Communication Complexity Notes

Caleb Koch

November 2019

These were notes taken during an independent study course with Eshan Chattopadhyay in complexity theory during Fall 2019 at Cornell.

Contents

Purpose/Roadmap

The goal of this set of notes is threefold:

1. introduce communication complexity via a careful work through of [RY19] and [AB09]
2. highlight the connections between circuit complexity and communication complexity – particularly via Karchmer-Wigderson Games and communication theoretic lower bounds on the monotone circuit (depth) complexity of MATCH
3. survey and work through the basics of information complexity including the relatively recent result of [GMWW14] applying communication/information theoretic techniques to the KRW conjecture.

Preliminaries

We use the communication model presented in [RY19, Chapter 1]. In particular, Alice and Bob are two agents trying to compute $f : X \times Y \to \{0,1\}$ on some input (x, y) where Alice receives x and Bob receives y by exchanging as few bits as possible. They agree beforehand on a communication model which is a binary tree such that each node is labeled with either Alice or Bob, indicating which agent "owns" the node. If Alice owns the node v then v is equipped with a function $f_v : X \to \{0,1\}$ indicating which branch (child node) to take on input x – if $f_v(x) = 1$ then we take the high (right) branch and otherwise we take the lo (left) branch. Likewise if Bob owns the node v then there is an associated map $f_v : Y \to \{0,1\}$ indicating which branch to take. We denote this tree, specifying a protocol, as π.

The outcome of a protocol is a leaf in the tree which is labeled with 1 or 0 to denote the output of $f(x, y)$. The length of the protocol, $\|\pi\|$, is the depth of the protocol tree. For a function $f : X \times Y \to \{0,1\}$ the *deterministic communication complexity*, $\mathsf{P}^{\mathsf{cc}}(f)$ is the length of the protocol π that minimizes $\|\pi\|$, i.e. the depth of the shortest depth protocol tree for computing f.

Lower bound techniques

In this section, we introduce several lower bound techniques in communication complexity.

Rectangles

Rectangles are the "building blocks of communication protocols" [RY19]. A rectangle is simply a subset $A \times B \subseteq X \times Y$. Equivalently a subset $R \subseteq X \times Y$ is a rectangle iff for all $(x, y), (x', y') \in R$ we have $(x, y'), (x', y) \in R$ as well. In particular, if R satisfies the above property for some fixed $(x, y) \in R$ then we have

$$R = \{x' \in X : (x', y) \in R\} \times \{y' \in Y : (x, y') \in R\}.$$

Every protocol can be viewed as a set of rectangles where each path in the protocol tree yields a sequence of nested rectangles. In particular, we can associate with each vertex v, a rectangle $R_v \subseteq X \times Y$ that indicates exactly those inputs (x, y) that would pass through v during execution. A *monochromatic* rectangle is a rectangle $R \subseteq X \times Y$ such that for all $(x, y) \in X \times Y$ we have $f(x, y) = b$ for some $b \in \{0, 1\}$.[1]

For a leaf ℓ in a protocol, the associated rectangle R_ℓ consisting of those inputs $(x, y) \in X \times Y$ that would lead to ℓ during execution must necessarily be monochromatic since the label of the leaf determines the value of $f(x, y)$. Since every input (x, y) leads to a unique leaf in the protocol, the rectangles $\{R_\ell : \ell \text{ is a leaf}\}$ partitions $X \times Y$ into monochromatic rectangles. For a binary tree of depth $\|\pi\|$, the maximum number of leaves is bounded above by $2^{\|\pi\|}$ and so the size of the partition induced by π is also bounded above by $2^{\|\pi\|}$. In particular, let $\chi_1(f)$ denote the least number of 1-monochromatic rectangles needed to partition $\{(x, y) \in X \times Y : f(x, y) = 1\}$ and define $\chi_0(f)$ likewise, then the minimum number of monochromatic rectangles needed to partition $X \times Y$, $\chi(f) = \chi_1(f) + \chi_0(f)$ satisfies

$$\chi(f) \leq 2^{\mathsf{P}^{\mathsf{cc}}(f)}.$$

This is more often stated as $\log \chi(f) \leq \mathsf{P}^{\mathsf{cc}}(f)$.

because the the communication protocol π realizing $\mathsf{P}^{\mathsf{cc}}(f)$ induces a partition of $X \times Y$ into monochromatic rectangles of size at most $2^{\|\pi\|} = 2^{\mathsf{P}^{\mathsf{cc}}(f)}$.

