Boolean Functions

Boolean function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$

$$\text{AND}_n(x) = \begin{cases} -1 \quad \text{(TRUE)} & \text{if } x = (-1)^n \\ 1 \quad \text{(FALSE)} & \text{otherwise} \end{cases}$$
Approximate Degree

- A real polynomial $p$ $\epsilon$-approximates $f$ if

  $$|p(x) - f(x)| < \epsilon \quad \forall x \in \{-1, 1\}^n$$

- $\widetilde{\deg}_\epsilon(f) = \text{minimum degree needed to } \epsilon\text{-approximate } f$

- $\widetilde{\deg}(f) := \deg_{1/3}(f)$ is the approximate degree of $f$
Why Care About Approximate Degree?

Upper bounds on $\tilde{\deg}_\epsilon(f)$ yield efficient learning algorithms:

- $\epsilon \to 1$: PAC learning [KS01]
- $\epsilon$ “close to” 1: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon < 1$ a constant: Agnostic Learning [KKMS05]
Why Care About Approximate Degree?

Lower bounds on $\widetilde{\deg}_\epsilon(f)$ yield lower bounds on:

- Quantum query complexity [BBCMW98] [AS01] [Amb03] [KSW04]
- Communication complexity [BVdW08] [She07] [SZ07] [CA08] [LS08] [She12]
- Circuit complexity [MP69] [Bei93] [Bei94] [She08]
Example: What is the Approximate Degree of $\text{AND}_n$?

\[ \widetilde{\deg}(\text{AND}_n) = \Theta(\sqrt{n}). \]

- **Upper bound:** Use **Chebyshev Polynomials**.
- **Markov’s Inequality:** Let $G(t)$ be a univariate polynomial s.t. $\deg(G) \leq d$ and $\sup_{t \in [-1,1]} |G(t)| \leq 1$. Then
  \[ \sup_{t \in [-1,1]} |G'(t)| \leq d^2. \]
- **Chebyshev polynomials are the extremal case.**
Example: What is the Approximate Degree of $\text{AND}_n$?

\[
\widetilde{\deg}(\text{AND}_n) = O(\sqrt{n}).
\]

- After shifting a scaling, can turn degree $O(\sqrt{n})$ Chebyshev polynomial into a univariate polynomial $Q(t)$ that looks like:

\[
Q(-1+2/n) = 2/3
\]

- Define $n$-variate polynomial $p$ via $p(x) = Q(\sum_{i=1}^{n} x_i/n)$.
- Then $|p(x) - \text{AND}_n(x)| \leq 1/3 \quad \forall x \in \{-1, 1\}^n$. 
Example: What is the Approximate Degree of $\text{AND}_n$?

\[ \text{[NS92]} \quad \tilde{\deg}(\text{AND}_n) = \Omega(\sqrt{n}). \]

- Lower bound: Use \textit{symmetrization}.
- Suppose \(|p(x) - \text{AND}_n(x)| \leq 1/3 \quad \forall x \in \{-1, 1\}^n\).
- There is a way to turn \(p\) into a \textit{univariate} polynomial \(p^{\text{sym}}\) that looks like this:

\[ Q(-1+2/n) \geq 2/3 \]

- Claim 1: \(\deg(p^{\text{sym}}) \leq \deg(p)\).
- Claim 2: Markov’s inequality \(\implies\) \(\deg(p^{\text{sym}}) = \Omega(n^{1/2})\).
Beyond Symmetrization: Analyzing the OR-AND Tree
Symmetrization is “lossy”: in turning an $n$-variate poly $p$ into a univariate poly $p^{\text{sym}}$, we throw away information about $p$.

Challenge problem: What is $\tilde{\deg}(\text{OR-AND}_n)$?
History of the OR-AND Tree

Upper bounds

[HMW03] \( \widetilde{\deg}(\text{OR-AND}_n) = O(n^{1/2}) \)

Lower bounds

[NS92] \( \Omega(n^{1/4}) \)
[Shi01] \( \Omega(n^{1/4} \sqrt{\log n}) \)
[Amb03] \( \Omega(n^{1/3}) \)
[Aar08] Reposed Question
[She09] \( \Omega(n^{3/8}) \)
[BT13] \( \Omega(n^{1/2}) \)
[She13] \( \Omega(n^{1/2}), \) independently
What is best error achievable by any degree $d$ approximation of $f$?

