A note on the nondeterministic communication complexity of equality

Mathieu Hoyrup*

June 5, 2026

Abstract

This note gathers results estimating the nondeterministic communication complexity of equality. These results are only partially proved in the reference textbooks on communication complexity, and the purpose of this note is to state the results and their proofs in a hopefully clear and explicit manner. Moreover, the nondeterministic communication complexity of inequality is usually given approximately but its exact value turns out to be known.

The purpose of this note is to gather known results. It is partly based on the standard references on communication complexity by Kushilevitz and Nisan [KN96] and by Rao and Yehudayoff [RY20].

Let X, Y be two sets and $f : X \times Y \to \{0, 1\}$. There are various forms of communication complexity expressing the idea that two players, Alice and Bob, want to evaluate f on an input (x, y), but Alice only knows x and Bob only knows y, so they need to communicate and they want to minimize the number of bits of information that are exchanged in the process.

We focus on the particular notions of nondeterministic communication complexity. They can be explained in terms of non-deterministic protocols, but there is another way of thinking about them. We fix some output $v \in \{0, 1\}$ and think of $\{(x, y) : f(x, y) = v\}$ (called the v-set of f) as a set of allowed pairs, and the goal of Alice and Bob is to produce all the allowed pairs. In doing so, they need to coordinate their choices. They first gather to make some agreement about what they will produce, and then make independent choices respecting this agreement (Alice chooses some x, Bob chooses some y). We want that by varying the terms of

*Université de Lorraine, CNRS, Inria, LORIA, Nancy, 54000, France

the agreement, they can produce all the possible pairs. The question is: how many distinct agreements are needed?

For each agreement, the set of pairs (x, y) that are produced by Alice and Bob is a product, also called a rectangle, because their further choices are independent. Moreover, that rectangle is monochromatic in the sense that f takes constant value v on that rectangle. The question then becomes: how many monochromatic rectangles of color v are needed to cover the v-set of f?

We now recall the classical definitions.

1 Definitions

Let X, Y be two sets. A **rectangle** is $R = A \times B$ for some $A \subseteq X$ and $B \subseteq Y$. Note that an intersection of rectangles is a rectangle.

Let $f : X \times Y \to \{0, 1\}$ and $b \in \{0, 1\}$. The b-set of f is $f^{-1}(b)$. A set $R \subseteq X \times Y$ is b-**monochromatic** if f has constant value b on R, i.e. if R is contained in the b-set of f.

Definition 1.1. Let $f : X \times Y \to \{0, 1\}$:

- The **cover number** of f, $C(f)$, is the minimal number of monochromatic rectangles that cover $X \times Y$,
- For $b \in \{0, 1\}$, the b-**cover number** of f, $C^b(f)$, is the minimal number of b-monochromatic rectangles that cover $f^{-1}(b)$,
- The **nondeterministic communication complexity** functions of f are $N(f) = \log_2 C(f)$ and $N^b(f) = \log_2 C^b(f)$.

By definition, one has

$$C(f) = C^0(f) + C^1(f).$$

2 General bounds

We recall the standard techniques to prove lower and upper bounds on the cover numbers.

2.1 Fooling sets

A simple and powerful technique to prove lower bounds on communication complexity is given by fooling sets.

Definition 2.1. Let $b \in \{0,1\}$. A b**-fooling set** for $f : X \times Y \to \{0,1\}$ is a set $F = \{(x_i, y_i) : i \in I\} \subseteq X \times Y$ such that:

- $f(x_i, y_i) = b$ for all $i \in I$,
- For all $i \neq j$, $f(x_i, y_j) \neq b$ or $f(x_j, y_i) \neq b$.

Proposition 2.1 (Fooling set technique)**.** *If F is a b-fooling set for f, then*

$$C^b(f) \geq |F|.$$

Proof. If the b-set of f is covered by rectangles, then each rectangle can only contain one (x_i, y_i), so there are at least $|F|$ rectangles. □

2.2 Relating the 0-cover number and the 1-cover number

Another technique relates the quantities $N^0(f)$ and $N^1(f)$. It follows from results that are stated in [KN96] without proofs (Exercice 2.6 and Proposition 2.2, item 1.), involving the deterministic communication complexity $D(f)$. It can also be obtained by a direct argument that does not rely on this quantity, presented below.

