An XOR Lemma for Deterministic Communication Complexity

Siddharth Iyer*  siyer@cs.washington.edu
Anup Rao*  anuprao@cs.washington.edu

July 1, 2024

Abstract

We prove a lower bound on the communication complexity of computing the $n$-fold xor of an arbitrary function $f$, in terms of the communication complexity and rank of $f$. We prove that $D(f^\oplus n) \geq n \cdot \left(\Omega(D(f)) - \log \text{rk}(f)\right)$, where here $D(f), D(f^\oplus n)$ represent the deterministic communication complexity, and $\text{rk}(f)$ is the rank of $f$. Our methods involve a new way to use information theory to reason about deterministic communication complexity.

1. Introduction

How is the complexity of computing a Boolean function $f$ on 1 input related to the complexity of computing $f$ on $n$ inputs? In this work, we give new lower bounds for the deterministic communication complexity of computing $f$ on $n$ inputs, making the first progress on this question in many years. We refer the reader to the textbooks [KN97, RY20] for the broader context surrounding these problems and the model of communication complexity.

Given a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, define the functions

$$f^n(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(x_1, y_1), f(x_2, y_2), \ldots, f(x_n, y_n),$$

$$f^\oplus n(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(x_1, y_1) \oplus f(x_2, y_2) \oplus \ldots \oplus f(x_n, y_n).$$

So, $f^n$ computes $f$ on $n$ distinct inputs, and $f^\oplus n$ computes the parity of the outputs of $f$. Because every protocol computing $f^n$ is also a protocol for computing $f^\oplus n$, the complexity of computing $f^\oplus n$ can only be smaller. An important example to keep in mind is when $x, y$ are bits and $f(x, y) = x \oplus y$. Then the communication complexity of $f$ and $f^\oplus n$ are both 2, so there is no increase in the complexity of the xor for such functions.

The communication complexity of a function $f$ is related to the number of monochromatic rectangles needed to cover the inputs to $f$. A monochromatic rectangle is a pair $A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}$ such that $f$ is constant when restricted to $A \times B$. Let $D(f)$ denote the communication complexity of $f$, and let $C(f)$ denote the minimum number of monochromatic rectangles needed to cover the inputs of $f$. It is a standard fact that $D(f) \geq \log C(f)$. Prior to our work, the best known result concerning the complexity of computing these functions was proved by Feder, Kushilevitz, Naor and Nisan [FKN95], who showed that when $\sqrt{D(f)} > \log \log(|\mathcal{X}| \cdot |\mathcal{Y}|)$, $D(f^n)$ grows with $n$.

*Supported by NSF award 2131899.
Theorem 1 ([FKNN95]). \( D(f^n) \geq \log C(f^n) \geq n \cdot (\sqrt{D(f)} - \log \log(|X| \cdot |Y|)) \).

Another important parameter of \( f \) is its rank. The function \( f \) can be viewed as a Boolean matrix \( M \) whose \( (x,y) \)th entry is \((-1)^{f(x,y)}\). We write \( \text{rk}(f) \) to denote the rank of this matrix. Because \( M \) has \( \pm 1 \) entries, it can have at most \( 2^\text{rk}(f) \) distinct rows and at most \( 2^\text{rk}(f) \) distinct columns. This observation leads to the following corollary of Theorem 1:

Corollary 2. \( D(f^n) \geq n \cdot (\sqrt{D(f)} - \log \text{rk}(f) - 1) \).

There have been a number of results concerning the randomized communication complexity of \( f^n \) and \( f^\oplus n \) in recent years. These results rely on definitions from information complexity and simulations of protocols that have small information complexity. See [SK87, Raz92, Raz95, CSWY01, BJKS02, JRS03, BCR10, HJMR10, BR11, Bra15, Kol16, She18, JPY12, BRWY13b, BRWY13a, Yu22, GKR16, RR15, IR24]. However, communication complexity is a model where the deterministic and randomized complexity can be quite far from each other. For example, the randomized communication complexity of the equality function is a constant, but there is no deterministic protocol that beats the performance of the trivial protocol.