Primal LP (Linear in $\epsilon$ and coefficients of $p$):

$$\min_{p, \epsilon} \epsilon$$

s.t. \( |p(x) - f(x)| \leq \epsilon \) for all \( x \in \{-1, 1\}^n \)

\( \deg p \leq d \)

Dual LP:

$$\max_{\psi} \sum_{x \in \{-1, 1\}^n} \psi(x) f(x)$$

s.t. \( \sum_{x \in \{-1, 1\}^n} |\psi(x)| = 1 \)

\( \sum_{x \in \{-1, 1\}^n} \psi(x)q(x) = 0 \) whenever \( \deg q \leq d \)
Dual Characterization of Approximate Degree

**Theorem:** \( \deg_\epsilon(f) > d \) iff there exists a “dual polynomial” \( \psi \) with

1. \[ \sum_{x \in \{-1, 1\}^n} \psi(x)f(x) > \epsilon \] “high correlation with \( f \)”
2. \[ \sum_{x \in \{-1, 1\}^n} |\psi(x)| = 1 \] “\( L_1 \)-norm 1”
3. \[ \sum_{x \in \{-1, 1\}^n} \psi(x)q(x) = 0, \deg q \leq d \] “pure high degree \( d \)”

Key technique in, e.g., [She07] [Lee09] [She09]

A **lossless** technique. Strong duality implies any approximate degree lower bound can be witnessed by dual polynomial.
Goal: Construct an explicit dual polynomial $\psi_{\text{OR-AND}}$ for OR-AND
Constructing a Dual Polynomial

- By [NS92], there are dual polynomials
  \( \psi_{\text{OUT}} \) for \( \widetilde{\deg}(\text{OR}_{n^{1/2}}) = \Omega(n^{1/4}) \) and
  \( \psi_{\text{IN}} \) for \( \widetilde{\deg}(\text{AND}_{n^{1/2}}) = \Omega(n^{1/4}) \)

- Can we combine \( \psi_{\text{OUT}} \) and \( \psi_{\text{IN}} \) to obtain a dual polynomial \( \psi_{\text{OR-AND}} \) for OR-AND?
A First Attempt

\[
\psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := \psi_{\text{OUT}}(\ldots, \psi_{\text{IN}}(x_i), \ldots)
\]

Easy to check: \(\psi_{\text{OR-AND}}\) has pure high degree at least \(\frac{n^{1/2}}{4} \cdot \frac{n^{1/2}}{4} = \frac{n^{1/2}}{2}\).

E.g. If \(\psi_{\text{OUT}}(y_1, y_2) = y_1 \cdot y_2\) and \(\psi_{\text{IN}}(z_1, z_2) = z_1 \cdot z_2\), then \(\psi_{\text{OR-AND}}(x_{11}, x_{12}, x_{21}, x_{22}) = (x_{11} \cdot x_{12}) \cdot (x_{21} \cdot x_{22}) = x_{11} \cdot x_{12} \cdot x_{21} \cdot x_{22}\).

Does \(\psi_{\text{OR-AND}}\) have high correlation with \(\text{OR-AND}\)?

Problem: Proposed definition of \(\psi_{\text{OR-AND}}\) may feed non-Boolean values into \(\psi_{\text{OUT}}\). But we only have control over \(\psi_{\text{OUT}}\) on Boolean inputs.
A First Attempt

\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := \psi_{\text{OUT}}(\ldots, \psi_{\text{IN}}(x_i), \ldots) \]

- Easy to check: \( \psi_{\text{OR-AND}} \) has pure high degree at least \( n^{1/4} \cdot n^{1/4} = n^{1/2} \).
- E.g. If \( \psi_{\text{OUT}}(y_1, y_2) = y_1y_2 \) and \( \psi_{\text{IN}}(z_1, z_2) = z_1z_2 \), then
  \[ \psi_{\text{OR-AND}}(x_{11}, x_{12}, x_{21}, x_{22}) = (x_{11}x_{12})(x_{21}x_{22}) = x_{11}x_{12}x_{21}x_{22}. \]
A First Attempt

\[ \psi\text{OR-AND}(x_1, \ldots, x_{n^{1/2}}) := \psi\text{OUT}(\ldots, \psi\text{IN}(x_i), \ldots) \]

