Communication Complexity (Spring'26) May 8, 2026

Lecture: XOR Lemma

Instructor: *Toniann Pitassi* Presenter: *Jack Parkhouse*

1 Introduction

In the standard communication complexity model, Alice receives an input x and Bob receives an input y, and together they want to compute a Boolean function

$$f(x, y) : \{0,1\}^q \times \{0,1\}^q \to \{0,1\}$$

while minimizing the number of communicated bits.

A natural question is how the communication complexity of f changes when we want to compute many copies of the function simultaneously. In particular, define

$$f^n(x_1, \ldots, x_n, y_1, \ldots, y_n) = (f(x_1, y_1), \ldots, f(x_n, y_n))$$

and

$$f^{\oplus n}(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(x_1, y_1) \oplus \cdots \oplus f(x_n, y_n)$$

In both scenarios, Alice and Bob are given n different x_i and y_i strings respectively. The function f^n computes n independent copies of f, while $f^{\oplus n}$ computes the parity of the outputs. Since a protocol for f^n also computes $f^{\oplus n}$ and both functions can always be computed by independently running the deterministic protocol for f on each coordinate:

$$D(f^{\oplus n}) \leq D(f^n) \leq n \cdot D(f)$$

The main question studied in the paper is:

> How can we lower bound $D(f^{\oplus n})$ and $D(f^n)$ based on $D(f)$?

Prior work showed lower bounds for $D(f^n)$, but obtaining lower bounds for $D(f^{\oplus n})$ was much more difficult. The main contribution of the paper is a new deterministic XOR lemma that can also be used to lower bound $D(f^n)$:

$$D(f^{\oplus n}) \geq n \cdot \left(\frac{\Omega(D(f))}{\log \mathrm{rk}(f)} - \log \mathrm{rk}(f) \right)$$

The proof proceeds by converting a hypothetical low-communication protocol for $f^{\oplus n}$ into increasingly strong structural information about the communication matrix of f itself. First, a deterministic protocol for $f^{\oplus n}$ with communication m partitions the communication matrix into at most 2^m monochromatic rectangles. By averaging, one of these rectangles must be very large.

Theorem 1. *If $f^{\oplus n}$ can be computed using m bits of deterministic communication, then there exists a monochromatic rectangle for $f^{\oplus n}$ of size at least*

$$\frac{2^{2qn}}{2^m}.$$

The key technical contribution of the paper is then to show that large monochromatic rectangles for $f^{\oplus n}$ force the existence of large monochromatic rectangles for the original function f.

Theorem 2. *If $f^{\oplus n}$ has a monochromatic rectangle of size 2^k, then f has a monochromatic rectangle of size at least*

$$2^{k/n-2}.$$

Finally, the paper shows that if every sufficiently large submatrix of f contains a large monochromatic rectangle, then the communication matrix of f admits a recursive decomposition that yields an efficient deterministic protocol. This creates an upper bound on $D(f)$ in terms of the rectangle structure forced by $f^{\oplus n}$. Rearranging this implication yields the paper's main lower bound:

$$D(f^{\oplus n}) \geq n \cdot \left(\frac{\Omega(D(f))}{\log \mathrm{rk}(f)} - \log \mathrm{rk}(f) \right).$$

Conceptually, the argument shows that a low-communication protocol for $f^{\oplus n}$ would force so much monochromatic structure inside the communication matrix of f that one could recursively exploit this structure to construct an unexpectedly efficient deterministic protocol for f itself.

2 Background: Monochromatic Rectangles

A rectangle is a set of the form $A \times B$ where A is a subset of Alice's inputs and B is a subset of Bob's inputs. A rectangle is monochromatic if f is constant on every point in the rectangle. A monochromatic rectangle for f or $f^{\oplus n}$ is labeled by a single bit, while a monochromatic rectangle for f^n is labeled by an n-bit output vector.

Communication protocols naturally partition the communication matrix into monochromatic rectangles. Therefore, understanding large monochromatic rectangles gives information about communication complexity. We let $C(f)$ denote the minimum number of monochromatic rectangles needed to cover the communication matrix of f. A standard fact from communication complexity is:

$$D(f) \geq \log C(f)$$

3 Step 1: Protocols Give Large Monochromatic Rectangles

The first step converts a low-communication protocol for $f^{\oplus n}$ or f^n into a large monochromatic rectangle.

Theorem 1. *If $f^{\oplus n}$ can be computed using m bits of deterministic communication, then there exists a monochromatic rectangle for $f^{\oplus n}$ of size at least*

$$\frac{2^{2qn}}{2^m}.$$

The communication protocol partitions the communication matrix into at most 2^m monochromatic rectangles. The total number of inputs to f^n is

$$2^{qn} \cdot 2^{qn} = 2^{2qn}.$$

Thus, by an averaging argument, at least one monochromatic rectangle must contain at least

$$\frac{2^{2qn}}{2^m}$$

inputs. Therefore, if communication is small, there must exist a very large monochromatic rectangle.