Interestingly, we also have a converse result stating that $\mathsf{P}^{\mathsf{cc}}(f)$ is $O(\log^2(\chi(f)))$. Namely we have the following proposition.

Proposition 0.1. *Let $\mathcal{R}$ be a partition of $X \times Y$ into n rectangles. Then there is a protocol π that on input (x, y) allows Alice and Bob to determine which $R \in \mathcal{R}$ contains (x, y) and $\|\pi\| \leq \log^2 n + \log n$.*

Proof. The protocol works by alternating between Bob and Alice who announces during a turn a rectangle in $\mathcal{R}$ that allows the other agent to eliminate at least half of the feasible rectangles. Thus the total number of rounds is $\log n$ and each round involves communicating a $\log n$ bitstring so the total length of the protocol is $O(\log^2 n)$.

Specifically, let $R = A \times B$ and $R' = A' \times B'$ be two rectangles, then R intersects R' *horizontally* if $A \cap A' \neq \varnothing$ and R intersects R' *vertically* if $B \cap B' \neq \varnothing$. Furthermore, for an input (x, y), we say R is

[1] Similarly, a rectangle is 1-monochromatic if $b = 1$ and 0-chromatic otherwise.

horizontally good if $x \in A$ and R horizontally intersects at most $n/2$ rectangles in $\mathcal{R}$ and likewise we say R is *vertically good* if $y \in B$ and it intersects at most $n/2$ rectangles in $\mathcal{R}$.

We observe that the unique rectangle $R \in \mathcal{R}$ with $(x, y) \in R$ is either vertically or horizontally good. To see this, note that any $R' \in \mathcal{R}$ with $R' \neq R$ cannot intersect R both horizontally and vertically since then (x, y) would be in R', contradicting $R' \cap R = \varnothing$. We can thus stick R' into one of three sets: $\mathcal{R}_h$ if it horizontally intersects R, $\mathcal{R}_v$ if it vertically intersects R, or $\mathcal{R}_n$ if neither hold. Note then $\mathcal{R}_h, \mathcal{R}_n, \mathcal{R}_v$ partition $\mathcal{R} \setminus \{R\}$ so $|\mathcal{R}_h| + |\mathcal{R}_n| + |\mathcal{R}_v| = n - 1$ which shows that either $|\mathcal{R}_h| \leq n/2$ or $|\mathcal{R}_v| \leq n/2$.[2] It follows that R is either horizontally or vertically good.

Thus, the protocol can proceed in the following fashion. Let $P \subseteq \mathcal{R}$ denote the set of potential rectangles containing (x, y). Initially we have $P = \mathcal{R}$ and P is gradually pruned as Alice and Bob communicate. In a step of the protocol, Alice checks if P contains a vertically good rectangle. If yes, she sends Bob $\log|\mathcal{R}| = \log n$ bits identifying which rectangle is vertically good. Otherwise, Alice sends 1 bit indicating that she doesn't have such a rectangle, in which case Bob checks to see if P contains a horizontally good rectangle and sends Alice the identifier for that rectangle. Thus we require $\log n + 1$ bits of communication to account for the extra bit Alice may need to use if she doesn't end up having a vertically good rectangle. Every step is possible since R is always in P and is either horizontally good or vertically good. Once the "good" rectangle has been communicated both agents can cut P in half (at least) by removing all the rectangles that do not properly intersect the specified rectangle. That is, if the communicated rectangle is horizontally good (which means it was communicated by Bob) then the agents remove from P all the rectangles which do not horizontally intersect the communicated rectangle.