- Easy to check: \( \psi\text{OR-AND} \) has pure high degree at least \( n^{1/4} \cdot n^{1/4} = n^{1/2} \).
- E.g. If \( \psi\text{OUT}(y_1, y_2) = y_1y_2 \) and \( \psi\text{IN}(z_1, z_2) = z_1z_2 \), then
  \[ \psi\text{OR-AND}(x_{11}, x_{12}, x_{21}, x_{22}) = (x_{11}x_{12})(x_{21}x_{22}) = x_{11}x_{12}x_{21}x_{22}. \]
- Does \( \psi\text{OR-AND} \) have high correlation with \( \text{OR-AND}_n \)?
- Problem: Proposed definition of \( \psi\text{OR-AND} \) may feed non-Boolean values into \( \psi\text{OUT} \). But we only have control over \( \psi\text{OUT} \) on Boolean inputs.
A Second (and Final) Attempt [She09, Lee09]

\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

(\(C\) chosen to ensure \(\psi_{\text{OR-AND}}\) has \(L_1\)-norm 1).
A Second (and Final) Attempt [She09, Lee09]

\[
\psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)|
\]

\((C \text{ chosen to ensure } \psi_{\text{OR-AND}} \text{ has } L_1\text{-norm 1}).\)

Must verify:

1. \(\psi_{\text{OR-AND}}\) has pure high degree \(\geq n^{1/4} \cdot n^{1/4} = n^{1/2}.\)
2. \(\psi_{\text{OR-AND}}\) has high correlation with OR-AND.
\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

\((C \text{ chosen to ensure } \psi_{\text{OR-AND}} \text{ has } L_1\text{-norm 1}).\)

Must verify:

1. \(\psi_{\text{OR-AND}}\) has pure high degree \(\geq n^{1/4} \cdot n^{1/4} = n^{1/2}. \checkmark [\text{She09}]\)

2. \(\psi_{\text{OR-AND}}\) has high correlation with OR-AND. [BT13]
(Sub)Goal: Show $\psi_{\text{OR-AND}}$ has pure high degree at least $n^{1/2}$ [She09]
\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

- Intuition: Consider \( \psi_{\text{OUT}}(y_1, y_2, y_3) = y_1 y_2 \). Then
  \( \psi_{\text{OR-AND}}(x_1, x_2, x_3) \) equals:

  \[ C \cdot \text{sgn}(\psi_{\text{IN}}(x_1)) \cdot \text{sgn}(\psi_{\text{IN}}(x_2)) \cdot \prod_{i=1}^{3} |\psi_{\text{IN}}(x_i)| = \psi_{\text{IN}}(x_1) \cdot \psi_{\text{IN}}(x_2) \cdot |\psi_{\text{IN}}(x_3)| \]
Pure High Degree Analysis [She09]

\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

- **Intuition:** Consider \( \psi_{\text{OUT}}(y_1, y_2, y_3) = y_1 y_2 \). Then
  \( \psi_{\text{OR-AND}}(x_1, x_2, x_3) \) equals:

  \[
  C \cdot \text{sgn}(\psi_{\text{IN}}(x_1)) \cdot \text{sgn}(\psi_{\text{IN}}(x_2)) \cdot \prod_{i=1}^{3} |\psi_{\text{IN}}(x_i)|
  \]

  \[
  = \psi_{\text{IN}}(x_1) \cdot \psi_{\text{IN}}(x_2) \cdot |\psi_{\text{IN}}(x_3)|
  \]

- Each term of \( \psi_{\text{OR-AND}} \) is the product of \( \text{PHD}(\psi_{\text{OUT}}) \) polynomials over disjoint variable sets, each of pure high degree at least \( \text{PHD}(\psi_{\text{IN}}) \).

- So \( \psi_{\text{OR-AND}} \) has pure high degree at least \( \text{PHD}(\psi_{\text{OUT}}) \cdot \text{PHD}(\psi_{\text{IN}}) \).
(Sub)Goal: Show $\psi_{\text{OR-AND}}$ has high correlation with OR-AND
Correlation Analysis

\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

- Idea: Show

\[ \sum_{x \in \{-1,1\}^n} \psi_{\text{OR-AND}}(x) \cdot \text{OR-AND}_{n}(x) \approx \sum_{y \in \{-1,1\}^{n^{1/2}}} \psi_{\text{OUT}}(y) \cdot \text{OR}_{n^{1/2}}(y). \]

