Communication complexity tools on recognizable picture languages

Véronique Terrier

Communication complexity tools on recognizable picture languages

Véronique Terrier
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
14000 Caen, France

Abstract

The paper deals with the class `REC` of recognizable picture languages, `UREC` its unambiguous variant and `co-REC` the complement class of `REC`. The aim of this paper is two-fold:

First, the paper focuses on some necessary conditions for a language to be recognizable. Such conditions have already been identified in several works [13, 1, 5, 2, 3]. Here, we revisit them in the light of communication complexity arguments.

Second, we use communication complexity measures in order to construct a language which is a recognizable and co-recognizable language but not an unambiguous one. This answers a question raised in [14].

1 Introduction

The class `REC` of recognizable picture languages generalizes the notion of regular languages to two-dimensional setting. Initially introduced by D. Giammarresi and A. Restivo in [4], this class has been proved to be very robust and has been characterized in several different ways, extending classic specifications of regular string languages. For instance the recognizable picture languages can be described by tiling systems (i.e., as projections of local sets of tiles), by the on-line tessellation automata, or by a fragment of existential monadic second-order logic [4, 10, 6, 7].

However the situation for recognizable picture languages is not as uniform as that for regular string languages. Unlike one-dimensional case, adding non-determinism increases the expressive power; indeed, the class `REC` is larger than its deterministic variant. In particular, `REC` contains NP-complete problems [12]. Further many other properties proved in the one-dimensional case, are no longer true; among other, `REC` is not closed under complement [10].

Here, we are mainly interested in a subclass of `REC`, the class `UREC` of unambiguous recognizable picture languages. Several studies have already made explicit the reduction of expressive power induced by the constraint of non-ambiguity and, notably, have exhibited a necessary condition for any language to be unambiguously recognizable [1]. In this paper, we will see how such a criterion can be directly obtained from communication complexity arguments. In addition we will focus on the following open question which was first raised in a more general context of grid graphs [14, 5, 2]: "Is there a recognizable and co-recognizable language which is not unambiguous?" We will give a positive answer. For that, we will use communication complexity results to identify a candidate L with a significant gap between unambiguous communication complexity on one side and non-deterministic and co-non-deterministic complexities on the other side.

The paper is organized as follows. Section 2 contains the definitions of recognizable picture language, tiling system and its unambiguous variant. Section 3 recalls the necessary background about communication complexity that we will need later on. Section 4 revisits already known lower bounds for tiling systems in the light of communication complexity. Finally, the presentation of a recognizable and co-recognizable language which is not unambiguous, is given in Section 5.

2 Tiling system

We recall the standard definitions about recognizable picture languages.

Definition 1 (picture)**.** For $m, n \in \mathbb{N}$ and Σ an alphabet, a *picture* p of *size* (m, n) is an array with m rows and n columns of elements over Σ. The pixel $p(i, j)$ refers to the symbol on the i-th row and the j-th column.
Let $\sharp$ be a special symbol not in Σ. The *bordered picture* of p is the picture $\hat{p}$ of size $(m+2, n+2)$ obtained by surrounding p with $\sharp$.

$\Sigma^{m,n}$ denotes the set of all pictures over Σ of size (m, n) and Σ^{**} the set of all pictures over Σ. A *picture language* over Σ is a subset of Σ^{**}.

Definition 2 (tile, tilable picture)**.** A *tile* is a picture of size $(2, 2)$.
For Σ an alphabet, Θ a set of tiles over $\Sigma \cup \{\sharp\}$ and p a picture over Σ, Θ *tiles* p if all the sub-pictures of size $(2, 2)$ of the bordered picture $\hat{p}$ belong to Θ.

A language which can be described by the set of its $(2, 2)$-sub-pictures, is said local.

Definition 3 (local language)**.** A picture language L over Σ is *local* if there exists a set of tiles Θ over $\Sigma \cup \{\sharp\}$ such that L contains exactly those pictures tilable by Θ. If so, we write $L = \mathrm{Local}(\Theta)$.

The key notion of recognizable languages is defined in terms of local language and projection, and so involves both locality and non-determinism.

Definition 4 (projection)**.** For Γ and Σ two alphabets and $\pi : \Gamma \to \Sigma$ a map, the projection by π of a picture $p \in \Gamma^{m,n}$ is the picture $p' \in \Sigma^{m,n}$ such that $p'(i, j) = \pi(p(i, j))$ for all pixels. The projection $\pi(L)$ of a language L over Γ is the language over Σ defined as $\{\pi(p) : p \in L\}$.

Definition 5 (tiling system)**.** A tiling system is a quadruple $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$ where

- Σ and Γ are finite alphabets;
- Θ is a set of tiles over $\Sigma \cup \{\sharp\}$;
- $\pi : \Sigma \to \Gamma$ is a projection.

Definition 6 (recognizable language)**.** A picture language L over Γ is *tiling recognizable* if there exist a tiling system $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$ such that $L = \pi(\mathrm{Local}(\Theta))$.
`REC` denotes the family of all recognizable picture languages and `co-REC` the family of languages whose complement is in `REC`.