Since each step cuts P by at least half, there are at most $\log n$ rounds of communication. And each round requires at most $\log n + 1$ bits of communication so that total number of bits communicated (which is the length of the protocol) is $\log n(\log n + 1) = O(\log^2 n)$. □

Corollary 0.2. *We have $\mathsf{P}^{\mathsf{cc}}(f)$ is $O(\log^2 \chi(f))$.*

Proof. The smallest partition of $X \times Y$ into monochromatic rectangles has size $\chi(f)$. By Proposition 0.1, there is a protocol of length $O(\log^2 \chi(f))$ that given input (x, y) allows Alice and Bob to determine which rectangle contains (x, y). Since each rectangle is monochromatic, this allows Alice and Bob to compute $f(x, y)$ using $O(\log^2 \chi(f))$ bits of communication. It follows that $\mathsf{P}^{\mathsf{cc}}(f)$ is $O(\log^2 \chi(f))$. □

[2] If both were $> n/2$ we would have $|\mathcal{R} \setminus \{R\}| > n$, a contradiction.

Corollary 0.3. *For all $f : X \times Y \to \{0,1\}$, we have*

$$\log \chi(f) \leq \mathsf{P}^{\mathsf{cc}}(f) \leq \log^2 \chi(f) + \log \chi(f).$$

Proof. The result follows from Proposition 0.1 and the remark before the proposition that $\chi(f) \leq 2^{\mathsf{P}^{\mathsf{cc}}(f)}$. □

Many other lower bound techniques for communication complexity are established via some relationship to partition number. We shall see some examples below. But first, let us apply 0.3 to derive some communication complexity lower bounds.

Disjointness

Let DISJ : $2^{[n]} \times 2^{[n]} \to \{0,1\}$ be the set disjointness function given by

$$\mathrm{DISJ}(A,B) = \begin{cases} 1 & A \cap B = \varnothing \\ 0 & \text{otherwise} \end{cases}.$$

We have $\log \chi(\mathrm{DISJ}) \leq \mathsf{P}^{\mathsf{cc}}(\mathrm{DISJ})$ by Corollary 0.3 and hence it is sufficient lower bound $\chi(\mathrm{DISJ})$ to get a lower bound on the communication complexity of set disjointness. Intuitively, if we can show that all rectangles must be small then that means the partition size has to be large. More precisely we can show that every 1-monochromatic rectangle of DISJ cannot be bigger than 2^n.

Proposition 0.4. *Let R be a 1-monochromatic rectangle of* DISJ*. Then $|R| \leq 2^n$.*

Proof. Let $R = A \times B \subseteq 2^{[n]} \times 2^{[n]}$. Define

$$X' = \bigcup_{X \in A} X \qquad Y' = \bigcup_{Y \in A} Y.$$

Then we observe that as R is 1-monochromatic we have $\mathrm{DISJ}(X', Y') = 1$ and thus as $X', Y' \subseteq [n]$ are disjoint we must have $|X'| + |Y'| \leq n$. We also have $A \subseteq 2^{X'}$ and so $|A| \leq 2^{|X'|}$. Likewise $|B| \leq 2^{|Y'|}$. It follows that

$$\begin{aligned} |R| &\leq |A \times B| \\ &= |A| \cdot |B| \\ &\leq 2^{|X'|+|Y'|} \\ &\leq 2^n. \end{aligned}$$

□

How many subsets $A, B \subseteq [n]$ satisfy $\mathrm{DISJ}(A,B) = 1$? There are exactly 3^n since for any $x \in [n]$ we can choose to put it in A, or B, or neither and there are n such x's each with three distinct placements showing that there are 3^n ways to pick A, B with $A \cap B = \varnothing$.

Thus the task we are presented with is to partition a space of 3^n elements using partitions whose individual size is at most 2^n. Hence in particular the number of partitions is at least $3^n/2^n$. It follows that $\chi_1(\mathrm{DISJ}) \geq 3^n/2^n$. We can thus prove the following proposition.

