  - **[Kolountzakis Lipton Markakis Mehta Vishnoi]** $|S| \leq O(k/log(k))$
- Learning symmetric juntas
  - **[MOS]** $n^{\frac{2}{3} \cdot k} \cdot \mathrm{poly}(n, 2^k)$
  - **[KLMMV]** $n^{O\left(\frac{k}{\log(k)}\right)} \cdot \mathrm{poly}(n, 2^k)$
- **Same algorithm, different analysis**

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Our Results

- Degree of minimal term in Fourier spectrum of f
  - $|S| \leq O(k^{0.525})$ (improving $|S| \leq O(k/\log(k))$)
  - Assuming ERH: $|S| \leq k^{0.5+o(1)}$

- Learning symmetric juntas:
  - Fourier learning algorithm runs in time $n^{O(k^{0.525})} \cdot \mathrm{poly}(n, 2^k)$ (previous analysis: $n^{O(k/\log(k))} \cdot \mathrm{poly}(n, 2^k)$)

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Preliminaries

Notations: Bias, Correlation Immune Functions, Fixing Variables

- **Definition**: $\text{bias}(f) = Pr_{\vec{x} \in_R \{0,1\}^k}[f(\vec{x}) = 1]$

Preliminaries

Notations: Bias, Correlation Immune Functions, Fixing Variables

- **Definition**: $\text{bias}(f) = Pr_{\vec{x} \in_R \{0,1\}^k}[f(\vec{x}) = 1]$
- **Definition**: $f|_{(m,r)}$: f after fixing (any) r variables to 1 and $m - r$ variables to 0
- **Examples**:
  - $MAJORITY^5|_{(2,0)} = x_1 \wedge x_2 \wedge x_3$
  - $PARITY^{20}|_{(5,3)} = PARITY^{15} \oplus 1$
- Clearly, $F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r)$ for $|x| = 0, 1, \ldots, k - m$

Preliminaries

Notations: Bias, Correlation Immune Functions, Fixing Variables

- **Definition**: $\text{bias}(f) = Pr_{\vec{x} \in_R \{0,1\}^k}[f(\vec{x}) = 1]$
- **Definition**: $f|_{(m,r)}$: f after fixing (any) r variables to 1 and $m - r$ variables to 0
- **Examples**:
  - $MAJORITY^5|_{(2,0)} = x_1 \wedge x_2 \wedge x_3$
  - $PARITY^{20}|_{(5,3)} = PARITY^{15} \oplus 1$
- Clearly, $F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r)$ for $|x| = 0, 1, \ldots, k - m$
- **Definition**: f is t-**correlation immune** if $\forall 0 \leq r \leq m \leq t$

$$\text{bias}(f) = \text{bias}(f|_{(m,r)})$$

.

Preliminaries

Notations: Bias, Correlation Immune Functions, Fixing Variables

- **Definition**: $\text{bias}(f) = Pr_{\vec{x} \in_R \{0,1\}^k}[f(\vec{x}) = 1]$
- **Definition**: $f|_{(m,r)}$: f after fixing (any) r variables to 1 and $m - r$ variables to 0
- **Examples**:
  - $MAJORITY^5|_{(2,0)} = x_1 \wedge x_2 \wedge x_3$
  - $PARITY^{20}|_{(5,3)} = PARITY^{15} \oplus 1$
- Clearly, $F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r)$ for $|x| = 0, 1, \ldots, k - m$
- **Definition**: f is t-**correlation immune** if $\forall 0 \leq r \leq m \leq t$