Example 1 (the language `INTERVAL`)**.** The language of all pictures p over $\{0, 1\}$ such that, given m the height of p, there exist two columns of p at distance m which are equal.

$$\texttt{INTERVAL} = \Big\{ p \in \{0,1\}^{**} : \ p \text{ is of the size } (m, n) \text{ and} \\ \text{it exists } j \text{ such that for all } i = 1, \ldots, m \, : \, p(i, j) = p(i, j+m) \Big\}$$

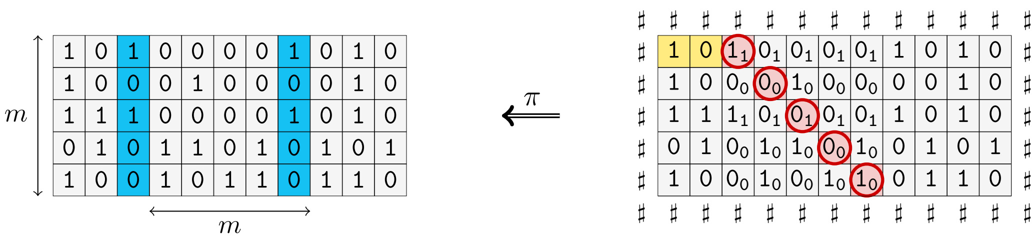


Figure 1: A picture of `INTERVAL` and its tiling

INTERVAL is in REC. We just give an informal description of the tiling system. As depicted in Figure 1, the tiling system will

(i) choose non-deterministically one column: it is materialized by the line drawn from the upper left corner;

(ii) transfer the chosen column to the right: the column elements are recorded in the subscripts;

(iii) determine the column $i+m$ at distance m of the chosen column i: it is achieved by the mean of a diagonal which links the top of the column i with the bottom of the column $i+m-1$;

(iv) verify that the columns i and $i+m$ are equal: the tiling ensures that each subscript of the column $m+i-1$ matches the value on its right.

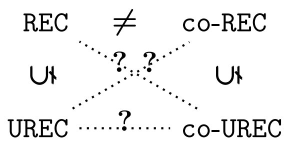
Here is another tiling for the same picture. That means the tiling is ambiguous: one picture admits more than one local pre-images.

$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$
$\sharp$	1	0	1	0_0	0_0	0_0	0_0	1_0	0	1	0	$\sharp$
$\sharp$	1	0	0	0_0	1_0	0_0	0_0	0_0	0	1	0	$\sharp$
$\sharp$	1	1	1	0_0	0_0	0_0	0_0	1_0	0	1	0	$\sharp$
$\sharp$	0	1	0	1_1	1_1	0_1	1_1	0_1	1	0	1	$\sharp$
$\sharp$	1	0	0	1_1	0_1	1_1	1_1	0_1	1	1	0	$\sharp$
$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$	$\sharp$

A tiling system is unambiguous if it does not have two different tilings for any picture.

Definition 7 (unambiguous)**.** A tiling system $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$ that recognizes a language L is *unambiguous* if for every picture $p \in L$ there exists exactly one picture $q \in \mathrm{Local}(\Theta)$ such that $p = \pi(q)$.
Any picture language recognized by an unambiguous tiling system is said *unambiguous*. UREC denotes the family of all unambiguous recognizable picture languages and co-UREC the family of languages whose complement is in UREC.

To end this section, let us mention the currently known relationships among these classes REC, co-REC, UREC and co-UREC. REC and co-REC are incomparable because REC is not closed under complement [10]. UREC (resp. co-UREC) is properly contained in REC (resp. co-REC) [1]. Unfortunately, no relation is known between UREC and co-UREC, and even, between UREC and co-REC.



3 Communication complexity

The communication complexity framework provides efficient tools for proving lower bounds. Introduced by Yao [15], the communication complexity measures the amount of communication needed to compute a global function whose inputs are distributed among several parties. Typically, two players Alice and Bob wish to jointly compute a function $f : X \times Y \to \{0, 1\}$. Alice gets $x \in X$ and Bob gets $y \in Y$. They exchange bits according to some fixed protocol and stop when both know the value $f(x, y)$. Several different models of protocols exist.

Now we just list the definitions and results needed for our study. A comprehensive treatment may be found in the book by Kushilevitz and Nisan [11]. For our context, since tiling systems are intrinsically a non-deterministic device, the non-deterministic model will be appropriate.

Definition 8 (non-deterministic and unambiguous protocols)**.** For $f : X \times Y \to \{0,1\}$ a function, a *non-deterministic protocol* for f of *cost* k consists of two functions $A : X \times \{0,1\}^k \to \{0,1\}$ and $B : Y \times \{0,1\}^k \to \{0,1\}$ such that

- $f(x,y) = 1 \Rightarrow \exists z \in \{0,1\}^k : A(x,z) = 1$ and $B(y,z) = 1$
- $f(x,y) = 0 \Rightarrow \forall z \in \{0,1\}^k : A(x,z) = 0$ or $B(y,z) = 0$

Such a protocol is *unambiguous* if for each tuple (x,y) such that $f(x,y) = 1$ there is exactly one z: $A(x,z) = 1$ and $B(y,z) = 1$.

The above protocol can be understood in this following way. Alice and Bob both share the functions f, A and B. If Alice receives x and Bob y, then the information required to convince them that $f(x,y) = 1$, corresponds to the certificate z of size k. Now the communication complexity of f quantifies the information transfer.