Proposition 0.5. $n \log(3/2) \leq \mathsf{P}^{cc}(\mathrm{DISJ})$

Proof. We write

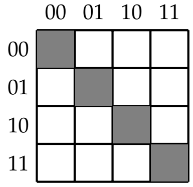
$$\begin{aligned}(3/2)^n &\leq \chi_1(\mathrm{DISJ}) \\ &\leq \chi_1(\mathrm{DISJ}) + \chi_0(\mathrm{DISJ}) \\ &= \chi(\mathrm{DISJ}) \\ n\log(3/2) &\leq \log \chi(\mathrm{DISJ}) \\ &\leq \mathsf{P}^{\mathsf{cc}}(\mathrm{DISJ}).\end{aligned}$$

□

We can give a similar analysis of EQ : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, the equality function, given by

$$\mathrm{EQ}(x,y) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}.$$

Below is a visualisation of EQ for $n = 2$



	00	01	10	11
00	■	□	□	□
01	□	■	□	□
10	□	□	■	□
11	□	□	□	■

We observe that any 1-monochromatic rectangle, R, satisfies $|R| = 1$. This is because if $(x,y), (x',y') \in R$ with $\mathrm{EQ}(x,y) = \mathrm{EQ}(x',y') = 1$ then either $\mathrm{EQ}(x,y') = 1$ and $\mathrm{EQ}(x',y) = 1$ showing that $x = x' = y = y'$.

Note that in this case there are exactly 2^n string pairs $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ satisfying $\mathrm{EQ}(x,y) = 1$ (in particular all pairs (x,x) for $x \in \{0,1\}^n$). Thus as every 1-monochromatic rectangle has size 1, any partition has size exactly 2^n. In other words, $\chi_1(\mathrm{EQ}) = 2^n$. Hence we can prove an analogous lower bound as the one for set disjointness.

Proposition 0.6. $n \leq \mathsf{P}^{cc}(\mathrm{EQ})$.

Proof. We write

$$\begin{aligned}2^n &\leq \chi_1(\mathrm{EQ}) \\ &\leq \chi(\mathrm{EQ}) \\ n &\leq \log \chi(\mathrm{EQ}) \\ &\leq \mathsf{P}^{\mathsf{cc}}(\mathrm{EQ}).\end{aligned}$$

□

Matrices

We have left X, Y unspecified in the function $f : X \times Y \to \{0,1\}$. However, often these can be taken to be $\{0,1\}^n$ (in particular they are finite sets so every element can be uniquely identified by $\log |X|$ bits or $\log |Y|$ bits). In doing so, we can form a communication matrix indexed by bitstrings whose entries are 0 or 1 depending on the evaluation of the function on the corresponding input. For example, the communication matrix for the equality function, EQ, is always the identity matrix. Let $M(f)$ be the communication matrix for f. In particular $M(f)_{x,y} = f(x,y)$. Then the following relates the rank of the matrix with the partition number of f.

Proposition 0.7. *For every f, we have $\chi_1(f) \geq \operatorname{rank}(M(f))$.*

Proof. Let $\mathcal{R}$ be the minimum partition of $X \times Y$ into 1-monochromatic rectangles. The intuition is that each element of $\mathcal{R}$ can contribute at most 1 to the overall rank of $M(f)$. For each $R \in \mathcal{R}$ we construct a matrix M_R such that $(M_R)_{x,y} = 1$ iff $(x,y) \in R$ and otherwise $(M_R)_{x,y} = 0$. Note that $\operatorname{rank}(M_R) = 1$ since M_R will have a single block of 1s surrounded by $0s$. Also the R are disjoint so

$$M = \sum_{R \in \mathcal{R}} M_R$$

evaluates to a matrix with $|\mathcal{R}|$ blocks of 1 and the rest of the entries 0. Hence in particular $\operatorname{rank}(M) \leq |\mathcal{R}|$ (some blocks may horizontally/vertically intersect making the rank less than $\mathcal{R}$). Note that $M = M(f)$ since the 0-monochromatic rectangles will consist of 0 blocks. Hence we have $\operatorname{rank}(M(f)) \leq |\mathcal{R}| = \chi_1(f)$. □

The infamous *log-rank conjecture* states that this bound is essentially tight.