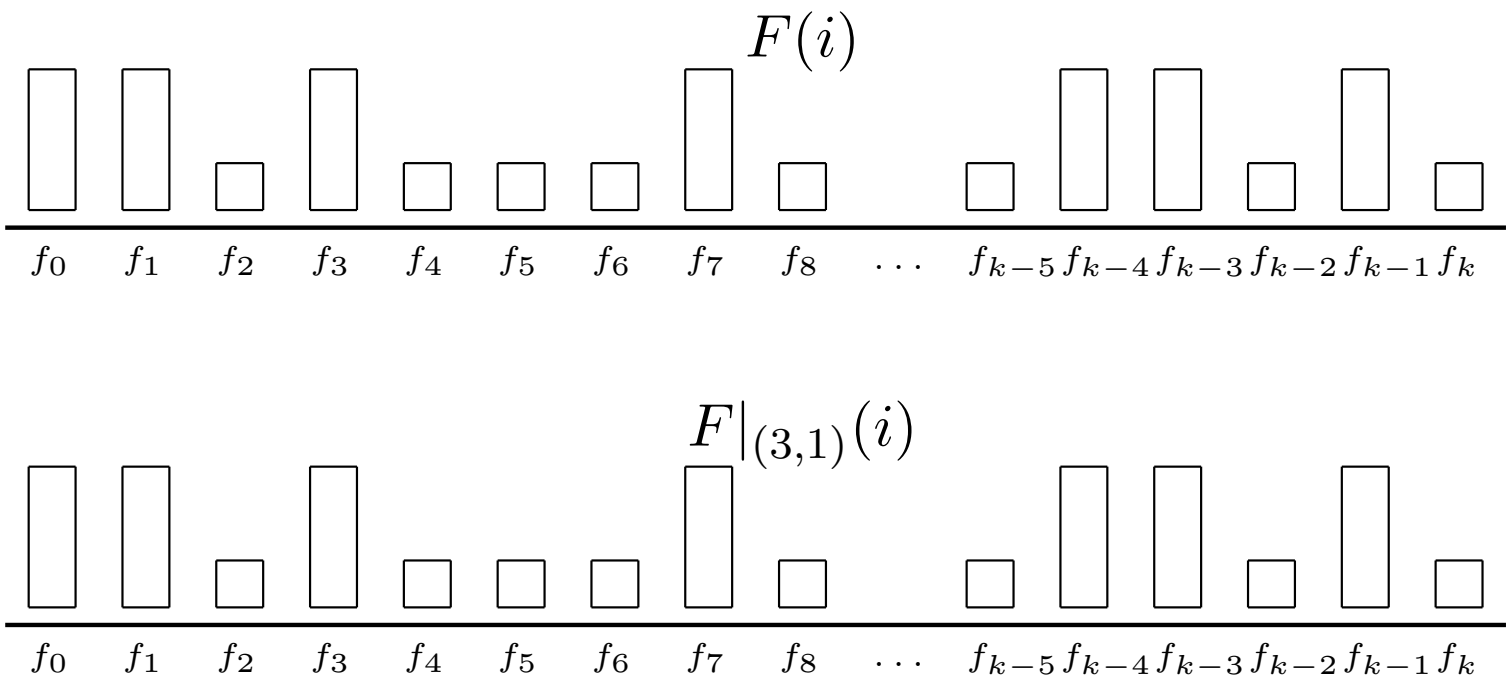
$$\text{bias}(f) = \text{bias}(f|_{(m,r)})$$

.

- **Theorem**: **[Xiao Massey]** f is t-correlation immune iff $\forall 1 \leq |S| \leq t \ : \ \hat{f}(S) = 0$

Preliminaries

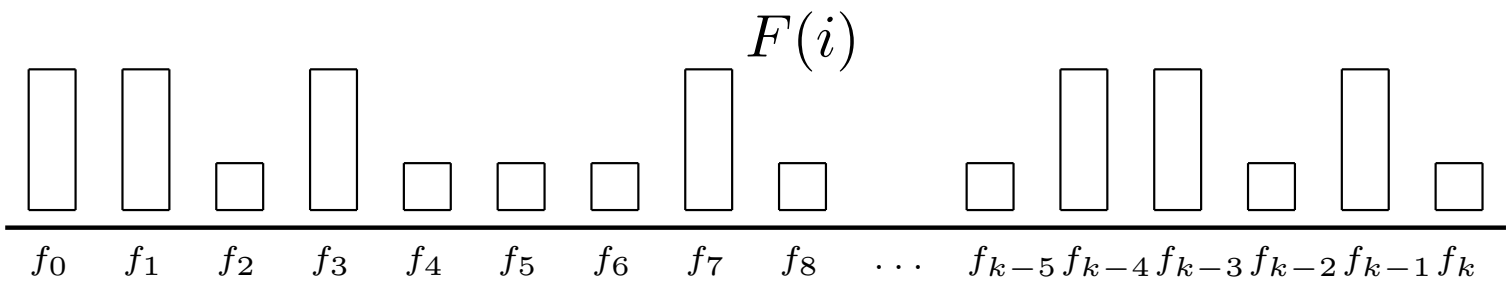
Fixing Variables Illustration

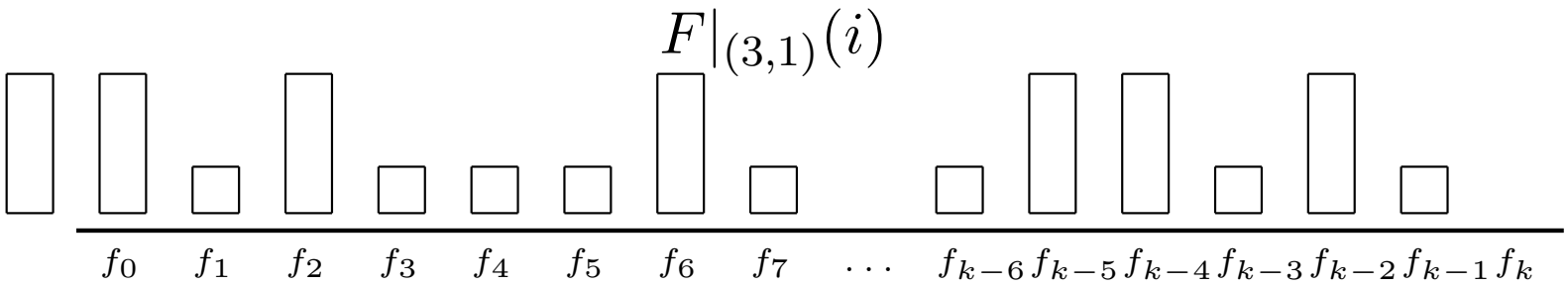


$$F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r) \text{ for } |x| = 0, 1, \ldots, k - m$$