- Intuition: We are feeding \( \text{sgn}(\psi_{\text{IN}}(x_i)) \) into \( \psi_{\text{OUT}}. \)

- \( \psi_{\text{IN}} \) is **correlated** with \( \text{AND}_{n^{1/2}} \), so \( \text{sgn}(\psi_{\text{IN}}(x_i)) \) is a “decent predictor” of \( \text{AND}_{n^{1/2}}. \)

- But there are errors. Need to show errors don’t “build up”.
Correlation Analysis

\[ \psi_{\text{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\text{IN}}(x_i)| \]

- **Goal:** Show

\[ \sum_{x \in \{-1,1\}^n} \psi_{\text{OR-AND}}(x) \cdot \text{OR-AND}_n(x) \approx \sum_{y \in \{-1,1\}^{n^{1/2}}} \psi_{\text{OUT}}(y) \cdot \text{OR}_{n^{1/2}}(y). \]

- **Case 1:** Consider any \( y = (\text{sgn} \psi_{\text{IN}}(x_1), \ldots, \text{sgn} \psi_{\text{IN}}(x_{n^{1/2}})) \neq \text{All-False}. \)

- There is some coordinate of \( y \) that equals TRUE. Only need to “trust” this coordinate to force \( \text{OR-AND}_n \) to evaluate to True on \( (x_1, \ldots, x_{n^{1/2}}) \). So errors do not build up!
Case 2: Consider $y = \text{All-False}$.

$\text{OR}_{n^{1/2}}(y) = \text{OR-AND}_n(x_1, \ldots, x_{n^{1/2}})$ only if all coordinates of $y$ are “error-free”.

Fortunately, $\psi_{\text{IN}}$ has a special one-sided error property: If $\text{sgn}(\psi_{\text{IN}}(x_i)) = 1$, then $\text{AND}_{n^{1/2}}(x_i)$ is guaranteed to equal 1.
Two Cases.

In first case (feeding at least one TRUE into $\psi_{\text{OUT}}$), errors did not build up, because we only needed to “trust” the TRUE value.

In second case (all values fed into $\psi_{\text{OUT}}$ are FALSE), we needed to trust all values. But we could do this because $\psi_{\text{IN}}$ had one-sided error.
A real polynomial $p$ is a \textbf{negative one-sided $\epsilon$-approximation} for $f$ if

\begin{align*}
|p(x) - 1| &< \epsilon \quad \forall x \in f^{-1}(1) \\
p(x) &\leq -1 \quad \forall x \in f^{-1}(-1)
\end{align*}

\textbf{$\tilde{\text{odeg}}_{-,\epsilon}(f)$} = \text{min degree of a one-sided $\epsilon$-approximation for $f$.}
Dual Formulation of $\tilde{\text{odeg}}_{-,\epsilon}$

Primal LP (Linear in $\epsilon$ and coefficients of $p$):

$$\min_{p, \epsilon} \quad \epsilon$$

s.t.  \quad \begin{align*}
|p(x) - 1| &\leq \epsilon & \text{for all } x \in f^{-1}(1) \\
p(x) &\leq -1 & \text{for all } x \in f^{-1}(-1) \\
\deg p &\leq d
\end{align*}$$

Dual LP:

$$\max_\psi \quad \sum_{x \in \{-1, 1\}^n} \psi(x) f(x)$$

s.t.  \quad \begin{align*}
\sum_{x \in \{-1, 1\}^n} |\psi(x)| &= 1 \\
\sum_{x \in \{-1, 1\}^n} \psi(x) q(x) &= 0 & \text{whenever } \deg q \leq d \\
\psi(x) &\leq 0 & \forall x \in f^{-1}(-1) \\
\psi(x) &\leq 0 & \forall x \in f^{-1}(-1)
\end{align*}$$
Proof that $\overline{\text{odeg}}_{-1/3}(\text{AND}_n) = \Omega(\sqrt{n})$

We argued that the symmetrization of any $1/3$-approximation to $\text{AND}_n$ had to look like this:

$$Q(-1+2/n) \geq 2/3$$
Hardness Amplification for Constant-Depth Circuits [BT14]
Main Theorem

- Given: A “simple” Boolean function $f$ that is “hard to approximate to low error” by degree $d$ polynomials.
- Can we turn $f$ into a “still-simple” $F$ that is hard to approximate even to very high error?
Main Theorem

- Given: A “simple” Boolean function $f$ that is “hard to approximate to low error” by degree $d$ polynomials.
- Can we turn $f$ into a “still-simple” $F$ that is hard to approximate even to very high error?