Definition 9 (non-deterministic and unambiguous communication complexities)**.** The *non-deterministic communication complexity* of f, $\mathcal{N}_1(f)$, is the minimal cost k of a protocol, over all non-deterministic protocols for f.
The *co-non-deterministic communication complexity* of f, $\mathcal{N}_0(f)$, equals $\mathcal{N}_1(\neg f)$ where $\neg f(x,y) = 1 - f(x,y)$.
The *unambiguous communication complexity* of f, $\mathcal{UN}_1(f)$, is the minimal cost k of a protocol, over all unambiguous protocols for f.

To evaluate these complexities, the fundamental ingredients are the characteristic matrix of the function f, also named communication matrix or Hankel matrix, and its all ones or all zeros sub-matrices.

Definition 10 (characteristic matrix, monochromatic rectangle)**.** For $f : X \times Y \to \{0,1\}$ a function, the *characteristic matrix* M_f is a $|X| \times |Y|$ matrix whose (x,y) entry is $f(x,y)$. Let v be 0 or 1. A *v-monochromatic rectangle* of M_f is a subset $R = A \times B \subseteq X \times Y$ such that $f(x,y) = v$ for all $x \in A$ and $y \in B$.

Several attributes describe the structure of the characteristic matrix:

Definition 11 (cover and partition numbers, rank)**.** For $f : X \times Y \to \{0,1\}$ a function and M_f its characteristic matrix

- The *v-cover number*, $C^v(f)$, is the minimum number of v-monochromatic rectangles needed to cover all the v in M_f (possibly with overlaps).
- The *v-partition number*, $X^v(f)$, is the minimum number of v-monochromatic rectangles needed to cover all the v in M_f without overlap.
- The *rank*, $\mathrm{rank}(f)$, is the rank of M_f.

As emphasized in the following key lemma, a non-deterministic protocol for f of cost k corresponds to a cover of all the 1 with at most 2^k rectangles. For unambiguous protocol, the cover involves rectangles with no overlaps.

Lemma 1.

- $\mathcal{N}_1(f) = \log C^1(f)$ and $\mathcal{N}_0(f) = \log C^0(f)$
- $\mathcal{UN}_1(f) = \log X^1(f)$ and $\mathcal{UN}_0(f) = \log X^0(f)$

Based on linear algebra, the next results relate the partition number and the ranks.

Fact 1.

- $X^1(f) \geq \mathrm{rank}(f)$ and $X^0(f) \geq \mathrm{rank}(\neg f)$
- $|\,\mathrm{rank}(f) - \mathrm{rank}(\neg f)| \leq 1$

It gives us lower bounds on unambiguous communication complexities in terms of rank.

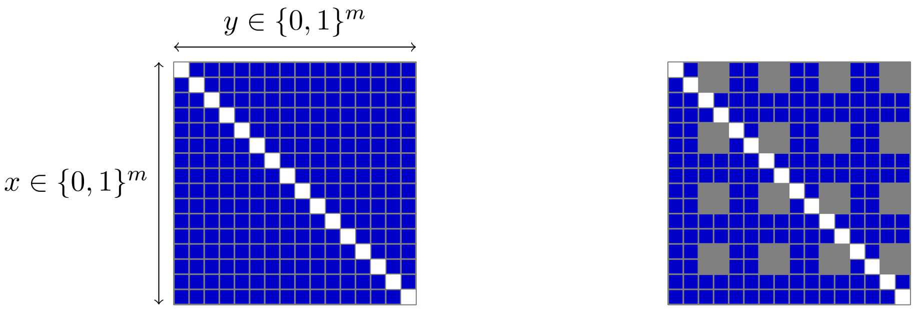
Lemma 2. $\mathcal{UN}_1(f) \geq \log \operatorname{rank}(f)$ and $\mathcal{UN}_0(f) \geq \log(\operatorname{rank}(f) - 1)$

Here is an example to illustrate these notions.

Example 2 (the function $\texttt{NEQ}$)**.** The non-equality function is as following.

$$\begin{array}{rccl} \texttt{NEQ}_m: & \{0,1\}^m \times \{0,1\}^m & \to & \{0,1\} \\ & (x,y) & \to & \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{else} \end{cases} \end{array}$$

The characteristic matrix of $\texttt{NEQ}_m$ is the complement of the $2^m \times 2^m$ identity matrix.



(a) $M_{\texttt{NEQ}_4}$
■ stands for 1
□ stands for 0

(b) A 1-monochromatic rectangle drawn in gray
With the inputs in lexicographic order,
it corresponds to $R_{2,0} = \{x_0x_1x_2x_3 \in \{0,1\}^4 : x_2 = 0\} \times \{y_0y_1y_2y_3 \in \{0,1\}^4 : y_2 = 1\}$

Figure 2: The characteristic matrix of $\texttt{NEQ}_m$ with $m = 4$.