Conjecture 0.8. *There is a universal constant c such that for every f we have*

$$c \log^c \operatorname{rank}(M(f)) \geq \mathsf{P}^{cc}(f)$$

Relation to Circuits

Karchmer-Wigderson Games

Communication complexity relates nicely to circuit complexity via Karchmer-Wigderson games. Suppose we have a function $f : \{0,1\}^n \to \{0,1\}$. Alice gets some $x \in \{0,1\}^n$ such that $f(x) = 0$ whereas Bob gets some $y \in \{0,1\}^n$ such that $f(y) = 1$. The goal is to find $i \in [n]$ with $x_i \neq y_i$. In the case where monotone, they seek an index such that $x_i < y_i$. Quite surprisingly, circuit-depth is equivalent to communication complexity.

Proposition 0.9. *Let $C : \{0,1\}^n \to \{0,1\}$ be a Boolean circuit of depth d. Then $\mathsf{P}^{cc}(C) \leq d$. Likewise if C is monotone, then C yields a protocol solving the monotone version of the game.*

Proof. The proof is by induction starting at the topmost (i.e. the output gate).[3] Suppose the topmost gate of C is an AND gate so that $C = g \wedge h$ for some $g : \{0,1\}^n \to \{0,1\}$ and $h : \{0,1\}^n \to \{0,1\}$. If $x \in \{0,1\}^n$ is such that $C(y) = 1$ then we have $g(y) = h(y) = 1$. Otherwise if $C(x) = 0$ then we must have either $g(x) = 0$ or $h(x) = 0$. We thus have Bob communicate a bit indicating whether $g(x) = 0$ or $h(x) = 0$ so that Alice and Bob both pick the subcircuit C' (corresponding to either g or h) satisfying $C'(x) \neq C'(y)$. Likewise in the OR case we can write $C = g \vee h$. Here, we have either $g(y) = 1$ or $h(y) = 1$ but $g(x) = h(x) = 0$. Thus Alice communicates a bit indicating which subcircuit C' satisfies $C'(y) \neq C'(x)$. If the gate is a negation gate then we can write $C = \neg g$ in which case both Alice and Bob can recurse on g without communicating anything. We repeat this procedure until we get to an input gate $C' = x_i$ satisfying $C'(x) \neq C'(y)$ which implies $x_i \neq y_i$.

In the monotone case, instead of picking C' with just $C'(x) \neq C'(y)$ we can enforce the stronger condition that $C'(x) < C'(y)$ since there are no negations. Thus, at the input gate level we get that $x_i < y_i$.

In this protocol we decrease the depth of the circuit at each step by 1 until we've traced a path from the output node to an input node. Thus the length of any protocol is d where d is the depth of the circuit. □

Proposition 0.10. *Suppose that a function $f : \{0,1\}^n \to \{0,1\}$ has a Karchmer-Wigderson game using d bits of communication. Then, there is a depth-d Boolean circuit computing f. Likewise if the game solves the monotone case then the corresponding circuit is monotone.*

Proof. Our proof is by induction on d. We prove a slightly stronger claim which is: given nonempty $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$ if there is a protocol of length d such that Alice on input $x \in A$ and Bob on input $y \in B$ can find an index i such that $x_i \neq y_i$ then there is a circuit, C, of depth d such that $C(A) = 0$ and $C(B) = 1$.

For the base case $d = 0$. In this case, the protocol always outputs some fixed i with $x_i \neq y_i$ for all $x \in A$ and for all $y \in B$. Hence it must be the case that either

$$x_i = 0 \qquad y_i = 1$$

for all $x \in A$ and $y \in B$ in which case we take the circuit consisting of the single output gate x_i or we have

$$x_i = 1 \qquad y_i = 0$$

[3] The idea here is a we start at the output gate and trace a path backwards through the circuit to an input gate.

in which case we take our circuit to consist of the single output gate $\neg x_i$.