A: Yes.

Theorem

Let $f$ be a Boolean function with $\widehat{\text{odeg}}_{-,1/2}(f) \geq d$. Let $F = \text{OR}_t(f, \ldots, f)$. Then $\widehat{\text{odeg}}_{-,1-2^{-t}}(F) \geq d$. 
Proof of Main Theorem

- Define $\psi_{\text{IN}}$ to be any dual witness to the fact that $\overline{\deg}_{-,1/3}(f) \geq d$.
- Define $\psi_{\text{OUT}} : \{-1, 1\}^t \to \mathbb{R}$ via:
  \[
  \psi_{\text{OUT}}(y) = \begin{cases} 
  1/2 & \text{if } y = \text{ALL-FALSE} \\
  -1/2 & \text{if } y = \text{ALL-TRUE} \\
  0 & \text{otherwise}
  \end{cases}
  \]
- Combine $\psi_{\text{OUT}}$ and $\psi_{\text{IN}}$ exactly as before to obtain a dual witness $\psi_F$ for $F$.

Must verify:
1. $\psi_F$ has pure high degree $d$.
2. $\psi_F$ has correlation at least $1 - 2^{-t}$ with $F$. 
Proof of Main Theorem: Pure High Degree

- Notice $\psi_{\text{OUT}}$ is balanced (i.e., it has pure high degree 1).
- So previous analysis shows $\psi_F$ has pure high degree at least $1 \cdot d = d$. 
Proof of Main Theorem: Correlation Analysis

\( \psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\text{IN}}(x_i)| \)

- **Idea:** Show

\[
\sum_{x \in \{-1,1\}^n} \psi_F(x) \cdot F(x) \geq \sum_{y \in \{-1,1\}^t} \psi_{\text{OUT}}(y) \cdot \text{OR}_t(y) - 2^{-t} = 1 - 2^{-t}.
\]

- **Case 1:** Consider \( y = (\text{sgn} \psi_{\text{IN}}(x_1), \ldots, \text{sgn} \psi_{\text{IN}}(x_t)) = \text{All-True} \).

  - If even a single coordinate \( y_i \) of \( y \) is “error-free”, then
    \( F(x) = \text{OR}_t(f(x_1), \ldots, f(x_t)) = -1 \). :-D

  - Any individual coordinate of \( y \) is in error with probability at most \( 1/2 \), since \( \psi_{\text{IN}} \) is well-correlated with \( f \).

  - So **all** coordinates of \( y \) are in error with probability only \( 2^{-t} \).
Proof of Main Theorem: Correlation Analysis

\[ \psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\text{IN}}(x_i)| \]

- Idea: Show

\[ \sum_{x \in \{-1,1\}^n} \psi_F(x) \cdot F(x) \geq \sum_{y \in \{-1,1\}^t} \psi_{\text{OUT}}(y) \cdot \text{OR}_t(y) - 2^{-t} = 1 - 2^{-t}. \]

- Case 2: Consider \( y = (\text{sgn} \psi_{\text{IN}}(x_1), \ldots, \text{sgn} \psi_{\text{IN}}(x_t)) = \text{All-False} \). Then \( \text{sgn}(\psi_F(x)) = \text{sgn}(\psi_{\text{OUT}}(y)) = 1 \).

- Then \( F(y) = \text{OR}_t(f(x_1), \ldots, f(x_t)) = 1 \) only if all coordinates of \( y \) are “error-free”.

- Fortunately, \( \psi_{\text{IN}} \) has one-sided error: If \( \text{sgn}(\psi_{\text{IN}}(x_i)) = 1 \), then \( f(x_i) \) is guaranteed to equal 1.
We want to apply amplification to functions in AC$^0$, getting out very “hard” functions that are still in AC$^0$.

Let $ED : \{-1, 1\}^n \rightarrow \{-1, 1\}$ denote the Element Distinctness function.

[AS04] showed $\tilde{\deg}(ED) = \Omega(N^{2/3})$.

This is the best known lower bound on the approximate degree of an AC$^0$ function.