Observe that for any couple $(x, y) \in \{0,1\}^m \times \{0,1\}^m$, a certificate that $x \neq y$ is an integer i in range $0, \ldots, m-1$ and a value $v = 0$ or 1 such that $x_i = v$ and $y_i = 1 - v$. The size of the certificate is $\log m + 1$ ($\log m$ bits to code i and 1 bit to code v). In parallel, the couples (x, y) such that $x_i = v$ and $y_i = 1 - v$ draw a 1-monochromatic rectangle $R_{i,v} = \{x \in \{0,1\}^m : x_i = v\} \times \{y \in \{0,1\}^m : y_i = 1 - v\}$. And the union of the $2m$ rectangles $R_{i,v}$: $\bigcup_{0 \leq i < m, v=0,1} R_{i,v}$ is a 1-cover of the characteristic matrix. In other words, $\mathcal{N}_1(\texttt{NEQ}_m) \leq \log m + 1$. On the other hand, these rectangles overlap and do not form a 1-partition. As a matter of fact, the characteristic matrix has full rank 2^m. Thus $\mathcal{UN}_1(\texttt{NEQ}_m) \geq \log \operatorname{rank}(\texttt{NEQ}_m) = m$. Note also that $\mathcal{N}_0(\texttt{NEQ}_m) = \mathcal{UN}_0(\texttt{NEQ}_m) = m$ since all 0-monochromatic rectangles are of size 1 and form a 0-partition.

In the sequel, to specify the communication complexities bounds, we will use the standard asymptotic notation. The *big O* defines the upper bounds: $O(f)$ is the set of functions bounded above by f. The *big omega* defines the lower bounds: $\Omega(f)$ is the set of functions bounded below by f. The *little omega* defines the strict lower bounds: $\omega(f)$ is the set of functions dominating f.

4 Tiling system limits in light of communication complexity

Several different lower bounds on picture language recognition have been demonstrated in a set of works [13, 1, 5, 2, 3]. The common approach is based on a technique that consists in reducing two-dimensional languages to string languages over the alphabet of the columns. Then the results are obtained in the setting of regular string languages. As a matter of fact, all these proofs involve indirectly communication complexity arguments. Here, our aim is to make explicitly use of the communication complexity tools to yield direct and simple proofs of these results. Let us mention that communication complexity has already proved to be fruitful in the context of tiling systems over infinite picture [8].

Besides the techniques, as considered below on finite pictures, simply generalize the communication complexity techniques specified for regular string languages in [9].

The starting point rests on the following common observation. When we divide a picture into two parts, the size of the cut sets a bound on the information that can pass from one side to the other. For a tiling system, the communication consists of the tiled boundary between the two parts. Now the question is in what extent this bottleneck will entail limit on its recognition capacity. The communication complexity framework will help us to formalize this.

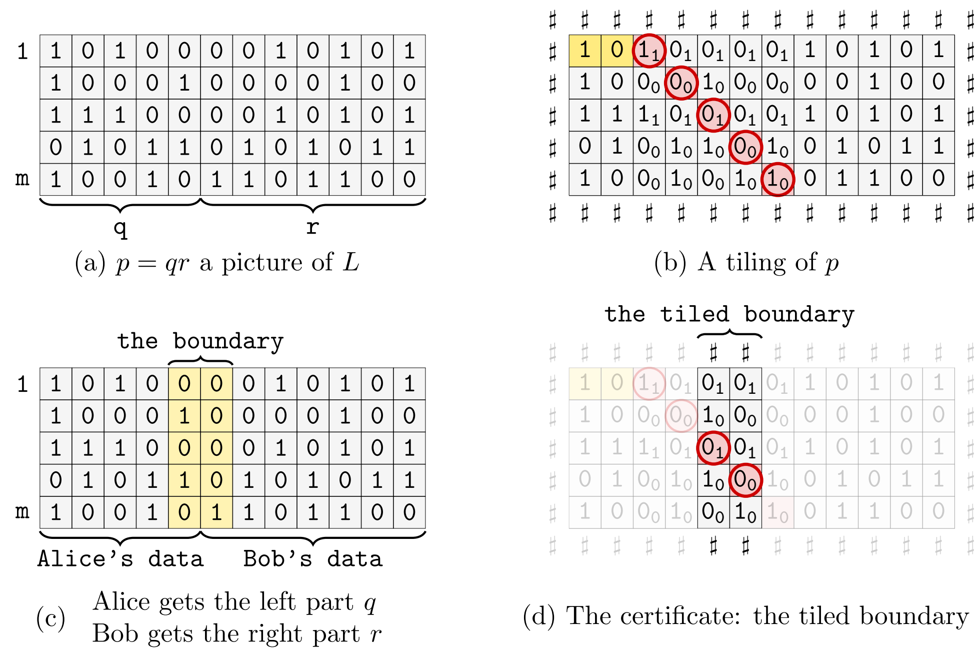
The capacity of a tiling system to recognize some language L will be limited by the communication complexity of the following related functions.

Definition 12. Let L be a picture language over Σ. We associate to L a family of boolean functions $(g^L_{m,n})$

$$\begin{array}{rccl} g^L_{m,n}: & \bigcup\limits_{i=1}^{n} \Sigma^{m,i} \times \bigcup\limits_{i=1}^{n} \Sigma^{m,i} & \to & \{0,1\} \\ & (q,r) & \to & \begin{cases} 1 & \text{if } qr \in L \\ 0 & \text{else} \end{cases} \end{array}$$

Proposition 1. Let L be a recognizable picture language. Then, for any n, it exists a non-deterministic protocol for $g^L_{m,n}$ whose cost is in $O(m)$.