Now suppose $d > 0$ and Alice is the first to communicate. We can write $A = A_0 \cup A_1$ such that A_0 consists of those inputs that lead her to communicate a 0 while A_1 consists of those that lead her to communicate a 1. If either are empty then we can reduce the length of the communication protocol by 1 simply by ignoring Alice's first message. The result will then follow inductively. Otherwise, both sets are nonempty inductively[4] we get two subcircuits C_0 and C_1 satisfying

$$C_0(A_0) = 0 \qquad C_0(B) = 1$$

$$C_1(A_1) = 0 \qquad C_1(B) = 1.$$

We build the circuit $C = C_0 \wedge C_1$. We then have $C(B) = C_0(B) \wedge C_1(B) = 1$ and

$$C(x) = \begin{cases} C_0(A_0) \wedge C_1(A_0) = 0 & \text{if } x \in A_0 \\ C_0(A_1) \wedge C_1(A_1) = 0 & \text{if } x \in A_1 \end{cases}$$

showing that $C(A) = 0$. In the case where Bob sends the first bit □

We also present a similar, yet perhaps simpler proof based on protocol trees. Note that we can modify our definition of protocol trees to work with arbitrary functions $F : \mathcal{X} \times \mathcal{Y} \to S$ for any finite set S by allowing leaf labels to be any element in S as opposed to just 0 or 1.

Lemma 0.11. *Let $f : \{0,1\}^n \to \{0,1\}$ and let $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$ be nonempty. Suppose there is a protocol π computing $F : A \times B \to [n]$ such that if $F(x,y) = i$ then $x_i \neq y_i$. Then there is a circuit of depth $\|\pi\|$ such that for all $x \in A$ we have $C(x) = 0$ and for all $y \in B$ we have $C(y) = 1$. Moreover, ignoring vertex labels we have $C = \pi$.*

Proof. We construct the following circuit inductively from the protocol tree π.

- if $\pi = \ell$ is a leaf for $\ell \in [n]$ we have either for all $x \in A$ and $y \in B$ that

  $$x_\ell = 0 \qquad y_i = \ell$$

  or

  $$x_\ell = 1 \qquad y_\ell = 0.$$

  In the first case we define $C := x_\ell$ as the single input gate for the ℓth input and in the second case we let $C := \neg x_\ell$ as the negation of the ℓth input.

[4] Note we can applying the inductive hypothesis since the protocol corresponding to A_0/A_1 is one bit shorter.

- If $\pi = (v, H, L)$ where v is a vertex, H is the right subtree and L is the left subtree, then we inductively have circuits C_H and C_L for the sub-protocols and then define

$$C := \begin{cases} C_H \wedge C_L & \text{if } \textsc{owner}(v) = \text{Alice} \\ C_H \vee C_L & \text{if } \textsc{owner}(v) = \text{Bob}. \end{cases}$$

Note that it is immediate that the depth of C is $\|\pi\|$. Thus it remains to show that $C(x) = 0$ for all $x \in A$ and $C(y) = 1$ for all $y \in B$.

We prove this statement by induction on $\|\pi\|$. If $\|\pi\| = 0$ then $\pi = \ell$ is a leaf for some $\ell \in [n]$. It follows that $F(x, y) = \ell$ for all $(x, y) \in A \times B$ and so $x_\ell \neq y_\ell$ for all $(x, y) \in A \times B$. Suppose $x_\ell = 0$ for all $x \in A$. Then our circuit is $C = x_\ell$ and we have $C(x) = 0$ for all $x \in A$ and $C(y) = 1$ for all $y \in B$. The other case when $x_\ell = 1$ for all $x \in A$ is symmetric.

Now suppose we have a protocol π with $\|\pi\| > 0$. Let $\pi = (v, H, L)$ for a vertex v and sub-protocols H and L. Suppose $\textsc{owner}(v) =$ Alice. Let $f_v : A \to \{0, 1\}$ be the function associated with v which encapsulates Alice's communication protocol at v. Then we partition A into $A_0 = f_v^{-1}(0)$ and $A_1 = f_v^{-1}(1)$. If one of A_0 or A_1 is empty then we can just replace π with the subprotocol corresponding to the nonempty set of inputs to form an equivalent protocol of depth $\|\pi\| - 1$ and then apply the inductive hypothesis. Otherwise, suppose both A_0 and A_1 are nonempty. Then we observe H is a protocol for the function $F : A_1 \times B \to [n]$ and L is a protocol for the function $F : A_0 \times B \to [n]$. Thus inductively we have a circuit C_L corresponding to L and a circuit C_H corresponding to H. Moreover, these circuits both have depth $\leq \|\pi\|$ and satisfy