In fact, a number of connections between the model of communication complexity and other models of computation are only meaningful when using deterministic protocols. A good example is the connection between circuit depth and communication complexity observed by Karchmer and Wigderson [KW90]. The randomized communication complexity of every Karchmer-Wigderson game is small, because the game can efficiently be solved by hashing. So, lower bounds on circuit depth can only be obtained by studying deterministic communication complexity. Karchmer, Raz and Wigderson [KRW95] conjectured that the communication complexity of this problem increases when the function is composed with itself. Recently, there have been attempts toward this conjecture and on understanding Karchmer-Wigderson games using ideas from information theory [GMWW17, MW19]. If the conjecture is true, this would imply that there is no way to simulate every polynomial time algorithm in logarithmic time with a parallel algorithm. Achieving such tantalizing results motivates us to study the questions about deterministic communication complexity we consider in this paper.

Before the present paper, techniques from information theory had not led to results about the deterministic communication complexity of \( f^n \) or \( f^\oplus n \). That is because known methods to simulate protocols with small information lead to simulations that introduce errors, even if the protocols being simulated do not make errors. In the present paper, we use information theory to obtain results about deterministic communication complexity without introducing any errors. That is the key technical contribution of our work. Our proofs are short, but they circumvent a barrier to applying information theory in the setting of deterministic communication protocols.

Lovász and Saks [LS88] conjectured that there is a constant \( c \) such that \( D(f) \leq (\log \text{rk}(f))^c \). This is called the log-rank conjecture. To date, the best known upper bound is \( D(f) \leq \sqrt{\text{rk}(f)} \) [Lov14, ST23], and it is known that there are \( f \) with \( D(f) \geq (\log \text{rk}(f))^{2-o(1)} \) [GPW18]. Recall that \( D(f) \geq \log \text{rk}(f) \). Our main result gives stronger lower bounds when \( D(f) \gg (\log \text{rk}(f))^2 \):

Theorem 3. \( D(f^\oplus n) \geq \log C(f^\oplus n) \geq n \cdot \left( \frac{\Omega(D(f))}{\log \text{rk}(f)} - \log \text{rk}(f) \right) \).

In comparison to Theorem 1, our result gives lower bounds even for computing the xor \( f^\oplus n \). The key new step of our proof is the following theorem, whose proof uses the sub-additivity of entropy in an essential way:
Theorem 4. If \( f^{\otimes n} \) has a monochromatic rectangle of size \( 2^k \), then \( f \) has a monochromatic rectangle of size \( 2^{k/n-2} \).

The above theorem allows us to use a monochromatic rectangle of large density in \( f^{\otimes n} \) to find a monochromatic rectangle of even larger density in \( f \). Combined with some reasoning about the rank of the function, we are able to use Theorem 4 to obtain a deterministic protocol that proves Theorem 3. In the rest of this paper, we give the details of the proofs of these two theorems.

2. Preliminaries and Notation

For a variable \( X = X_1, \ldots, X_n \), we write \( X_{<i} \) to denote \( X_1, \ldots, X_{i-1} \). We define \( X_{>i} \) similarly. All logarithms are taken base 2. We recall some basic definitions regarding entropy of random variables. Let \( A \) be a random variable distributed according to \( p(a) \). The entropy of \( A \) is defined as

\[
H(A) := \mathbb{E}_{p(a)} \left[ \log \frac{1}{p(a)} \right].
\]

Proposition 5. For any random variable \( A \) with finite support, we have \( H(A) \leq \log |\text{supp}(A)| \), with equality if \( A \) is distributed according to the uniform distribution.

If \( A \) and \( B \) are two jointly distributed random variables distributed according to \( p(ab) \) then the entropy of \( A \) conditioned on \( B \) is defined as

\[
H(A|B) := \mathbb{E}_{p(a,b)} \left[ \log \frac{1}{p(a|b)} \right].
\]

The entropy of jointly distributed random variables satisfy the chain rule:

\[
H(A, B) = H(A) + H(B|A).
\]

Additionally, it is known that the conditional entropy of a random variable cannot exceed its entropy.

Lemma 6. For any two jointly distributed random variables, \( A, B \), we have \( H(A|B) \leq H(A) \), with equality if \( A \) and \( B \) are independent.

We need the following basic fact about rank:

Proposition 7. For any two matrices \( A_1 \) and \( A_2 \), we have \( \text{rk}(A_1 + A_2) \leq \text{rk}(A_1) + \text{rk}(A_2) \).