Proof. A non-deterministic protocol for $g^L_{m,n}$ in which Alice and Bob share a tiling system recognizing L is as following. See Figure 3. Alice gets $q \in \bigcup\limits_{i=1}^{n} \Sigma^{m,i}$ and Bob gets $r \in \bigcup\limits_{i=1}^{n} \Sigma^{m,i}$. If the concatenation $p = rq$ is in L, it exists a tiling of p. In that case, the certificate z is the tiled boundary between the two parts r and q. Indeed Alice and Bob can respectively check that the certificate z matches their own picture part. On the other hand, if p is not in L, no tiling for p exists. Hence, Alice and Bob can not agree whatever z, the tiled boundary between the two parts q and r, is. Thus the cost of this non-deterministic protocol corresponds to the size of the tiled boundary, that is in $O(m)$.



(a) $p = qr$ a picture of L

(b) A tiling of p

(c) Alice gets the left part q
Bob gets the right part r

(d) The certificate: the tiled boundary

Figure 3: A non-deterministic protocol for $g^L_{m,n}$ when $L \in \mathtt{REC}$

□

Proposition 2. Let L be an unambiguous recognizable picture language. Then, for any n, it exists an unambiguous protocol for $g^L_{m,n}$ whose cost is in $O(m)$.

Proof. Under the additional assumption that it exits a tiling system for L which is unambiguous, the above protocol turns out to be unambiguous. Indeed, providing there exists exactly one tiling for every $p = rq$ in L, the certificate z is also unique. □

Putting together Lemmas 1 & 2 and Propositions 1 & 2, we find again the results stated in [3] and [1] that give recognizability conditions through measures of the characteristic matrix:

Theorem 1. (i) If $L \in \texttt{REC}$ then, whatever n is, $\log C^1(g^L_{m,n})$ is in $O(m)$.

(ii) If $L \in \texttt{UREC}$, then, whatever n is, $\log \mathrm{rank}(g^L_{m,n})$ is in $O(m)$.

Let us notice that the protocols described in the above propositions, consider all possible vertical cuts of the pictures. In the following, it will be sufficient to adopt a simplified view with only a single cut taken into account, the one that divides the pictures in two halves. Incidentally only pictures of even width will be involved. Concretely, we will consider the subfunctions $f^L_{m,n}$ that are the restrictions of $g^L_{m,n}$ on the domain $\Sigma^{m,n} \times \Sigma^{m,n}$

Definition 13. Let L be a picture language over Σ. The family of functions $(f^L_{m,n})$ is defined by:

$$\begin{array}{rccl} f^L_{m,n}: & \Sigma^{m,n} \times \Sigma^{m,n} & \to & \{0,1\} \\ & (q,r) & \to & \left\{ \begin{array}{ll} 1 & \text{if } qr \in L \\ 0 & \text{else} \end{array} \right. \end{array}$$

Obviously, the functions $g^L_{m,n}$ inherit the lower bounds of the subfunctions $f^L_{m,n}$: $C^1(f^L_{m,n}) \leq C^1(g^L_{m,n})$ and $\mathrm{rank}(f^L_{m,n}) \leq \mathrm{rank}(g^L_{m,n})$. Thus a weak form of Theorem 1 is as following.

Corollary 1. (i) If $L \in \texttt{REC}$ then, whatever n is, $\log C^1(f^L_{m,n})$ is in $O(m)$.

(ii) If $L \in \texttt{UREC}$, then, whatever n is, $\log \mathrm{rank}(f^L_{m,n})$ is in $O(m)$.

5 A recognizable and co-recognizable language which is not unambiguous

In this section we consider the question raised in [14]. Is there a recognizable and co-recognizable language which is not unambiguous? We will see that the answer is positive: it exists a language inside $(\texttt{REC} \cap \texttt{co-REC}) \setminus (\texttt{UREC} \cup \texttt{co-UREC})$. A candidate must be a language L easy to compute non-deterministically and co-non-deterministically but difficult to compute unambiguously. In terms of communication complexity, it means that the non-deterministic and co-non-deterministic communication complexities of $f^L_{m,n}$ are both in $O(m)$ but the rank of $f^L_{m,n}$ is in $2^{\omega(m)}$.

We can find in [11] the list-non equality predicate with a significant gap between unambiguous and non-deterministic communication complexities. It takes as input: two lists of length n of words of m bits each and tests whether the two respective j-th words differ for all j. In short, it simply corresponds to the family of boolean functions $(f^{\texttt{LNE}}_{m,n})$ associated to the following picture language LNE.

Example 3 (the language LNE)**.** LNE is the set of all pictures over $\{0,1\}$ of size $(m,2n)$ such that the j-th and $(n+j)$-th columns differ for all j.

$$\begin{aligned} \texttt{LNE} = \big\{ p \in \{0,1\}^{**} : \ & p \text{ is of the size } (m,2n) \text{ and} \\ & \text{for all } j = 1,\ldots,n, \text{ it exists } i \text{ such that } p(i,j) \neq p(i,j+n) \big\} \end{aligned}$$

0	1	0	0	0	0	0	0	1	0
1	0	0	1	0	0	0	0	1	0
1	1	0	0	0	0	1	0	1	0
0	0	1	1	0	1	0	0	0	1

(a) A picture in LNE

0	1	0	0	0	0	1	0	1	0
1	0	0	1	0	0	0	0	1	0
1	1	0	0	0	0	1	0	1	0
0	0	1	1	0	1	0	0	0	1

(b) A picture not in LNE

Figure 4: The language LNE

The three following facts present some bounds on the communication complexities of LNE. They come from [11] and are rephrased to fit our context. On the one hand, upper bounds on the non-deterministic communication complexities of LNE and its complement are obtained by describing appropriate protocols.