Preliminaries

Fixing Variables Illustration

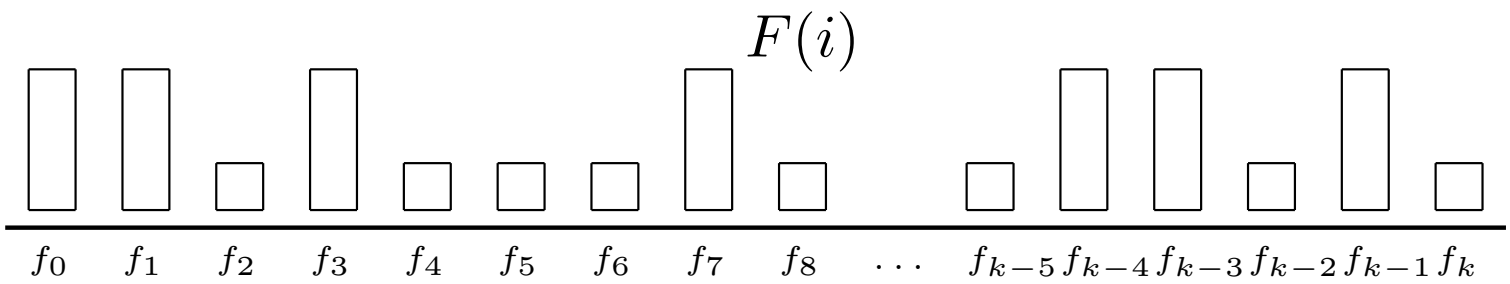


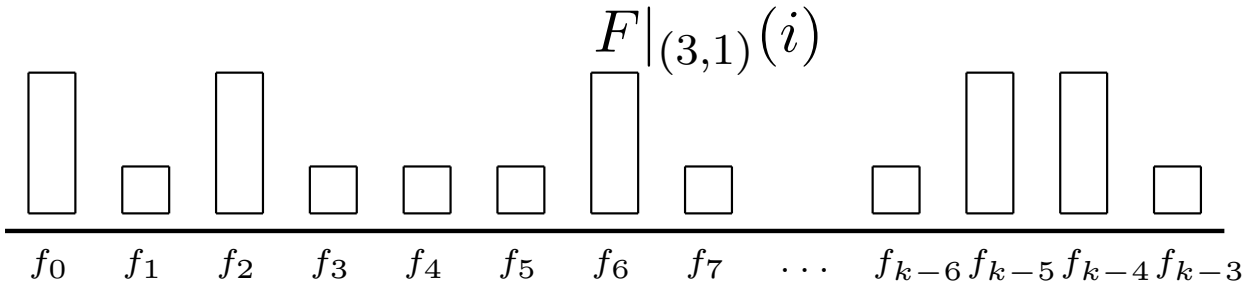


$$F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r) \text{ for } |x| = 0, 1, \ldots, k - m$$

Preliminaries

Fixing Variables Illustration





$$F|_{(m,r)}(|\vec{x}|) = F(|\vec{x}| + r) \text{ for } |x| = 0, 1, \ldots, k - m$$

Outline

1. Motivation & Background
   - Problem Definition
   - Application: Learning Symmetric Juntas
   - Previous Work
2. Our Results
   - Our Results
   - Proof Preliminaries
   - Proof Sketch

Proof Sketch
Achieving Linear Equations

Theorem: For any non-linear symmetric Boolean function, f, $\exists S \subseteq \{1,\ldots,k\};\ 1 \leq |S| = O(k^{0.525})$ such that $f(S) \neq 0$

- Assume, by contradiction, that there exists such f which is $t = \Theta(k^{0.525})$ correlation immune
- By **bias** definition:

$$\text{bias}(f) = Pr_{\vec{x} \in_R \{0,1\}^k}[f(\vec{x}) = 1] = \sum_{i=0}^{k} \frac{\binom{k}{i}}{2^k} \cdot F(i)$$

$$\text{bias}(f|_{(m,r)}) = \sum_{i=0}^{k-m} \frac{\binom{k-m}{i}}{2^{k-m}} \cdot F(i+r)$$

- By **[Xiao Massey]** for $r \leq m \leq t$: $\text{bias}(f) = \text{bias}(f|_{(m,r)})$
- This yields many linear equations on $\{F(i)\}_{i=0,\ldots,k}$
- Main obstacle: How to use the fact that the solutions are $0/1$?

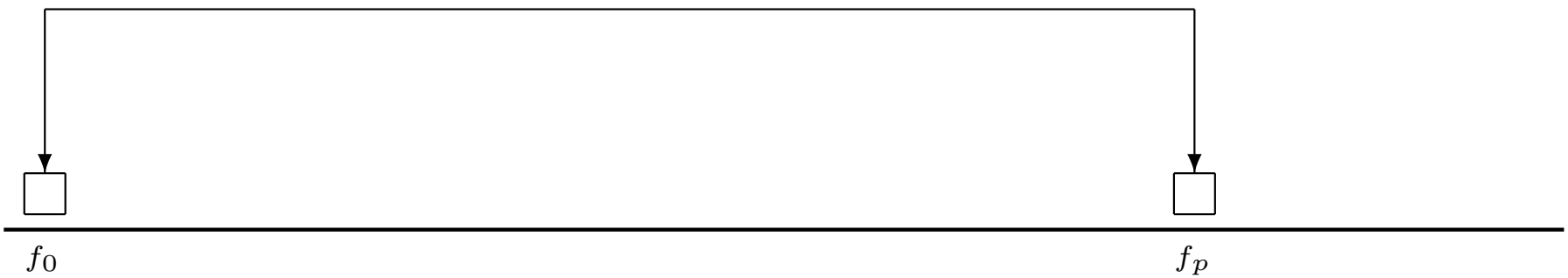
$$\text{bias}(f) \cdot 2^{k-m} = \text{bias}(f|_{(m,r)}) \cdot 2^{k-m} = \sum_{i=0}^{k-m} \binom{k-m}{i} F(i+r)$$