4 Step 2: Large Rectangles for $f^{\oplus n}$ Give Large Rectangles for f

The key technical contribution of the paper is the following theorem.

Theorem 2. *If $f^{\oplus n}$ has a monochromatic rectangle of size 2^k, then f has a monochromatic rectangle of size at least*

$$2^{k/n-2}.$$

Entropy Background

Recall several important facts about entropy.

1. If a random variable X is supported on N equally likely values, then

$$H(X) = \log_2 N.$$

2. If random variables X, Y are independent, then

$$H(X,Y) = H(X) + H(Y).$$

3. (Chain Rule)

$$H(X,Y) = H(X) + H(Y|X).$$

4. (Extended Chain Rule)

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{<i}).$$

Intuitively, entropy measures the amount of information contained in a random variable.

Setup

Suppose R is a monochromatic rectangle for $f^{\oplus n}$ of size $|R| = 2^k$. Create a random variable that chooses a uniformly random point in R. Since R is a rectangle, the conditional random variables X and Y are independent.

$$X = (X_1, \ldots, X_n) \qquad \text{and} \qquad Y = (Y_1, \ldots, Y_n).$$

Because (X, Y) is uniform over a set of size 2^k,

$$H(X, Y) = k.$$

Applying independence and the chain rule gives

$$\begin{aligned} k &= H(X, Y) \\ &= H(X) + H(Y) \\ &= \sum_{i=1}^{n} H(X_i | X_{<i}) + \sum_{i=1}^{n} H(Y_i | Y_{>i}) \\ &= \sum_{i=1}^{n} H(X_i, Y_i | X_{<i}, Y_{>i}). \end{aligned}$$

Therefore, by averaging, there exists some coordinate i such that

$$H(X_i, Y_i | X_{<i}, Y_{>i}) \geq \frac{k}{n}.$$

Fix values

$$x_{<i}, y_{>i}$$

achieving this bound.

Intuitively, this means that even after fixing all earlier coordinates of Alice and all later coordinates of Bob, the pair (X_i, Y_i) still contains a large amount of entropy. In the case of f^n, we are guaranteed that for all $x \in X_i$ and $y \in Y_i$, $f(x, y)$ is the same due to the fact that we are in a monochromatic rectangle. Therefore, we achieve the theorem for f^n.

The Problem with XOR

For ordinary direct products f^n, this would already be enough. However, for $f^{\oplus n}$, knowing that $f^{\oplus n}(X, Y)$ is the same does *not* immediately imply that $f(X_i, Y_i)$ is monochromatic. The xor could vary across coordinates while keeping the total parity fixed.

The idea is to reveal two more bits that represent the xor of the function values before i and the xor of the function values after i. In doing so, we lower the entropy by at most 2 bits while now ensuring that f is monochromatic. Define:

$$U = f(x_1, Y_1) \oplus \cdots \oplus f(x_{i-1}, Y_{i-1})$$

$$V = f(X_{i+1}, y_{i+1}) \oplus \cdots \oplus f(X_n, y_n).$$

The variables U and V each contain only one bit of information. Using the chain rule again,

$$H(X_i, Y_i | x_{<i}, y_{>i}, U, V) + 2 \geq H(X_i, Y_i | x_{<i}, y_{>i}) \geq \frac{k}{n}$$

Thus,

$$H(X_i, Y_i | x_{<i}, y_{>i}, U, V) \geq \frac{k}{n} - 2.$$

Fix values $u \in U$ and $v \in V$ that achieve this bound. The remaining support contains at least $2^{k/n-2}$ possible pairs (x_i, y_i). Define T to be the support of (X_i, Y_i) after conditioning on the fixed values

$$x_{<i}, y_{>i}, u, v.$$

Since the original rectangle R is monochromatic for $f^{\oplus n}$, the quantity

$$u \oplus f(x_i, y_i) \oplus v$$

must equal the value of the rectangle. Since u, v, and the color of the original rectangle are fixed, the value $f(x_i, y_i)$ must also be fixed throughout T. Therefore, T is a monochromatic rectangle for f of size at least

$$2^{k/n-2}.$$

5 Step 3: Large Rectangles Give Efficient Protocols

The final step shows how the existence of large monochromatic rectangles can be converted into an efficient deterministic communication protocol.

