THE STRUCTURE OF THE FOURIER SPECTRUM OF BOOLEAN FUNCTIONS, AND THEIR COMPLEXITY

Ben Lee Volk (Technion)

Based on joint work with Amir Shpilka and Avishay Tal

Main Theme:

Boolean functions with **simple** Fourier transform have **small** complexity.

There are several

1. ways to measure the complexity of the Fourier transform
2. relevant computational models

OUTLINE

- Boolean functions with small spectral norm
  - Circuit Complexity
  - Decision Trees
- Boolean functions with very few non-zero coefficients
  - Communication Complexity of XOR functions
  - Decision Trees

BOOLEAN FUNCTIONS

- Consider the vector space of functions:
$$\{ f \mid f : \mathbb{Z}_2^n \to \mathbb{R} \}.$$
- $\chi_\alpha(x) = (-1)^{\langle \alpha, x \rangle}$ for all $\alpha \in \mathbb{Z}_2^n$ is an orthonormal basis with respect to the inner product
$$\langle f, g \rangle = \mathbb{E}_x[f(x)g(x)]$$
- $f(x) = \sum_\alpha \hat{f}(\alpha)\chi_\alpha(x)$.
- We're interested in functions that only take the values $\{\pm 1\}$ (aka boolean functions).

SPECTRAL NORM OF BOOLEAN FUNCTIONS

The **spectral norm** (ℓ_1 norm) of $f: \mathbb{Z}_2^n \to \{-1,1\}$ is:

$$\|\hat{f}\|_1 = \sum_{\alpha} |\hat{f}(\alpha)|.$$

Parseval and Cauchy-Schwartz imply:
For every boolean function, $\|\hat{f}\|_1 \leq 2^{n/2}$.

For a random boolean function f, $\|\hat{f}\|_1 = 2^{\Omega(n)}$.

FUNCTIONS WITH SMALL SPECTRAL NORM

If $f: \mathbb{Z}_2^n \rightarrow \{-1,1\}$ is an indicator function of an affine subspace $V \subseteq \mathbb{Z}_2^n$, $\|\hat{f}\|_1 \leq 3$.

(Examples of such functions: AND, OR, XOR)

FUNCTIONS WITH SMALL SPECTRAL NORM

Theorem (**[Green-Sanders08]**): Suppose f is a boolean function with $\|\hat{f}\|_1 \leq M$. Then

$$f = \sum_{i=1}^{L} \pm \mathbf{1}_{V_i},$$

where $V_i \subseteq \mathbb{Z}_2^n$ are affine subspaces and $L \leq 2^{2^{O(M^4)}}$.

CIRCUIT COMPLEXITY OF FUNCTIONS WITH SMALL SPECTRAL NORM

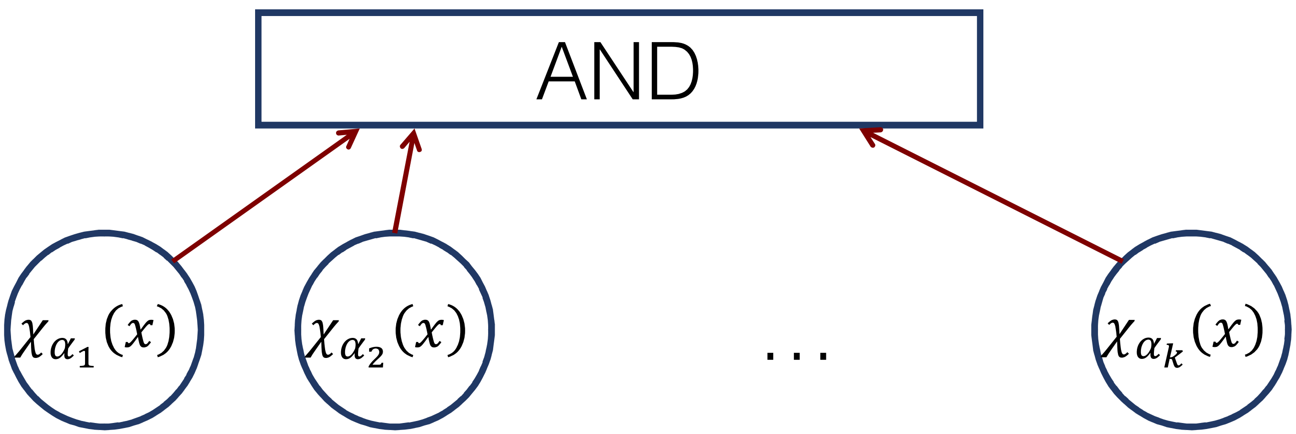
$\mathrm{AC}^0[2]$: Class of boolean functions computed by circuits with polynomial size, constant depth, and unbounded fan-in AND, OR, NOT and “MOD 2” gates.