Fact 2. $\mathcal{N}_1(f^{\texttt{LNE}}_{m,n}) \leq O(n \log m)$

Proof. A non-deterministic protocol where Alice and Bob get each an half of the picture p, is as following. A certificate to convince Alice and Bob that p is in LNE provides for each column $j = 1, \cdots, n$ a row $\alpha(j)$ and a bit v such that $p(\alpha(j), j) = v$ and $p(\alpha(j), n+j) = 1 - v$. The certificate contains the list of n couples composed of the $\log m$ bits of the value $\alpha(j)$ and the corresponding bit $p(\alpha(j), j)$. It gives a certificate of size $n(\log m + 1)$ bits. □

Fact 3. $\mathcal{N}_0(f^{\texttt{LNE}}_{m,n}) \leq O(\log n + m)$

Proof. A certificate to convince Alice and Bob that p is not in LNE consists of a column j such that $p(i, j) = p(i, n+j)$ for all i. The certificate contains the $\log n$ bits of the value j and the m bits of the column. □

On the other hand, lower bounds on the unambiguous communication complexities of LNE and its complement are obtained through the rank approach.

Fact 4. $\mathcal{UN}_1(f^{\texttt{LNE}}_{m,n}) \geq \Omega(mn)$ and $\mathcal{UN}_0(f^{\texttt{LNE}}_{m,n}) \geq \Omega(mn)$

Proof. Recall Lemma 2 that provides lower bounds on unambiguous communication complexities through the rank of the characteristic matrix: $\mathcal{UN}_1(f^{\texttt{LNE}}_{m,n}) \geq \log \operatorname{rank}(f^{\texttt{LNE}}_{m,n})$ and $\mathcal{UN}_0(f^{\texttt{LNE}}_{m,n}) \geq \log(\operatorname{rank}(f^{\texttt{LNE}}_{m,n}) - 1)$. Now algebraic techniques will allow to compute the rank of $M_{f^{\texttt{LNE}}_{m,n}}$. First consider the characteristic matrix $M_{f^{\texttt{LNE}}_{m,1}}$ where $n = 1$. It is the complement of the $2^m \times 2^m$ identity matrix and has full rank 2^m. Then the characteristic matrix $M_{f^{\texttt{LNE}}_{m,n}}$ for $f^{\texttt{LNE}}_{m,n}$ corresponds to the $2^{mn} \times 2^{mn}$ matrix obtained by the Kronecker product of n matrices $M_{f^{\texttt{LNE}}_{m,1}}$. Hence its rank is $(2^m)^n = 2^{mn}$. □

In light of these facts, we observe that if n is in $O(m/\log m)$ then $\mathcal{N}_1(f^{\texttt{LNE}}_{m,n})$ and $\mathcal{N}_0(f^{\texttt{LNE}}_{m,n})$ are upper bounded by $O(m)$. On the other hand, if n is in $\omega(1)$ then $\mathcal{UN}_1(f^{\texttt{LNE}}_{m,n})$ and $\mathcal{UN}_0(f^{\texttt{LNE}}_{m,n})$ are lower bounded by $\omega(m)$. We can now define a subset of LNE, containing pictures of appropriate size, which has low non-deterministic and co-non-deterministic complexities but high unambiguous complexity.

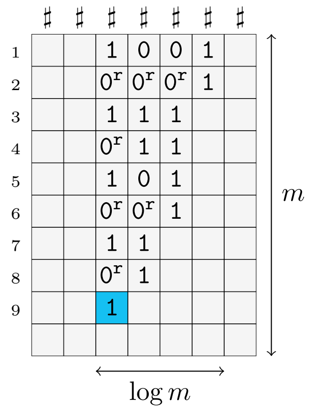
Example 4 (the language CANDIDATE)**.** It is the intersection of LNE with the set of pictures of size $(m, 2\lfloor m/\lceil \log m \rceil \rfloor)$.

$$\begin{aligned}\texttt{CANDIDATE} = \big\{ p \in \{0,1\}^{**} : & \text{ the size of } p \text{ is } (m, 2n) \text{ where } n = \lfloor m/\lceil \log m \rceil \rfloor \text{ and} \\ & \text{for all } j = 1, \ldots, \lfloor m/\lceil \log m \rceil \rfloor, \text{ it exists } i \text{ s.t. } p(i,j) \neq p(i, j+n) \big\}\end{aligned}$$

The task is now to make sure that CANDIDATE is effectively in REC and co-REC. More specifically, we have to show that CANDIDATE and its complement can be implemented by tiling systems. Actually we will see that CANDIDATE needs some adjusting to be tiling recognizable.

Facts 2 and 3 specify two non-deterministic protocols for CANDIDATE and its complement with low communication complexities. But, during the protocols, Alice and Bob have unlimited power to process the information they own while it is not the case for tiling systems. In particular the protocol for CANDIDATE requires the most resources: the tiling system must compute the $\log m$ bits of the row index of the pixel selected in each of the n columns.

For one specific pixel, a tiling system can set up a binary counter which starts from the location of the pixel and counts up to the top of the picture. The binary value reached is therefore its row index. But the binary counter spreads over several columns, possibly up to $O(\log m)$ for a picture of height m. And, with a binary counter for each column, the bug is that the counters overlap.