We show that in fact $\tilde{\odeg}_{-1/3}(ED) = \Omega(N^{2/3})$. 
New Lower Bounds for AC$^0$

**Theorem**

Let $F = \text{OR}_{n^{2/5}}(ED_{n^{3/5}}, \ldots, ED_{n^{3/5}})$ and $\epsilon = 1 - 2^{-n^{2/5}}$. Then $
 \tilde{\text{odeg}}_{-,\epsilon}(F) = \tilde{\Omega}(n^{2/5}).$

**Proof:** Combine lower bound on $\tilde{\text{odeg}}_{-, 1/3}(ED)$ with Main Theorem.
New Lower Bounds for $\text{AC}^0$

Definition

Let $f : X \times Y \to \{-1, 1\}$ be a function, and $\mu$ a probability distribution on $X \times Y$. The discrepancy of $f$ under $\mu$ is

$$\text{disc}_\mu(f) := \max_{S \subseteq X, T \subseteq Y} \left| \sum_{x \in S} \sum_{y \in T} \mu(x, y) f(x, y) \right|.$$ 

The discrepancy of $f$ is: $\text{disc}(f) := \min_\mu \text{disc}_\mu(f)$.

- Low discrepancy implies high communication complexity in nearly every communication model.
- Also a central quantity in learning theory and circuit complexity.
**New Lower Bounds for AC⁰**

**Theorem (She08, “Pattern Matrix Method”)**

Let \( F : \{-1, 1\}^n \) be any function satisfying \( \widetilde{\deg}_{1-1/W}(F') \geq d \). Let \( F' : \{-1, 1\}^{4n} \times \{-1, 1\}^{4n} \to \{-1, 1\} \) by

\[
F'(x, y) = F(\ldots, \bigvee_{j=1}^{4} (x_{i,j} \land y_{i,j}), \ldots).
\]

Then \( \text{disc}(F') \lesssim \max\{1/W, 2^{-d}\} \).

**Corollary**

There is an AC⁰ function \( f \) (computed by a depth four circuit) with discrepancy \( \exp\left(-\Omega(n^{2/5})\right) \).

Proof: Apply Pattern Mat. Meth. to \( \text{OR}_{n^{2/5}}(\text{ED}_{n^{3/5}}, \ldots, \text{ED}_{n^{3/5}}) \).

Previous best bound: \( \exp\left(-\Omega(n^{1/3})\right) \) [She08, BVdW08].
Corollary

There is an $AC^0$ function $f$ that cannot be computed by $MAJ \circ THR$ circuits of size $\exp(\Omega(n^{2/5}))$.

Corollary

There is an $AC^0$ function $f$ with threshold weight $\exp(\Omega(n^{2/5}))$. 
Subsequent Work by Sherstov [She14]
Definition

Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be a Boolean function. A polynomial $p$ sign-represents $f$ if $\text{sgn}(p(x)) = f(x)$ for all $x \in \{-1, 1\}^n$.

Definition

The threshold degree of $f$ is $\min \deg(p)$, where the minimum is over all sign-representations of $f$. 
Threshold Degree of $\text{AC}^0$

- Minsky and Papert [MP69] proved an $\Omega(n^{1/3})$ lower bound on the threshold degree of a specific DNF.
- It had been open ever since to prove a lower bound of $\Omega(n^{1/3+\delta})$ for any function in $\text{AC}^0$.
- Only progress: $\Omega(n^{1/3} \log^k n)$ for any constant $k$ [OS03].
- We conjectured in [BT14] that $\text{OR}_{n^{2/5}}(\text{ED}_{n^{3/5}}, \ldots, \text{ED}_{n^{3/5}})$ has threshold degree $\Omega(n^{2/5})$. 
Subsequent Work

- Sherstov [She14] proved our conjecture.

**Theorem (She14)**

Let $f$ be a Boolean function with $\overline{\text{odeg}}_{-,1/2}(f) \geq d$. Let $F = \text{OR}_t(f, \ldots, f)$. Then $\text{deg}_\pm(F) = \Omega(\min\{d, t\})$.