We need the following lemma that shows that a protocol with a small number of leaves can be computed by a protocol with small communication (see [RY20], Theorem 1.7).

Lemma 8. If \( \pi \) is a deterministic protocol with \( \ell \) leaves, there exists a deterministic protocol computing \( \pi(x, y) \) with communication at most \( \lceil 2 \log_{3/2} \ell \rceil \).
3. Proof of Theorem 4

Let $R$ be a monochromatic rectangle for $f^\oplus n$ of size $2^k$, and let $(X, Y) \in R$ be uniformly random. Because $R$ is a rectangle, $X$ and $Y$ are independent. Using the chain rule, we get

$$k = H(XY) = H(X) + H(Y)$$

(because $X, Y$ are independent)

$$= \sum_{i=1}^{n} H(X_i | X_{<i}) + H(Y_i | Y_{>i})$$

(by the chain rule)

$$= \sum_{i=1}^{n} H(X_i | X_{<i} Y_{>i}) + H(Y_i | X_{<i} Y_{>i} X_i)$$

(because $X, Y$ are independent)

$$= \sum_{i=1}^{n} H(X_i Y_i | X_{<i} Y_{>i}).$$

(by the chain rule)

This implies there exist $i, x_{<i}, y_{>i}$ such that

$$H(X_i Y_i | x_{<i} y_{>i}) \geq k/n.$$

Define the random variables $U = f(x_1, Y_1) \oplus \ldots \oplus f(x_{i-1}, Y_{i-1})$ and $V = f(X_{i+1}, y_{i+1}) \oplus \ldots \oplus f(X_n, y_n)$. By the chain rule, and since $U, V$ are bits, we get

$$H(X_i Y_i | x_{<i} y_{>i} UV) + 2 \geq H(X_i Y_i | x_{<i} y_{>i} UV) + H(UV | x_{<i} y_{>i})$$

$$= H(X_i Y_i | UV | x_{<i} y_{>i})$$

$$\geq H(X_i Y_i | x_{<i} y_{>i})$$

$$\geq k/n,$$

so there is some fixed value of $u, v$ such that

$$H(X_i Y_i | x_{<i} y_{>i} uv) \geq k/n - 2.$$

The desired rectangle is the support of $(X_i, Y_i)$ conditioned on this fixed value of $(x_{<i}, y_{>i}, u, v)$, which we call $T$. Because $(X, Y)$ is distributed uniformly in $R$, the distribution of $(X_i, Y_i)$ conditioned on $(x_{<i}, y_{>i}, u, v)$ is a product distribution, and so $T$ is a rectangle. By Proposition 5, $|T| \geq 2^{k/n - 2}$. Because each input $(x_i, y_i) \in T$ corresponds to some input $(x, y) \in R$ with $f^\oplus n(x, y)$ fixed, and we have fixed $x_{<i}, y_{>i}$ and the xor of the function value in the first $i - 1$ as well as the last $n - i$ coordinates, $f(x_i, y_i)$ is determined within $T$, and $T$ is a monochromatic rectangle of $f$.

4. Proof of Theorem 3

The proof uses Theorem 4 and standard ideas along the lines of [NW95] to obtain a protocol for $f$. We shall prove that $f$ has a protocol tree whose number of leaves is bounded by

$$2^O((\log C(f^\oplus n)^{1/n} + \log rk(f)) \log rk(f))$$

(1)

By applying Lemma 8 to this protocol, we obtain a protocol with communication

$$O\left((\log C(f^\oplus n)^{1/n} + \log rk(f)) \log rk(f)\right) \geq D(f),$$

4
which proves that
\[ \log C(f^{\oplus n})^{1/n} \geq \frac{\Omega(D(f))}{\log \text{rk}(f)} - \log \text{rk}(f), \]
yielding the theorem.