An application of **[GS08]**: Functions with constant spectral norm are in $\mathrm{AC}^0[2]$.

CIRCUIT COMPLEXITY OF FUNCTIONS WITH SMALL SPECTRAL NORM

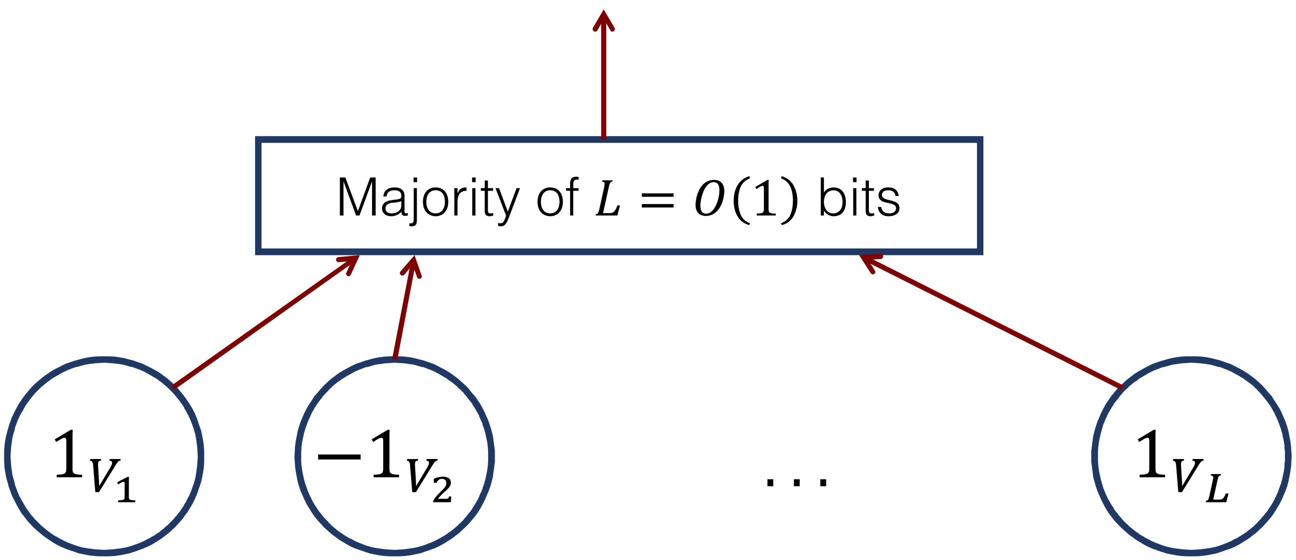
Proof:

Part #1: Every indicator of a subspace (AND of at most n parities or negation of parities) is in $\mathrm{AC}^0[2]$:



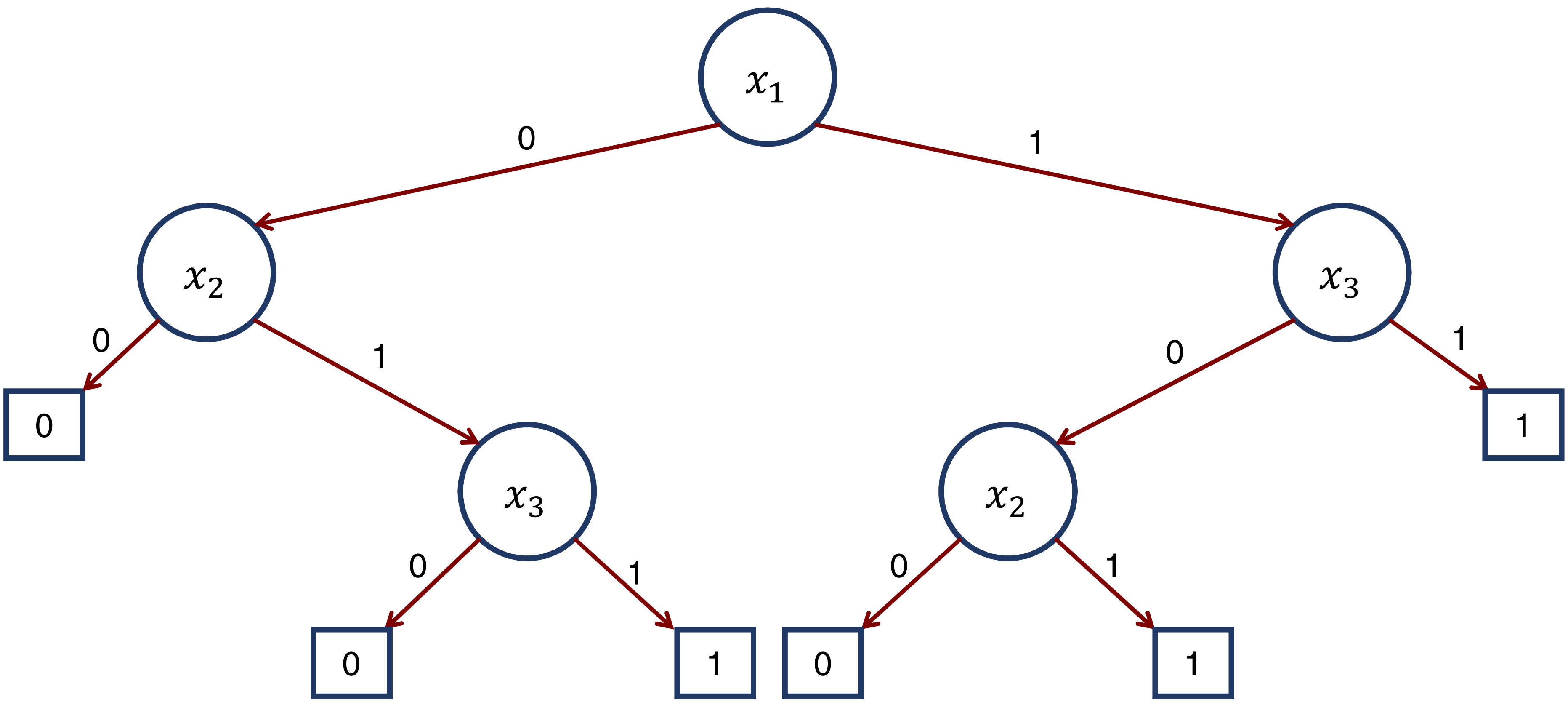
CIRCUIT COMPLEXITY OF FUNCTIONS WITH SMALL SPECTRAL NORM

Part #2:



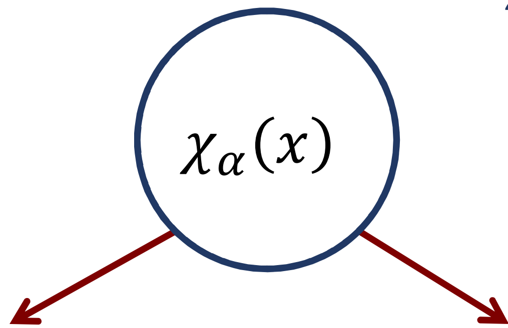
Number of gates: $2^{2^{O(M^4)}} \cdot \text{poly}(n)$. Depth = $O(1)$

DECISION TREES



PARITY DECISION TREES ($\bigoplus$-DT)

Same as decision tree, except that every internal node is labeled with a linear function over $\mathbb{Z}_2^n$:



$D^{\oplus}(f) :=$ minimal *depth* of a $\bigoplus$-DT for f
$\text{size}_{\oplus}(f) :=$ minimal *size* of a $\bigoplus$-DT for f
(minimal number of leaves).

PARITY DECISION TREES ($\oplus$-DT)

A function f computed by a parity decision tree of size s has $\|\hat{f}\|_1 \leq s$.

This inequality can be quite loose (e.g. $f = \mathrm{AND}$: $\|\hat{f}\|_1 \leq 3$, $\mathrm{size}_{\oplus}(f) = \Omega(n)$.