- More generally, he exhibits a depth $k$ circuit of polynomial size with threshold degree $\Omega(n^{(k-1)/(2k-1)})$. 
Additional Intuition and Results [Tha14]
Sherstov [She13] showed that for any polynomial \( p : \{-1, 1\}^t \rightarrow [-1, 1] \), and every \( \delta > 0 \), there is a polynomial \( p_{\text{robust}} : \mathbb{R}^t \rightarrow \mathbb{R} \) of degree \( O(\deg(p) + \log(1/\delta)) \) such that

\[
|p(y) - p_{\text{robust}}(y + e)| < \delta
\]

for all \( y \in \{-1, 1\}^t \) and \( e \in [1/3, 1/3]^t \).
Sherstov [She13] showed that for any polynomial 
\[ p : \{-1, 1\}^t \rightarrow [-1, 1], \] and every \( \delta > 0 \), there is a polynomial 
\[ p_{\text{robust}} : \mathbb{R}^t \rightarrow \mathbb{R} \] of degree \( O(\deg(p) + \log(1/\delta)) \) such that 
\[ |p(y) - p_{\text{robust}}(y + e)| < \delta \]
for all \( y \in \{-1, 1\}^t \) and \( e \in [1/3, 1/3]^t \).

Given functions \( g, f \), can construct an \( (\epsilon + \delta) \)-approximating polynomial for \( g(f, \ldots, f) \) via:
\[ p^* := p_{\text{robust}}(q, \ldots, q), \]
where \( p \) is an \( \epsilon \)-approximation for \( g \), and \( q \) is a \( (1/3) \)-approximation for \( f \).
Sherstov [She13] showed that for any polynomial $p : \{-1, 1\}^t \to [-1, 1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^t \to \mathbb{R}$ of degree $O(\deg(p) + \log(1/\delta))$ such that

$$|p(y) - p_{\text{robust}}(y + e)| < \delta$$

for all $y \in \{-1, 1\}^t$ and $e \in [1/3, 1/3]^t$.

Given functions $g, f$, can construct an $(\epsilon + \delta)$-approximating polynomial for $g(f, \ldots, f)$ via:

$$p^* := p_{\text{robust}}(q, \ldots, q),$$

where $p$ is an $\epsilon$-approximation for $g$, and $q$ is a $(1/3)$-approximation for $f$.

This technique is optimal for all block-composed functions discussed so far.
Robustification and Complementary Slackness

Throughout, we used a single “combining” technique to construct dual polynomials for block-composed functions. Namely, the dual polynomial for $F = g(f, \ldots, f)$ was:

$$
\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\text{IN}}(x_i)|.
$$

Handwavy Claim: $\psi_F$ is designed to witness the optimality of Robustification.

Specifically, $\psi_F$ “almost” obeys complementary slackness with respect to $p^* = p_{\text{robust}}(q, \ldots, q)$. 
Suppose $p_{\text{robust}}$ achieved exactly optimal error $\epsilon$ among all degree $d$ approximations to the outer function $g$.

Then $p_{\text{robust}}$ yields an optimal solution to the linear program capturing $\deg_{\epsilon}(g)$.

Complementary slackness guarantees a dual polynomial $\psi_{\text{OUT}}$ that places non-zero weight only on primal constraints that are made tight by $p_{\text{robust}}$.

i.e. $\psi_{\text{OUT}}(y) \neq 0$ only for “maximal error points” $y \in \{-1, 1\}^t$ satisfying $|p_{\text{robust}}(y) - g(y)| = \epsilon$. 
Robustification Is Optimal (Except When It’s Not)

Recall: $\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\text{OUT}}(\ldots, \text{sgn}(\psi_{\text{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\text{IN}}(x_i)|$

- If $\psi_{\text{IN}}$ were perfectly correlated with $f$, then $\psi_F(x_1, \ldots, x_t) \neq 0$ only for $(x_1, \ldots, x_t)$ such that
  $$|p_{\text{robust}}(q(x_1), \ldots, q(x_t)) - g(f(x_1), \ldots, f(x_t))| = \epsilon \pm \delta.$$

- That is, $\psi_F$ places non-zero weight only on points on which $p_{\text{robust}}(q, \ldots, q)$ has “nearly maximal error” of at least $\epsilon - \delta$. 
Q: Is Robustification Always Optimal?

- A: No.
- Define $\text{OMB}_t: \{-1, 1\}^t \rightarrow \{-1, 1\}$ via:

$$\text{OMB}_t(x_1, \ldots, x_t) = (-1)^{i^*-1},$$

where $i^*$ is the largest index such that $x_{i^*} = -1$. 

Consider the function $F = \text{OMB}_t(\text{OR}N, \ldots, \text{OR}N)$. Optimal $\epsilon$-approximating polynomial for $F$ is of the form $p^* = p(q, \ldots, q)$, where $p$ is a non-robust $\epsilon$-approximating polynomial for $\text{OMB}_t$ and $q$ is an approximation for $\text{OR}N$ (cf. [KS04, Beigel94]).
Q: Is Robustification Always Optimal?