We prove the bound by induction on \(|X| \cdot |Y|\) and \(\text{rk}(f)\). If \(\text{rk}(f) < 5\), or \(|X| \cdot |Y| \leq 1\), we obtain a protocol with a constant number of leaves. Otherwise, by averaging, \(f^{\oplus n}\) has a monochromatic rectangle of size
\[ \frac{|X|^n \cdot |Y|^n}{C(f^{\oplus n})}. \]
Theorem 4 then implies that \(f\) contains a monochromatic rectangle \(R\) of size at least
\[ \frac{|X| \cdot |Y|}{4 \cdot C(f^{\oplus n})^{1/n}}. \]

We can use \(R\) to partition the matrix corresponding to \(f\) as follows
\[ \begin{bmatrix} R & A \\ B & Z \end{bmatrix}. \]
Since \(R\) has rank 1, we have
\[
\begin{align*}
\text{rk}(f) & \geq \text{rk} \left( \begin{bmatrix} 0 \\ B & Z \end{bmatrix} \right) - 1 \quad \text{(Proposition 7)} \\
& \geq \text{rk} \left( \begin{bmatrix} 0 \\ A \end{bmatrix} \right) + \text{rk} \left( \begin{bmatrix} 0 \\ B \end{bmatrix} \right) - 1 \quad \text{by Gaussian elimination} \\
& \geq \text{rk} \left( \begin{bmatrix} R \\ A \end{bmatrix} \right) + \text{rk} \left( \begin{bmatrix} R \\ B \end{bmatrix} \right) - 3. \quad \text{(Proposition 7)}
\end{align*}
\]
So, we must have either
\[ \text{rk}(\begin{bmatrix} R \\ A \end{bmatrix}) \leq \frac{\text{rk}(f) + 3}{2}, \quad (2) \]
or
\[ \text{rk}\left( \begin{bmatrix} R \\ B \end{bmatrix} \right) \leq \frac{\text{rk}(f) + 3}{2}. \]
If Equation (2) holds, Alice sends a bit to Bob indicating whether her input is consistent with \(R\). Otherwise, Bob sends a bit indicating whether his input is consistent with \(R\). Without loss of generality, assume that Equation (2) holds.

Let \(f_0\) and \(f_1\) denote the sub-functions of \(f\) obtained by restricting to \([R \\ A]\) and \([B \\ Z]\) respectively. Since every rectangle cover of \(f^{\oplus n}\) yields a rectangle cover of \(f_0^{\oplus n}\) and a rectangle cover of \(f_1^{\oplus n}\), we have
\[ \max\{C(f_0^{\oplus n}), C(f_1^{\oplus n})\} \leq C(f^{\oplus n}). \]

If Alice’s input is consistent with \(R\), we may repeat the argument with the function \(f_0\) which satisfies \(\text{rk}(f_0) \leq \frac{\text{rk}(f) + 3}{2} \leq \frac{4 \text{rk}(f)}{5}\), so long as \(\text{rk}(f) \geq 5\). Otherwise, if Alice’s input is inconsistent with \(R\), we repeat the argument with the function \(f_1\) which has at most
\[ |X| \cdot |Y| \cdot \left(1 - \frac{1}{4 \cdot C(f^{\oplus n})^{1/n}}\right) \]

The number of recursive steps where the rank reduces by a factor of 4/5 is at most $O(\log \text{rk}(f))$. Moreover, since the matrix corresponding to $f$ has at most $2^{\text{rk}(f)}$ distinct rows and columns, the number of steps where the input space shrinks by a factor of $(1 - \frac{1}{4\cdot C(f^\oplus n)^{1/n}})$ is at most $8 \cdot \text{rk}(f) \cdot C(f^\oplus n)^{1/n}$. That is because after so many steps the number of inputs is at most
\[ 2^{2\cdot \text{rk}(f)} \cdot \left(1 - \frac{1}{4\cdot C(f^\oplus n)^{1/n}}\right)^{8\cdot \text{rk}(f) \cdot C(f^\oplus n)^{1/n}} \leq 2^{2\cdot \text{rk}(f)} \cdot e^{-2\cdot \text{rk}(f)} < 1. \]

The number of leaves in the protocol we have designed is at most
\[ \left(8 \cdot \text{rk}(f) \cdot C(f^\oplus n)^{1/n} + O(\log \text{rk}(f))\right)/O(\log \text{rk}(f)) \leq 2^{O((\log C(f^\oplus n)^{1/n} + \log \text{rk}(f)) \log \text{rk}(f))}, \]
since $C(f^\oplus n) \geq 1$. This proves Equation (1).

References