PARITY DECISION TREES ($\bigoplus$-DT)

Theorem: If f is a boolean function with $\|\hat{f}\|_1 \leq M$ then $\text{size}_{\bigoplus}(f) \leq n^{M^2}$.

Key Lemma: Can find a hyperplane such that the restriction of f to it has significantly smaller spectral norm.

KEY LEMMA

$\|\hat{f}\|_1 = M > 1$, $\hat{f}(\alpha), \hat{f}(\beta)$ two largest coefficients.

$f|_{\chi_{\alpha+\beta}=z}$:= restriction of f to $\{x \mid \chi_{\alpha+\beta}(x) = z\}$.

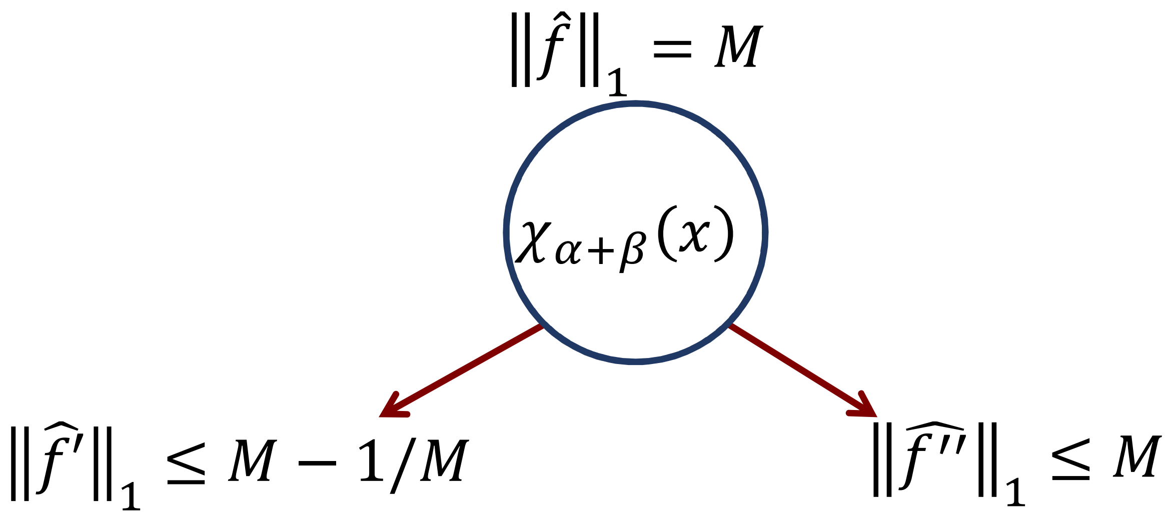
Then:

$$\left\|\widehat{f|_{\chi_{\alpha+\beta}=1}}\right\|_1 \leq M - |\hat{f}(\alpha)| \leq M - 1/M$$

$$\left\|\widehat{f|_{\chi_{\alpha+\beta}=-1}}\right\|_1 \leq M - |\hat{f}(\beta)|$$

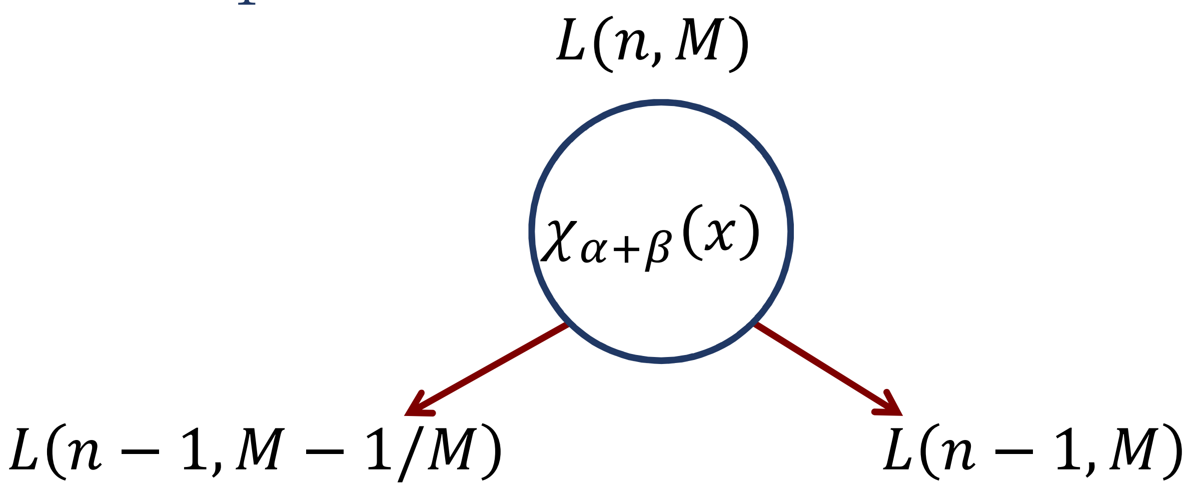
(*or the other way around)