- A: No.
- Define $\text{OMB}_t: \{-1,1\}^t \rightarrow \{-1,1\}$ via:
  \[
  \text{OMB}_t(x_1, \ldots, x_t) = (-1)^{i^*-1},
  \]
  where $i^*$ is the largest index such that $x_{i^*} = -1$.
- Consider the function $F = \text{OMB}_t(\text{OR}_N, \ldots, \text{OR}_N)$. 
Q: Is Robustification Always Optimal?

- A: No.
- Define $\text{OMB}_t : \{-1, 1\}^t \rightarrow \{-1, 1\}$ via:
  \[ \text{OMB}_t(x_1, \ldots, x_t) = (-1)^{i^* - 1}, \]
  where $i^*$ is the largest index such that $x_{i^*} = -1$.
- Consider the function $F = \text{OMB}_t(\text{OR}_N, \ldots, \text{OR}_N)$.
- Optimal $\epsilon$-approximating polynomial for $F$ is of the form $p^* = p(q, \ldots, q)$, where $p$ is a non-robust $\epsilon$-approximating polynomial for $\text{OMB}_t$ and $q$ is an approximation for $\text{OR}_N$ (cf. [KS04, Beigel94]).
A Final Result

Theorem

Let $f$ be a Boolean function with $\tilde{\text{odeg}}_{+,1/2}(f) \geq d$. Let $F = \text{OMB}_t(f, \ldots, f)$. Then $\tilde{\text{odeg}}_{+,1-2^{-t}}(F) \geq d$.
A Final Result

<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let $f$ be a Boolean function with $\widetilde{\text{odeg}}_{+,1/2}(f) \geq d$. Let $F = \text{OMB}<em>t(f, \ldots, f)$. Then $\widetilde{\text{odeg}}</em>{+,1-2^{-t}}(F) \geq d$.</td>
</tr>
</tbody>
</table>

Compare to our earlier theorem:

<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let $f$ be a Boolean function with $\widetilde{\text{odeg}}_{-,1/2}(f) \geq d$. Let $F = \text{OR}<em>t(f, \ldots, f)$. Then $\widetilde{\text{odeg}}</em>{-,1-2^{-t}}(F) \geq d$.</td>
</tr>
</tbody>
</table>

- Our new result applies to functions $F$ with low threshold degree (e.g., $\text{OMB}_t(\text{OR}, \ldots, \text{OR})$ has threshold degree 1).
- The earlier theorem only applied to functions with threshold degree at least $t$ [She14].
Corollary: we exhibit an \( AC^0 \) function \( F' \) with dimension complexity \( n^{O(\log n)} \) and discrepancy \( \exp(-\tilde{\Omega}(n^{2/5})) \).

Previous best separation for \( AC^0 \) was due to [BVdW08], who gave an \( AC^0 \) function with dimension-complexity \( \text{poly}(n) \) and discrepancy \( \exp(-\Omega(n^{1/3})) \).

Key technical ingredient for our result:

**Lemma**

Let \( F = \text{OMB}_{n^{2/5}}(\overline{\text{ED}}_{n^{3/5}}, \ldots, \overline{\text{ED}}_{n^{3/5}}) \) and \( \epsilon = 1 - 2^{-\tilde{\Omega}(n^{2/5})} \).

Then \( \deg_\pm(F) = O(\log n) \), while \( \deg_\epsilon(F) = \tilde{\Omega}(n^{2/5}) \).
The dual witness we construct for $F = \text{OMB}(f, \ldots, f)$ is:

$$
\psi_F(x_1, \ldots, x_t) = \sum_{i=1}^{t} \psi^{(i)}, \text{ where }
$$

$$
\psi^{(i)} = (-1)^{i-1}.
$$

$$
\left( \prod_{j<i} \mathbb{I}_E(x_j) \cdot |\psi_{\text{IN}}(x_j)| \right) \cdot \psi_{\text{IN}}(x_i) \cdot \left( \prod_{j=i+1}^{t} \mathbb{I}_{f^{-1}(1)}(x_j) \cdot |\psi_{\text{IN}}(x_j)| \right),
$$

where $E$ is set of inputs on which $\psi_{\text{IN}}$ “makes an error”.
Thank you!