To patch that, we simply reshape the language CANDIDATE: we enlarge the picture in inserting vertical stripes of width $\log m$ between each column. In this way, the tilling system will have enough space to implement the counters. In the same time, it does not change the height of the picture that is the size of our bottleneck.

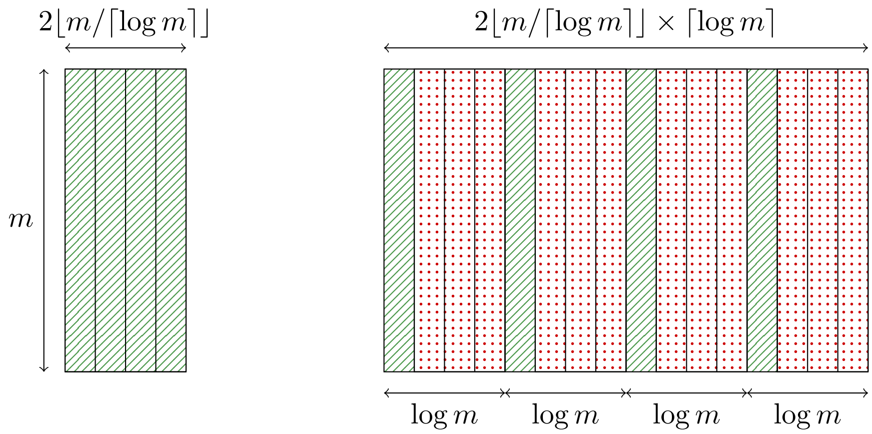


Figure 5: Reshaping CANDIDATE into PROPER

Example 5 (the language PROPER)**.**

$$\texttt{PROPER} = \big\{p \in \{0,1\}^{**} : \text{ the size } (m, 2n) \text{ of } p \text{ is such that } n = \lfloor m/\lceil \log m\rceil\rfloor \lceil \log m\rceil \\ \text{and for all } j = 0, \ldots, \lfloor m/\lceil \log m\rceil\rfloor - 1, \text{ it exists } i \text{ such that} \\ p(i, 1 + j\lceil \log m\rceil) \neq p(i, 1 + j\lceil \log m\rceil + n)\big\}$$

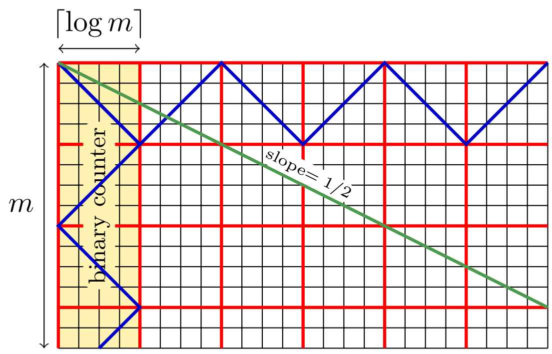
Once everything has been done to make it work well, let us verify that PROPER is tiling recognizable.

Proposition 3. PROPER is in REC.

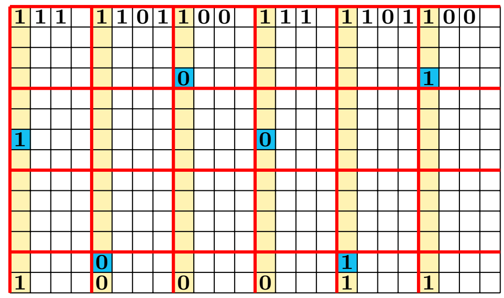
Proof. The recognition of PROPER can be split in two conditions:

1. The condition on the size (m, n) of the picture: n must be $2\lfloor m/\lceil \log m\rceil\rfloor \lceil \log m\rceil$.
2. The requirement inherited from LNE: the k-th and $(n + k)$-th columns must differ for all $k = 1 + r\lceil \log m\rceil$ when r ranges from 0 to $\lfloor m/\lceil \log m\rceil\rfloor - 1$.

To check the first condition, the tiling system makes use of a binary counter to evaluate $\lceil \log m \rceil$ and cuts the picture in squares of side length $\lceil \log m \rceil$. In this way, it can verify that n is a multiple of $\lceil \log m \rceil$ and $n/\lceil \log m \rceil = 2 \times \lfloor m/\lceil \log m \rceil \rfloor$.



To check the second condition, observe first that the above tiling system allows to identify every column $k = 1 + r\lceil \log m \rceil$ with $0 \le r < 2\lfloor m/\lceil \log m \rceil \rfloor$. Then the tiling system guesses one pixel in each of the identified column, computes the binary value of its row index and also transmits the pixel value to the bottom. Finally the tiling system verifies that the two picture halves hold the same sequence for the row indexes and two complement sequences for the bit values.



$\square$

Next let us check that the complement of `PROPER` is also in `REC`

Proposition 4. `PROPER` is in `co-REC`.