KEY LEMMA



BACK TO PARITY DECISION TREES ($\oplus$-DT)

Set $L(n, M) = \max_{\|\hat{f}\|_1 \leq M} \mathrm{size}_{\oplus}(f)$. By Key Lemma:



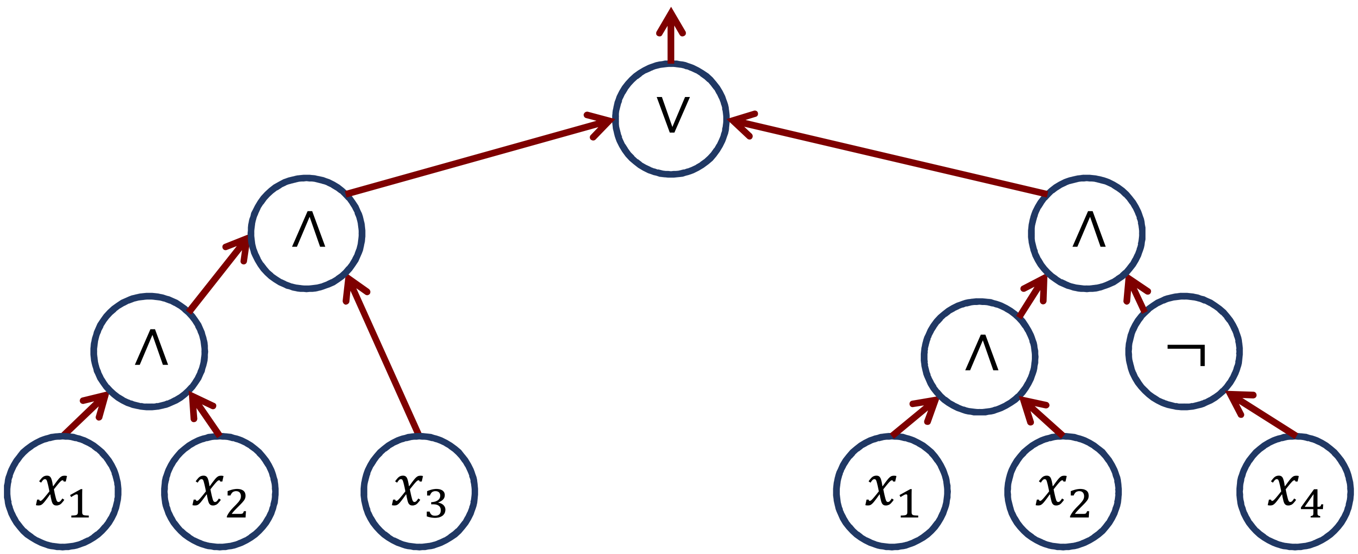
$\Rightarrow L(n, M) \leq L(n-1, M-1/M) + L(n-1, m)$

Remark: More careful analysis of Key Lemma gives $2^{M^2} n^M$.

FORMULAS

A **formula** is a **circuit** such that every gate has outdegree 1 (the underlying graph is a tree).

$$(x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge \neg x_4)$$



FORMULAS

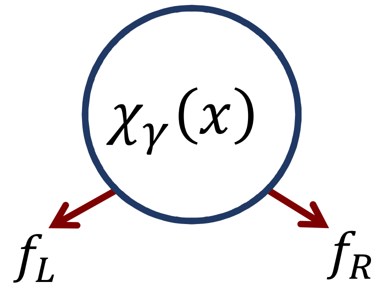
Let $L(f)$ be the size of a minimal De Morgan formula (gates allowed: fan-in 2 AND, OR, NOT) which computes f.