The idea is that if the 0-set of f can be covered by a few rectangles, then its 1-set can be covered by at most exponentially many more rectangles. It gives a coarse estimate in general, but is sometimes rather precise, as we will see in the case of $f = \mathsf{EQ}$.

Proposition 2.2 (Complement technique)**.** *One has*

$$\begin{aligned} N^1(f) &\leq C^0(f), \\ N^0(f) &\leq C^1(f), \\ N(f) &\leq \min(C^0(f), C^1(f)) + 1. \end{aligned}$$

Proof. The complement of a rectangle $R = A \times B$ is covered by two rectangles $S = \overline{A} \times Y$ and $T = X \times \overline{B}$. If the 0-set of f is covered by k rectangles, i.e. $f^{-1}(0) =$

$\bigcup_{i\in[k]} R_i$ with $R_i = A_i \times B_i$, then let $S_i = \overline{A_i} \times Y$ and $T_i = X \times \overline{B_i}$. The 1-set of f is covered by

$$\begin{aligned}\overline{\bigcup_{i\in[k]} R_i} &= \bigcap_{i\in[k]} \overline{R_i} \\ &= \bigcap_{i\in[k]} S_i \cup T_i \\ &= \bigcup_{E\subseteq[k]} \bigcap_{i\in E} S_i \cap \bigcap_{i\notin E} T_i\end{aligned}$$

so it is covered by at most 2^k rectangles. Therefore, $C^1(f) \leq 2^{C^0(f)}$ so $N^1(f) \leq C^0(f)$. The next inequality is obtained symmetrically. Finally,

$$\begin{aligned}N(f) = \log_2(C(f)) &= \log_2(C^0(f) + C^1(f)) \\ &\leq \log_2(C^0(f) + 2^{C^0(f)}) \\ &\leq \log_2(2^{C^0(f)} + 2^{C^0(f)}) \\ &= C^0(f) + 1,\end{aligned}$$

and symmetrically $N(f) \leq C^1(f) + 1$. □

In particular, lower bounds on $N^0(f)$ induce lower bounds on $N^1(f)$ and vice-versa.

3 Equality

Equality is a classical example whose communication complexity can be precisely calculated.

For a positive natural number n, let $[n] = \{1, \ldots, n\}$. Let $\mathsf{EQ} : [n] \to [n]$ send (x, y) to 1 iff $x = y$. Using the two techniques presented above, one can calculate almost exactly the cover numbers of EQ.

Proposition 3.1. *One has*

$$C^1(\mathsf{EQ}) = n,$$
$$\log_2(n) \leq C^0(\mathsf{EQ}) \leq 2\lceil\log_2(n)\rceil.$$

Proof. One has $C^1(\mathsf{EQ}) \leq n$ because the 1-set is covered by the n rectangles $\{x\} \times \{x\}$. The diagonal $\{(x, x) : x \in [n]\}$ is a 1-fooling set of size n for EQ, so $C^1(\mathsf{EQ}) \geq n$ by Proposition 2.1.

The upper bound on $C^0(\mathsf{EQ})$ follows from the decomposition of the 0-set into $2\lceil \log_2(n) \rceil$ rectangles, obtained by choosing the value of a bit of x written in binary, and choosing the opposite value for the corresponding bit of y. More precisely, to each $i < \lceil \log_2(n) \rceil$ and $b \in \{0, 1\}$, we associate the rectangle $R_{i,b} = \{(x, y) : x_i = b \text{ and } y_i = 1 - b\}$.

Lower bounding $C^0(\mathsf{EQ})$ is a well-know example for which the fooling set technique cannot be applied: the maximal size of a 0-fooling set for EQ is 3, for instance $F = \{(1, 2), (2, 3), (3, 1)\}$. Another argument is needed, and is provided by Proposition 2.2, giving $\log_2(n) = N^1(\mathsf{EQ}) \leq C^0(\mathsf{EQ})$. □

Intuitively, in order to produce a pair (x, x), the only option for Alice and Bob is to agree on the value of x, so they need a maximal amount communication. In order to produce a pair (x, y) with $x \neq y$, they need much less communication, agreeing for instance on one of the bits of x and y (and making those bit distinct).

The textbooks [KN96, RY20] on communication complexity do not mention more precise estimates of $C^0(\mathsf{EQ})$. It turns out that its exact value is known. We state the result and explain the context in which it was found.