- By **[Baker Harman Pintz]**, there exist a prime number p such that $p = k - \Theta(k^{0.525})$
- Assuming ERH we can pick $p = k - k^{0.5+o(1)}$
- Fixing $k - p$ and reducing modulo p yields: $\text{bias}(f) \cdot 2^p \equiv_p \sum_{i=0}^{p} \binom{p}{i} F(i+r) \equiv_p F(r) + F(p+r)$
- the LHS doesn't depends on r, let's denote it by c_p

Proof Sketch
Going Modular

$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

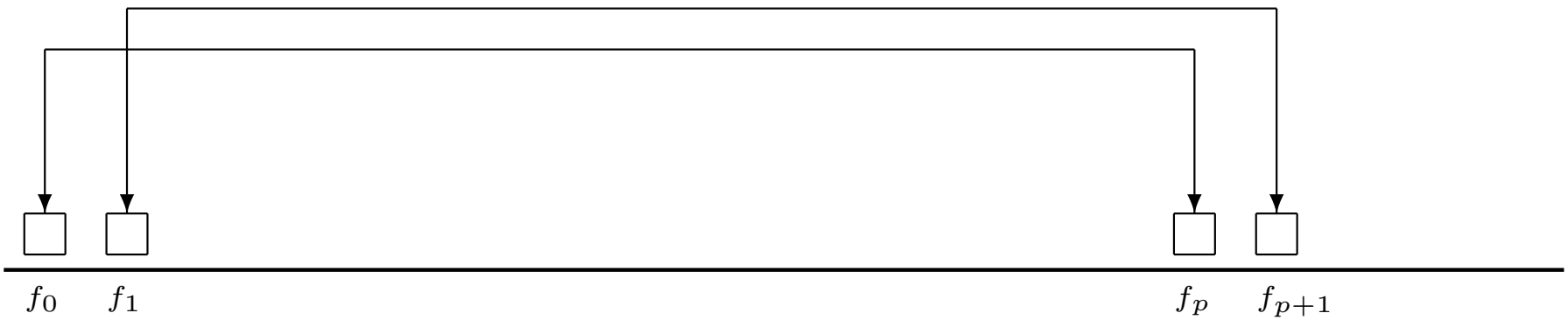
- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 0$:



Proof Sketch
Going Modular

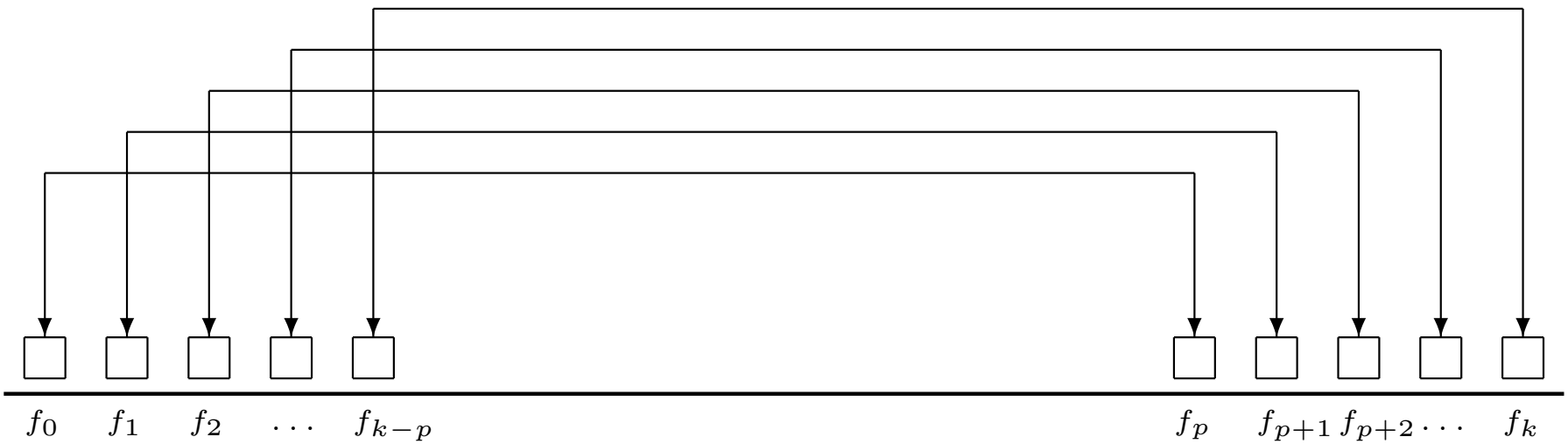
$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 0$:



$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

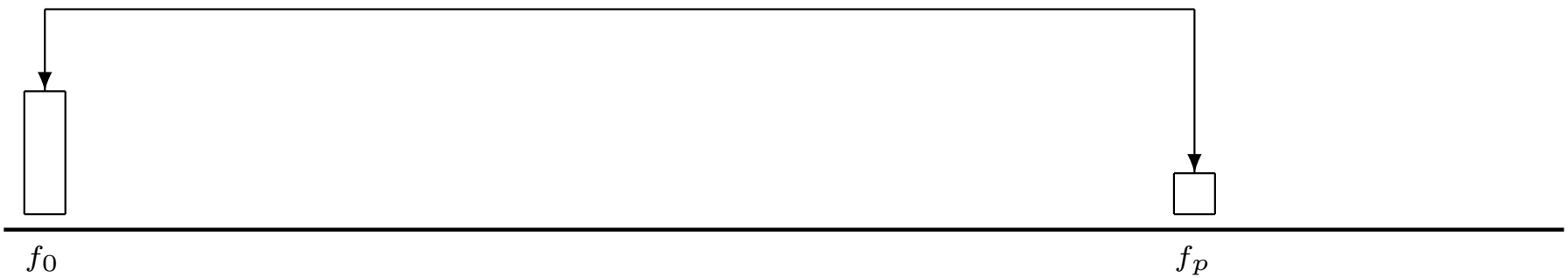
- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 0$:



Proof Sketch
Going Modular

$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

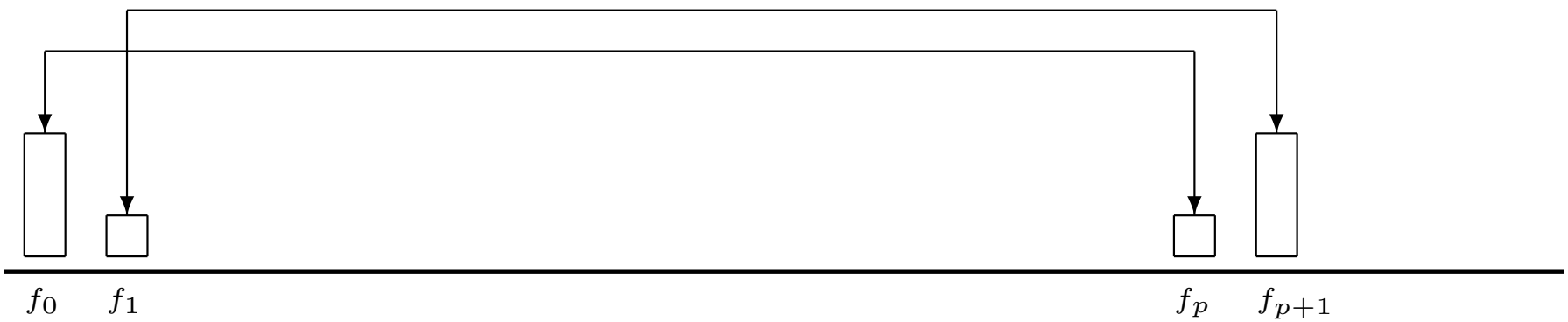
- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 1$:



Proof Sketch
Going Modular

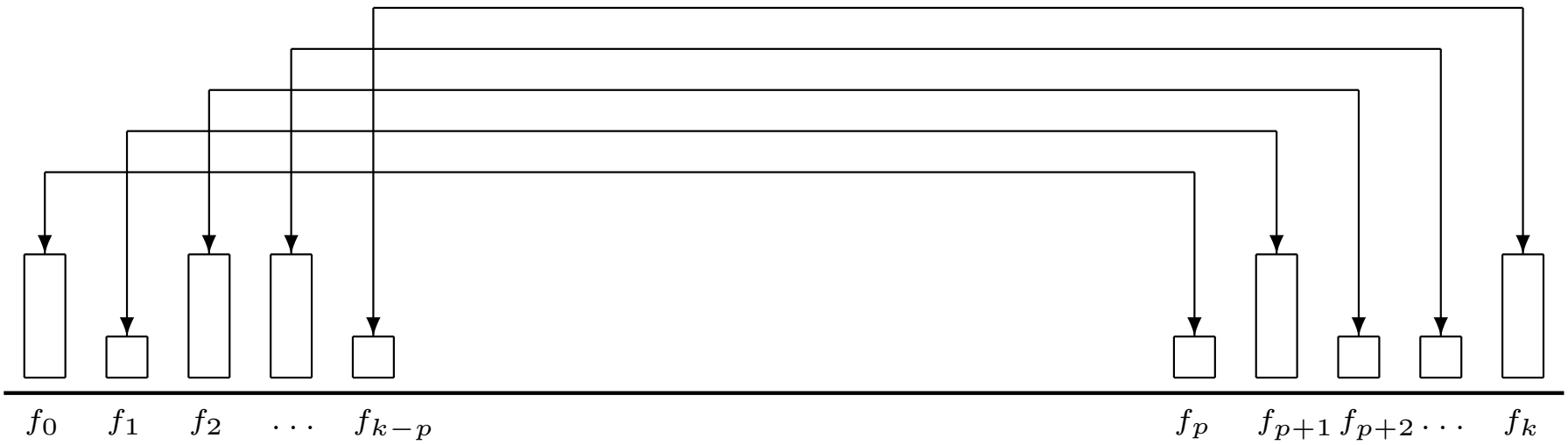
$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 1$:



$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

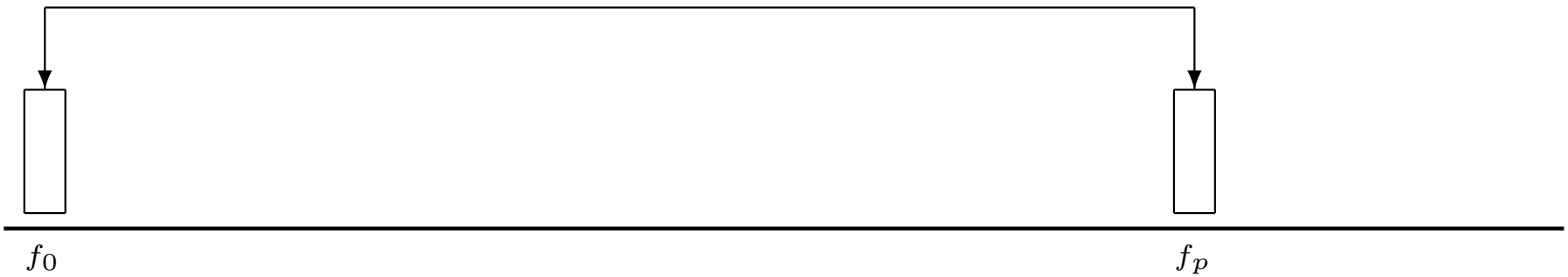
- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 1$:



Proof Sketch
Going Modular

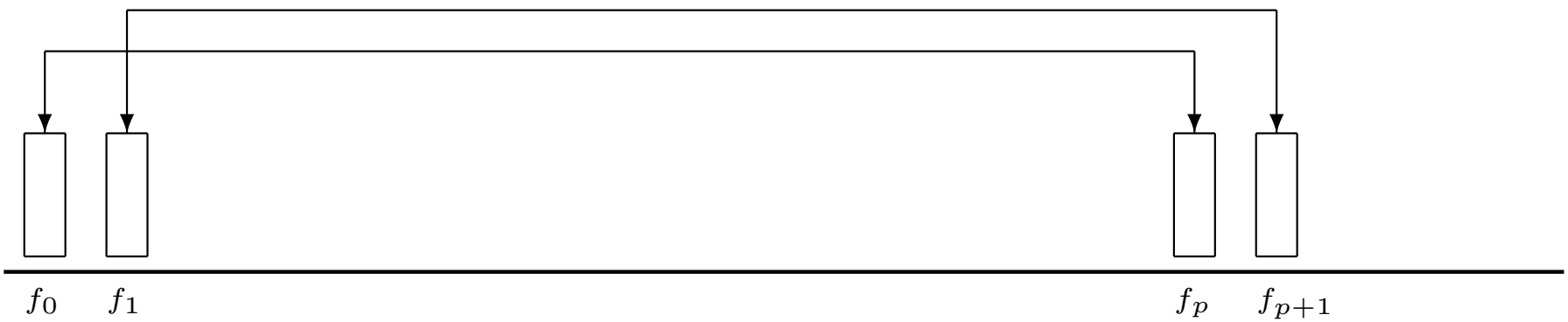
$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 2$:



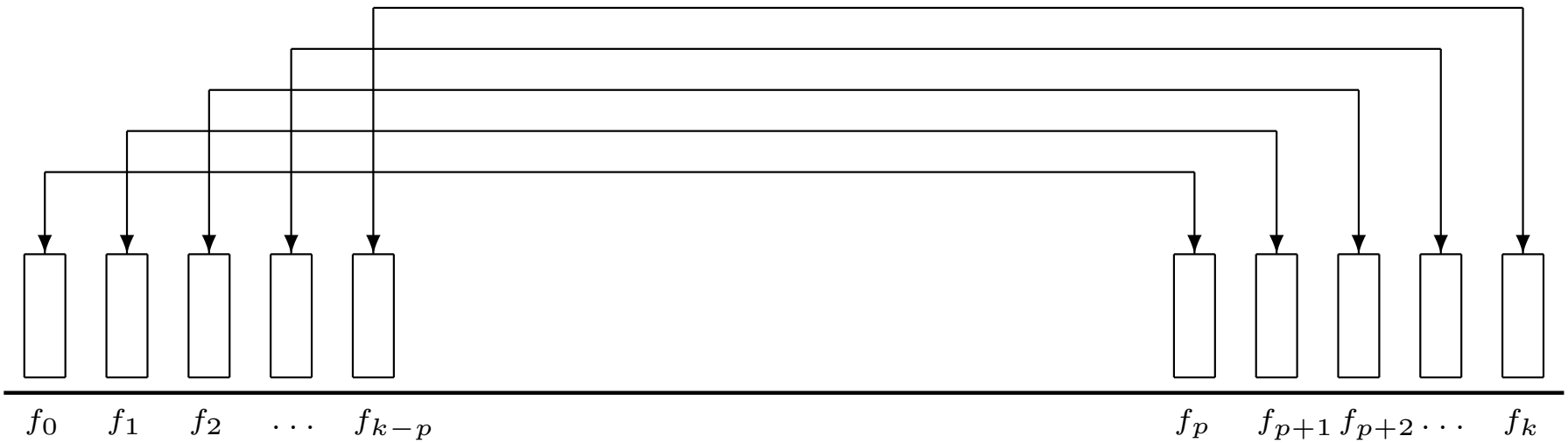
$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 2$:



$$\forall r = 0, 1, \ldots, k - p \ : \ F(r) + F(p + r) \equiv_p c_p$$

- Since $F(i)$ are 0/1 $\Rightarrow c_p (\text{mod } p) \in \{0, 1, 2\}$
- Case $c_p \equiv_p 2$:



- Fixing $k - p + 1$ bits and reducing modulo p yields:

$$1/2 \cdot c_p \equiv_p \sum_{i=0}^{p-1} \binom{p-1}{i} F(i+r) \equiv_p \sum_{i=0}^{p-1} (-1)^i F(i+r)$$

- Case $c_p \equiv_p 1$ implies that f is either $PARITY$ or $\neg PARITY$ which we excluded earlier 
- Cases $c_p \pmod p \in \{0, 2\}$: F is fixed on $[0, k-p] \cup [p, k]$

Proof Sketch

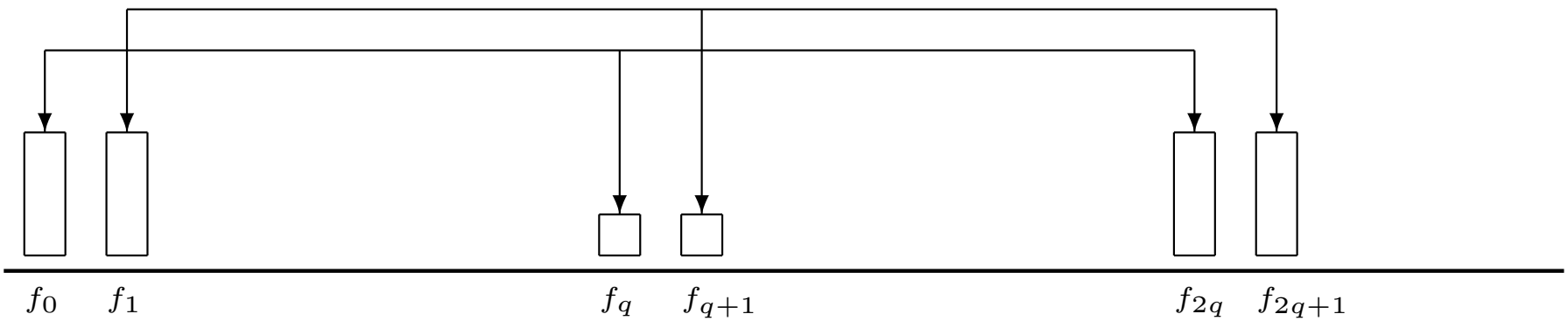
- By **[BHP]** there is a prime number, $q > \frac{p}{2}$, such that $q = \frac{k - \Theta(k^{0.525})}{2}$
- Fixing $k - 2q$ bits and reducing modulo q yields:

$$c_{2q} \equiv_q \sum_{i=0}^{2q} \binom{2q}{i} F(i+r) \equiv_q F(r) + 2 \cdot F(q+r) + F(2q+r)$$

Proof Sketch

- By **[BHP]** there is a prime number, $q > \frac{p}{2}$, such that $q = \frac{k - \Theta(k^{0.525})}{2}$
- Fixing $k - 2q$ bits and reducing modulo q yields:

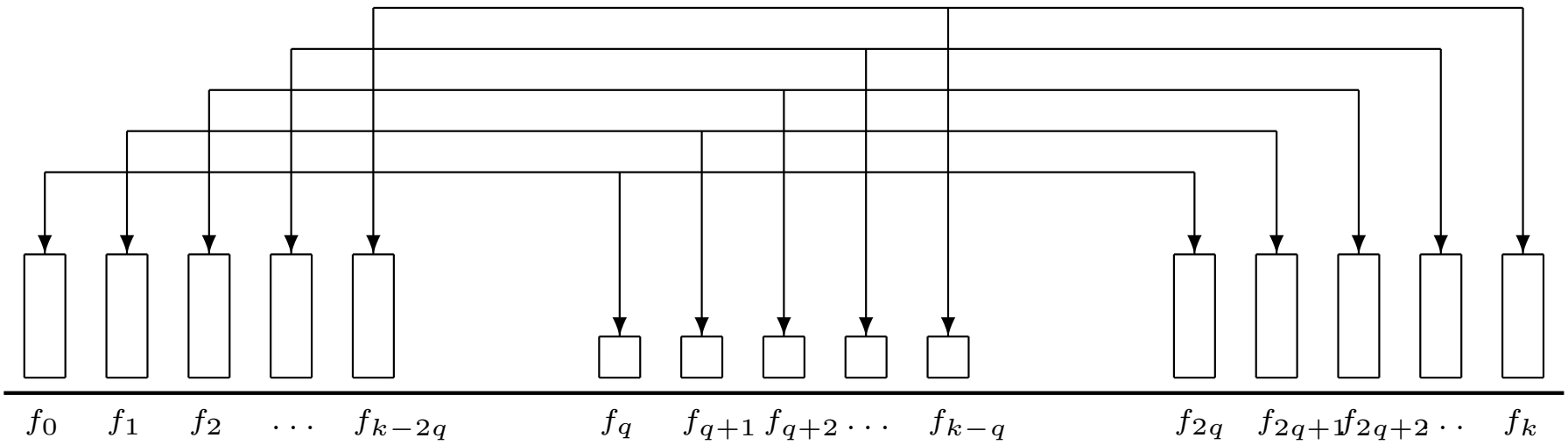
$$c_{2q} \equiv_q \sum_{i=0}^{2q} \binom{2q}{i} F(i+r) \equiv_q F(r) + 2 \cdot F(q+r) + F(2q+r)$$



Proof Sketch

- By **[BHP]** there is a prime number, $q > \frac{p}{2}$, such that $q = \frac{k-\Theta(k^{0.525})}{2}$
- Fixing $k - 2q$ bits and reducing modulo q yields:

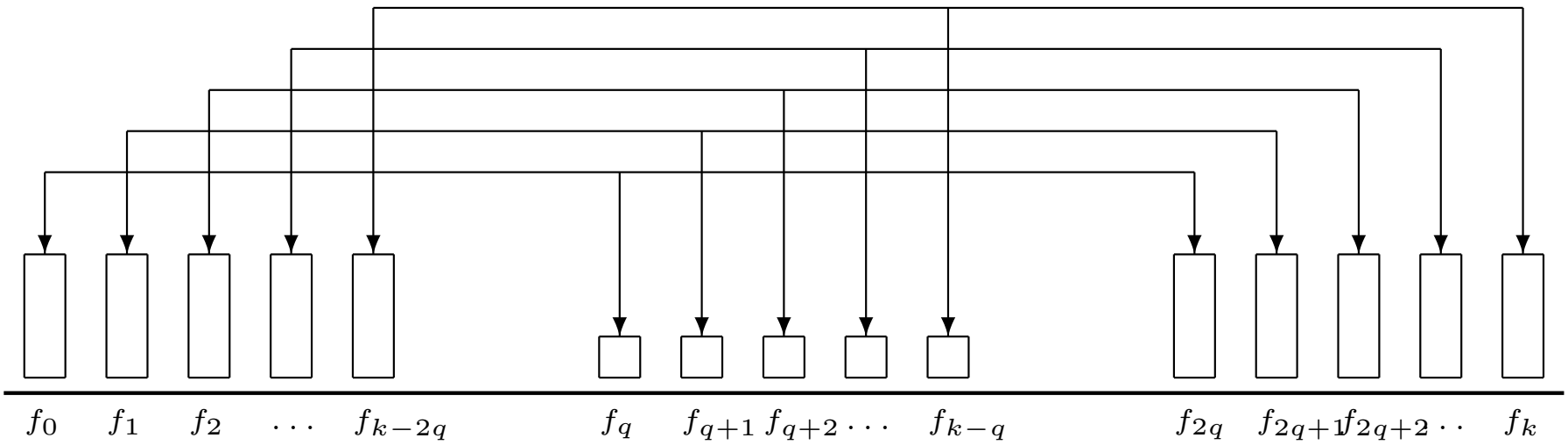
$$c_{2q} \equiv_q \sum_{i=0}^{2q} \binom{2q}{i} F(i+r) \equiv_q F(r) + 2 \cdot F(q+r) + F(2q+r)$$



Proof Sketch

- By **[BHP]** there is a prime number, $q > \frac{p}{2}$, such that $q = \frac{k - \Theta(k^{0.525})}{2}$
- Fixing $k - 2q$ bits and reducing modulo q yields:

$$c_{2q} \equiv_q \sum_{i=0}^{2q} \binom{2q}{i} F(i+r) \equiv_q F(r) + 2 \cdot F(q+r) + F(2q+r)$$



F is fixed on $[k/2 - \Theta(k^{0.525}), k/2 + \Theta(k^{0.525})]$
$\Rightarrow f$ is either constant or $bias(F) \neq \text{bias}(F|_{(2,1)})$. Q.E.D.

Summary

- Non-linear symmetric functions has $O(k^{0.525})$ size nonzero fourier coefficients.
- This shows that learning symmetric juntas using the Fourier based algorithm takes no more than $n^{O(k^{0.525})} \cdot \mathrm{poly}(n, 2^k)$

Summary

- Non-linear symmetric functions has $O(k^{0.525})$ size nonzero fourier coefficients.
- This shows that learning symmetric juntas using the Fourier based algorithm takes no more than $n^{O(k^{0.525})} \cdot \mathrm{poly}(n, 2^k)$

Thank You