Example: $L(\mathrm{XOR}) = O(n^2)$.

FORMULAS

Observation: If $\mathrm{size}_{\oplus}(f) = s$ then $L(f) = O(s \cdot n^2)$.

Proof: Induction on s.



$L(\chi_\gamma), L(\neg\chi_\gamma) = O(n^2)$.

$$f = (\chi_\gamma \wedge f_L) \vee (\neg\chi_\gamma \wedge f_R)$$
$$\Rightarrow L(f) \leq L(f_L) + L(f_R) + O(n^2).$$

FORMULAS

Corollary: Functions with small spectral norm not only have small $\mathrm{AC}^0[2]$ **circuits** but also small **formulas** (of size $O\left(2^{M^2} n^M \cdot n^2\right)$).

Furthermore: formulas, unlike trees, can be **balanced**.

So f also has a formula of depth $O(M \log n + M^2)$.

SPARSITY OF BOOLEAN FUNCTIONS

The **sparsity** of $f: \mathbb{Z}_2^n \to \{-1,1\}$ is the number of its non-zero Fourier coefficients:

$$\|\hat{f}\|_0 = \#\{\, \alpha \mid \hat{f}(\alpha) \neq 0 \,\}.$$

For a random function f, $\|\hat{f}\|_0 = \big(1 - o(1)\big)2^n$.

SPARSE FUNCTIONS: EXAMPLES

If f is computed by a $\bigoplus$-DT of depth d and size s, then $\|\hat{f}\|_0 \leq s \cdot 2^d \leq 4^d$.

Example: "Address function."

Input: | $x_1 \cdots x_{\log n}$ | $y_1 y_2 \quad \cdots \quad y_{n-1} y_n$ |

Output: $y_{x_1 \cdots x_{\log n}}$.

Sparsity: n^2.

SPARSE FUNCTIONS

Conjecture ([Zhang-Shi10],[Montanaro-Osborne09]):
$\exists c > 0$ such that for every boolean function f,