Proposition 3.2 (Exact value)**.** *One has*

$$\begin{aligned} C^0(\mathsf{EQ}) &= \min\left\{ m : \binom{m}{\lfloor \frac{m}{2} \rfloor} \geq n \right\} \\ &= \log_2(n) + \frac{1}{2}\log_2\log_2(n) + \frac{1}{2}\log_2\left(\frac{\pi}{2}\right) + o(1). \end{aligned}$$

Proof. There is a correspondence between 0-covers of size m and antichains of length n in the power set of $[m]$ ordered by inclusion:

- If $\bigcup_{i\in[m]} A_i \times B_i = \{(x, y) \in [n]^2 : x \neq y\}$, then let $\mathcal{S} = (S_1, \dots, S_n)$ be defined by $S_x = \{i \in [m] : x \in A_i\}$. One easily checks that $\mathcal{S}$ is an antichain w.r.t. inclusion: if $x \neq y$ then $(x, y) \in A_i \times B_i$ for some i, in particular $i \in S_x$; as $A_i \cap B_i = \emptyset$, one has $i \notin S_y$ hence $S_x \nsubseteq S_y$ (and symmetrically, $S_y \nsubseteq S_x$),
- Conversely, given an antichain $\mathcal{S} = (S_1, \dots, S_n)$ of subsets of $[m]$, define $A_i = \{x \in [n] : i \in S_x\}$ and $B_i = \overline{A_i}$. One easily checks that $(A_i \times B_i)_{i\in[m]}$ is a cover of the 0-set of EQ: if $x \neq y$, then let $S_x \nsubseteq S_y$ so there exists $i \in S_x \setminus S_y$, hence $(x, y) \in A_i \times B_i$.

Therefore, there exists a cover of the 0-set of EQ by m rectangles if and only if there exists an antichain of length n in $\mathcal{P}([m])$ ordered by inclusion. Sperner's theorem states that such an antichain exists if and only $\binom{m}{\lfloor \frac{m}{2} \rfloor} \geq n$, in which case it can be obtained by taking n subsets of $[m]$ of size $\lfloor \frac{m}{2} \rfloor$. □

This proof is due to Spencer [Spe70] who formulated it in the context of *completely separating systems*: such a system is a family $\mathcal{S}$ of subsets of $[n]$ such that for every pair of distinct elements $x, y \in [n]$, there exists $S \in \mathcal{S}$ containing x and not y (hence there also exists $T \in \mathcal{S}$ containing y and not x). Dickson [Dic69] introduced this notion as a strengthening of the asymmetric notion of separating system studied by Rényi [Rén61] and Katona [Kat66]. Dickson proved that their minimal size is of the order of $\log_2(n)$ and Spencer [Spe70] found their exact value by observing that they are in one-to-one correspondance with antichains and applying Sperner's theorem.

In terms of communication complexity, to each $x \in [n]$ we associate a binary string $s_x \in \{0,1\}^m$ in an injective way, so a proof that $x \neq y$ can be obtained by giving an index $i \in [m]$ such that $s_x(i) = 1$ and $s_y(i) = 0$. In order to cover the 0-set of EQ, we need the s_x's to be pairwise incomparable w.r.t. the bitwise ordering. The minimal m for which we can find such strings happens to be the one given in the statement.

The usual upper bound $\mathcal{C}^0(\mathsf{EQ}) \leq 2\lceil \log_2(n) \rceil$ given in Proposition 3.1 is actually an instance of this strategy, where $s_x = x\overline{x}$ is the binary expansion of x concatenated with the same string in which each bit is flipped.

References

[Dic69] T.J. Dickson. On a problem concerning separating systems of a finite set. *Journal of Combinatorial Theory*, 7(3):191–196, 1969.

[Kat66] G. Katona. On separating systems of a finite set. *Journal of Combinatorial Theory*, 1(2):174–194, 1966.

[KN96] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

[Rén61] Alfred Rényi. On random generating elements of a finite boolean algebra. *Acta Scientiarum Mathematicarum Szeged*, 22:75–81, 1961.

[RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[Spe70] Joel Spencer. Minimal completely separating systems. *Journal of Combinatorial Theory*, 8(4):446–447, 1970.