Proof. A picture of size (m, n) in $\overline{\texttt{PROPER}}$ either verifies that $n \neq 2\lfloor m/\lceil \log m \rceil \rfloor \lceil \log m \rceil$ or holds two identical columns of indexes $1 + r\lceil \log m \rceil$ and $n + 1 + r\lceil \log m \rceil$ for some $r < \lfloor m/\lceil \log m \rceil \rfloor$. As seen in the previous proposition, a tiling system can determine whether the picture size (m, n) satisfies $n = 2\lfloor m/\lceil \log m \rceil \rfloor \lceil \log m \rceil$, or not. Then, we can easily adapt the tiling system depicted in Example 1 to check whether there exists two identical columns of indexes $1 + r\lceil \log m \rceil$ and $n + 1 + r\lceil \log m \rceil$ for some $r < \lfloor m/\lceil \log m \rceil \rfloor$. $\square$

It remains to ascertain that `PROPER` is neither in `UREC` nor in `co-UREC`.

Proposition 5. `PROPER` is not in `UREC` $\cup$ `co-UREC`

Proof. The communication complexities of `PROPER` are essentially those of `LNE`. Indeed, adding columns to a picture modifies its width but leaves its height unchanged. So the amount of information that must be exchanged between the two halves of a `LNE` picture instance or its `PROPER` counterpart is the same. In particular, $\mathcal{UN}_1(f^{\texttt{PROPER}}_{m,\lfloor m/\lceil \log m \rceil \rfloor \lceil \log m \rceil}) = \mathcal{UN}_1(f^{\texttt{LNE}}_{m,m/\lceil \log m \rceil \rfloor})$. From Fact 4, it follows that $\mathcal{UN}_1(f^{\texttt{PROPER}}_{m,\lfloor m/\lceil \log m \rceil \rfloor \lceil \log m \rceil})$ is in $\Omega(m^2/\log m)$. In the same way, $\mathcal{UN}_0(f^{\texttt{PROPER}}_{m,\lfloor m/\lceil \log m \rceil \rfloor \lceil \log m \rceil})$ is in $\Omega(m^2/\log m)$. This, according to Corollary 1.ii, implies that neither `PROPER` nor $\overline{\texttt{PROPER}}$ is in `UREC`. $\square$

The above propositions are summarized in

Theorem 2. `PROPER` $\in$ (`REC` $\cap$ `co-REC`) $\setminus$ (`UREC` $\cup$ `co-UREC`)

6 Conclusion

The communication complexity framework is a simple and well-defined formalism for proving lower bounds. We have seen how it applies naturally for tiling systems and allows us to obtain new proofs of known limits. We could also use other types of protocols to further investigate the different subclasses of `REC`. For instance, one-round deterministic protocol seems well suited for deterministic variants of `REC`.

Another benefit is the existence of a large body of results on communication complexity. Thanks to these results, we were able to define a witness language `PROPER` that differentiates `REC` $\cap$ `co-REC` and `UREC` $\cup$ `co-UREC`, by calibrating its communication complexities.

Now, is there any similar strategy for the difficult question whether `UREC` is closed under complement or not? It seems less easy because the ranks of a matrix and its complement are of same order. So the rank lower bound approach can not be simply applied.

References

[1] Marcella Anselmo, Dora Giammarresi, Maria Madonia, and Antonio Restivo. Unambiguous recognizable two-dimensional languages. *RAIRO - Theoretical Informatics and Applications*, 40(2):277–293, 2006.

[2] Marcella Anselmo and Maria Madonia. Classes of two-dimensional languages and recognizability conditions. *RAIRO - Theoretical Informatics and Applications*, 44(4):471–488, 2010.

[3] Marcella Anselmo and Maria Madonia. A stronger recognizability condition for two-dimensional languages. *Discrete Mathematics & Theoretical Computer Science*, 15(2):139–156, 2013.

[4] Dora Giammarresi and Antonio Restivo. Recognizable picture languages. *IJPRAI*, 6(2&3):241–256, 1992.

[5] Dora Giammarresi and Antonio Restivo. Matrix-based complexity functions and recognizable picture languages. In *Logic and Automata: History and Perspectives [in Honor of Wolfgang Thomas].*, pages 307–330, 2008.

[6] Dora Giammarresi, Antonio Restivo, Sebastian Seibert, and Wolfgang Thomas. Monadic second-order logic over rectangular pictures and recognizability by tiling systems. *Information and Computation*, 125(1):32–45, 1996.

[7] Etienne Grandjean and Frédéric Olive. A logical approach to locality in pictures languages. *Journal of Computer and System Sciences*, 82(6):959–1006, 2016.

[8] Pierre Guillon and Emmanuel Jeandel. Infinite Communication Complexity. arXiv preprint, September 2014.

[9] Juraj Hromkovic, Sebastian Seibert, Juhani Karhumäki, Hartmut Klauck, and Georg Schnitger. Communication complexity method for measuring nondeterminism in finite automata. *Information and Computation*, 172(2):202–217, 2002.

[10] Katsushi Inoue and Itsuo Takanami. A characterization of recognizable picture languages. *IJPRAI*, 8(2):501–508, 1994.

[11] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[12] Kristian Lindgren, Cristopher Moore, and Mats Nordahl. Complexity of two-dimensional patterns. *Journal of Statistical Physics*, 91(5):909–951, Jun 1998.

[13] Oliver Matz. On piecewise testable, starfree, and recognizable picture languages. In Maurice Nivat, editor, *Foundations of Software Science and Computation Structure*, pages 203–210, 1998.

[14] Wolfgang Thomas, Sebastian Seibert, and Andreas Potthoff. Nondeterminism versus determinism of finite automata over directed acyclic graphs. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 1(2):285–298, 1994.

[15] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.