$$D^{\oplus}(f) \leq \left(\log \|\hat{f}\|_0\right)^c.$$

COMMUNICATION COMPLEXITY

$F: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$

Alice has $x \in \{0,1\}^n$

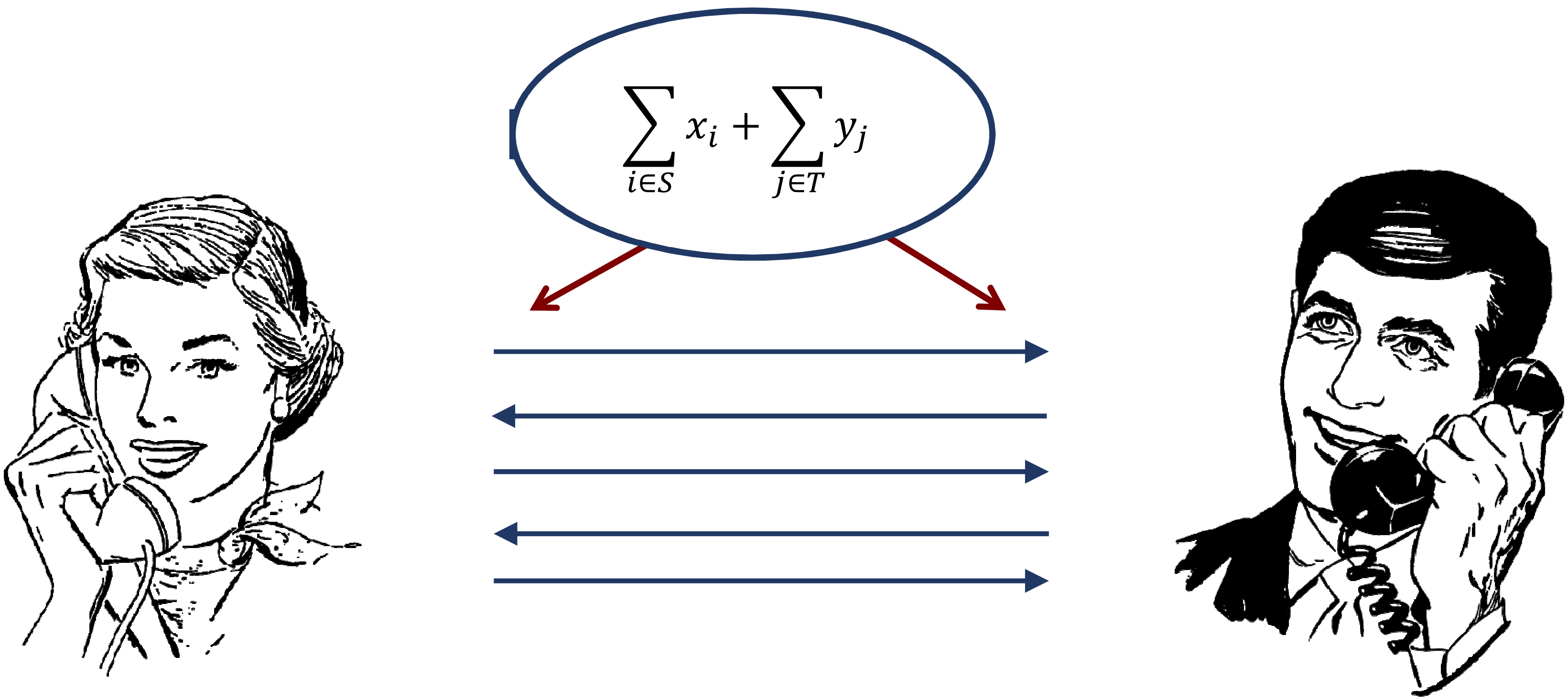
Bob has $y \in \{0,1\}^n$

Want to compute $F(x, y)$.

$\mathrm{cc}^{\mathrm{det}}(F) =$ minimal number of bits needed to communicate in order to compute F deterministically.

COMMUNICATION COMPLEXITY

Observation: A parity decision tree of depth d for $F \Rightarrow$ a protocol with at most $2d$ bits of communication.



COMMUNICATION COMPLEXITY: LOG-RANK CONJECTURE

Associate with every function F a real $2^n \times 2^n$ matrix M_F such that $M_F(x, y) = F(x, y)$.

Fact [Mehlhorn-Schmidt82]: $\mathrm{CC}^{\mathrm{det}}(F) \geq \log \mathrm{rank}(M_F)$.

Log-Rank Conjecture [Lovász-Saks88]: $\exists c$ such that $\mathrm{CC}^{\mathrm{det}}(F) \leq (\log \mathrm{rank}(M_F))^c$.

COMMUNICATION COMPLEXITY: SPARSITY

Suppose now $F(x, y) = f(x \oplus y)$, for $f: \mathbb{Z}_2^n \to \{-1,1\}$.
(Such functions are referred to as "**XOR functions.**")

The eigenvectors of M_F are the Fourier characters, and the eigenvalues are (up to normalization) the Fourier coefficients of f.

So $\mathrm{rank}(M_F) = \|\hat{f}\|_0$.

SPARSE FUNCTIONS AND $\bigoplus$-DTs

If follows that if

$$D^{\bigoplus}(f) = \text{poly}\log\|\hat{f}\|_0$$

Then the log-rank conjecture holds for XOR functions.

Best separation known:

a function f such that $D^{\bigoplus}(f) = \Omega\left(\log\|\hat{f}\|_0{}^{1.63\ldots}\right)$

[Nisan-Szegedy92, Nisan-Wigderson95, Kushilevitz94]

SPARSE FUNCTIONS: WHAT IT TAKES

When we look at f restricted to $\{\, x \mid \chi_\alpha(x) = \pm 1 \,\}$:

BEFORE	$\hat{f}(\beta_1)$	$\hat{f}(\beta_1 + \alpha)$	$\hat{f}(\beta_2)$	$\hat{f}(\beta_2 + \alpha)$	$\cdots$
AFTER	$\hat{f}(\beta_1) \pm \hat{f}(\beta_1 + \alpha)$		$\hat{f}(\beta_2) \pm \hat{f}(\beta_2 + \alpha)$		$\cdots$

We want to find α with many pairs $\hat{f}(\beta), \hat{f}(\beta + \alpha)$ in the support of $\hat{f}$.

SPARSE FUNCTIONS WITH SMALL SPECTRAL NORM

What is f has $\|\hat{f}\|_1 \leq M$ and $\|\hat{f}\|_0 = s$?