Toy Example

Suppose we are given the following extremely strong property: For every rectangle $X \times Y$ inside the communication matrix, there exists a monochromatic subrectangle of size at least

$$\frac{|X||Y|}{4}.$$

Suppose Alice and Bob know that their current inputs lie somewhere inside a rectangle $X \times Y$. By assumption, this rectangle contains a large monochromatic rectangle R. Rearrange the rows and columns so that R appears in the upper-left corner:

$$\begin{bmatrix} R & A \\ B & Z \end{bmatrix}.$$

Since R is monochromatic, it has rank 1. Using basic properties of matrix rank,

$$\mathrm{rk}(f) \geq \mathrm{rk} \begin{pmatrix} 0 & A \\ B & Z \end{pmatrix} - 1.$$

By Gaussian elimination,

$$\mathrm{rk} \begin{pmatrix} 0 & A \\ B & Z \end{pmatrix} \geq \mathrm{rk} \begin{pmatrix} 0 & A \end{pmatrix} + \mathrm{rk} \begin{pmatrix} 0 \\ B \end{pmatrix}.$$

Therefore,

$$\mathrm{rk}(f) \geq \mathrm{rk} \begin{pmatrix} R & A \end{pmatrix} + \mathrm{rk} \begin{pmatrix} R \\ B \end{pmatrix} - 3.$$

From this, we get that at least one of the following is true:

$$\operatorname{rk}\begin{pmatrix} R & A \end{pmatrix} \leq \frac{\operatorname{rk}(f)+3}{2},$$

$$\operatorname{rk}\begin{pmatrix} R \\ B \end{pmatrix} \leq \frac{\operatorname{rk}(f)+3}{2}.$$

Thus, one side of the matrix must have significantly smaller rank.

Communication Strategy

Assume without loss of generality that

$$\operatorname{rk}\begin{pmatrix} R & A \end{pmatrix} \leq \frac{\operatorname{rk}(f)+3}{2}.$$

Alice now sends one bit indicating whether her input lies among the rows intersecting R.

- If YES, then the protocol recursively continues on $\begin{pmatrix} R & A \end{pmatrix}$ whose rank has decreased by roughly a factor of 2.
- If NO, then the protocol recursively continues on $\begin{pmatrix} B & Z \end{pmatrix}$.

Notice that because R occupied at least one quarter of the original matrix,

$$|B \cup Z| \leq \frac{3|X||Y|}{4}.$$

Thus, every recursive step performs one of two useful operations:

1. Reduce the rank substantially.
2. Reduce the remaining input space by some factor.

Bounding the Number of Steps

The rank-reduction step can occur at most $O(\log \operatorname{rk}(f))$ times, since repeatedly halving the rank eventually reaches rank 1. The space-reduction step shrinks the remaining matrix size by a factor of at least $\frac{3}{4}$. Thus, after at most $O(\log(|X||Y|))$ such steps, the remaining matrix becomes extremely small.

Each root-to-leaf path can therefore be viewed as a sequence of:

- left moves (rank reductions),
- and right moves (space reductions).

The number of possible paths is bounded by

$$\binom{\operatorname{rk}(f) + O(\log \operatorname{rk}(f))}{\operatorname{rk}(f)}.$$

With the following helpful math:

$$\binom{a}{b} \leq a^b \leq 2^{log(a)b}$$

we obtain

$$\binom{\mathrm{rk}(f) + O(\log \mathrm{rk}(f))}{\mathrm{rk}(f)} \leq 2^{O((\log \mathrm{rk}(f))^2)}.$$

Thus, the protocol tree has at most $2^{O((\log \mathrm{rk}(f))^2)}$ leaves. Finally, balancing the protocol tree gives a deterministic protocol of depth

$$O((\log \mathrm{rk}(f))^2).$$

6 Tying Everything Together

The actual paper applies this same recursive strategy to the monochromatic rectangles obtained from Step 1 and Step 2. Step 2 showed that if $f^{\oplus n}$ has a small rectangle cover, then f contains a large monochromatic rectangle of size roughly

$$\frac{|X||Y|}{C(f^{\oplus n})^{1/n}}.$$

The proof then recursively partitions the communication matrix exactly as above. At each recursive step:

- either the rank decreases substantially,
- or the remaining input space shrinks substantially.

Carefully counting the number of possible recursive paths yields an upper bound on the number of leaves in the protocol tree. Balancing the protocol tree finally gives a deterministic protocol for f, which rearranges to the paper's main lower bound:

$$D(f^{\oplus n}) \geq n \cdot \left(\frac{\Omega(D(f))}{\log \mathrm{rk}(f)} - \log \mathrm{rk}(f) \right).$$

The key insight is that sufficiently large monochromatic rectangles force strong recursive structure in the communication matrix, and that recursive structure can be exploited to build efficient deterministic protocols.