Theorem: $D^{\oplus}(f) \leq M^2 \log s$
(**[Tsang-Wong-Xie-Zhang13]:** $M \log s$).

SPARSE FUNCTIONS WITH SMALL SPECTRAL NORM

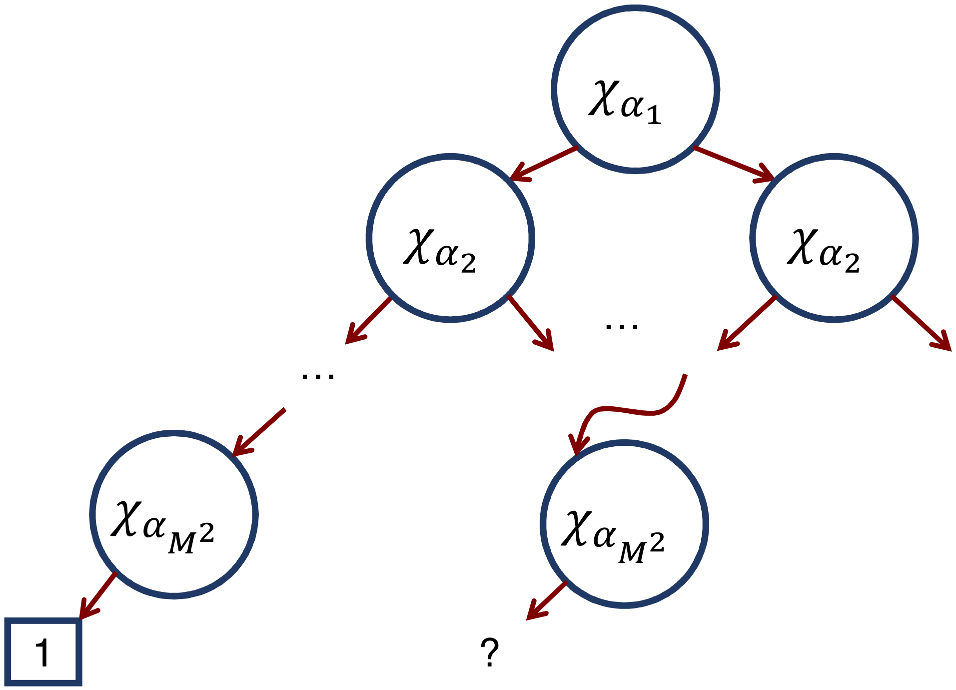
Proof:

Recall Key Lemma: Can find restriction with reduces the spectral norm by $M - 1/M$.

Apply Key Lemma M^2 times to obtain:

Theorem: For all f, $\exists$ affine subspace V of co-dimension $\leq M^2$ such that $f|_V$ is constant.

There exists M^2 linear functions $\chi_{\alpha_1}, \ldots, \chi_{\alpha_{M^2}}$ which can be fixed in a way which makes f constant. Consider the tree:



Because $f|_{\{\chi_{\alpha_i}=b_i\}}$ is constant, for any non-zero $\hat{f}(\beta)$ there is a non-zero $\hat{f}(\beta+\gamma)$ with $\gamma \in \mathrm{span}\{\alpha_i\}$.

Hence: $\hat{f}(\beta)$ and $\hat{f}(\beta+\gamma)$ collapse to the same coefficient under any settings of the χ_{α_i}'s:

$$\left\| \widehat{f\,|_{\{\chi_{\alpha_i}=b'_i\}}} \right\|_0 \leq \|\hat{f}\|_0 / 2.$$

Iterate at most $\log\|\hat{f}\|_0$ steps.

The same argument shows that in order to prove $D^{\oplus}(f) = \operatorname{poly}\log\|\hat{f}\|_0$, it's enough to prove:

Conjecture: For every boolean function f there is a subspace of co-dimension $\operatorname{poly}\log\|\hat{f}\|_0$ on which f is constant.

(since the reverse implication is immediate, this conjecture is in fact equivalent)

SUMMARY

Functions with small spectral norm have:

- Small circuits
- Small formulas
- Small $\oplus$-DTs
- (They also have small randomized **[Grolmusz97]** and deterministic **[Gavinsky-Lovett13]** communication complexity)

Sparse Functions:

- Open problem