$$C_H(y) = C_L(y) = 1$$

for all $y \in B$ and

$$C_H(x) = 0 \qquad C_L(x') = 0$$

for all $x \in A_1$ and $x \in A_0$. Thus we observe that $C = C_H \wedge C_L$ satisfies $C(x) = C_H(x) \wedge C_L(x) = 0$ for all $x \in A = A_1 \cup A_0$ and $C(y) = C_H(y) \wedge C_L(y) = 1 \wedge 1 = 1$ for all $y \in B$. Thus C has the desired properties. The case where $\textsc{owner}(v) =$ Bob is symmetrical. □

Note that Lemma 0.11 shows in particular that the underlying tree of the communication protocol for the KW game and the circuit computing the function corresponding to the KW game are precisely the same tree (ignoring vertex labels). This fact is also proven in [GMWW14, Theorem 2.12].

The monotone version of the theorem states analogously that if $f : \{0,1\}^n \to \{0,1\}$ is a monotone function and f has a monotone KW game using d bits of communication then there is a depth d monotone circuit computing f.

One nice application of KW games is a circuit lower bound for $\mathrm{MATCH}(G)$ which is a decision version of finding matching on a graph. Formally, a *matching* on a graph $G = (V, E)$ is a collection of edges, $S \subseteq E$, such that no two edges in E share a common vertex. Let G be a graph on n vertices so that G can be encoded as a bitstring $G \in \{0,1\}^{\binom{n}{2}}$ where the ith bit is a 1 if the ith edge is present in G or not (as there are at most $\binom{n}{2}$ edges in G). Given such a G we define

$$\mathrm{MATCH}(G) := \begin{cases} 1 & \text{if } G \text{ has a matching of size at least } \lfloor n/3+1 \rfloor \\ 0 & \text{otherwise} \end{cases}.$$

Note that MATCH is monotone since adding more edges to G will not destroy a matching that already exists. Thus, we can study the monotone circuit complexity of MATCH.

We leverage a lemma below whose proof is fairly nontrivial (see [RY19, Chapter 6])

Lemma 0.12. *Any randomized protocol that computes* DISJ *with error* $1/2 - \varepsilon$ *must have communication* $\Omega(\varepsilon^2 n)$.

Theorem 0.13 ([RY19],[RW92])**.** *If C is a monotone Boolean circuit computing* MATCH *then the depth of C is* $\Omega(n)$.

Proof. The strategy will be to show that any communication protocol for the monotone KW game for MATCH can be turned into a randomized protocol for DISJ and thus we can leverage the lower bound from Lemma 0.12 to get a lower bound on MATCH.

In particular let $\mathrm{DISJ} : 2^{[m]} \times 2^{[m]} \to \{0,1\}$ be the set disjointness function and let (X, Y) be an input. Alice and Bob each turn their respective sets into graphs G_X and G_Y respectively, using shared randomness. Each graph is on the vertex set $[3m+2]$. The agents start by randomly permuting the vertices. The constructions then are given as follows.

- *Alice's construction of* G_X: The edge set of G_X is given by

  $$\{\{3i, 3i-1\} : i \in X\} \cup \{\{3i, 3i-2\} : i \notin X\} \cup \{3m+1, 3m+2\}$$

  for all $i \in [m]$.

- *Bob's construction of* G_Y: The edge set of G_Y is given by

  $$\{\{3i-2, j\} : i \in Y\ \forall j \in [3m+1]\} \cup \{\{3i, j\} : i \notin Y\ \forall j \in [3m+1]\}$$

  where again i is taken as $i \in [m]$.

Now consider $\mathrm{MATCH}(G_Y)$. This function quantity is 1 if G_Y has a matching of size at least $\lfloor(3m+1)/3+1\rfloor = m+1$. Consider the vertex set $\{3i-2 : i \in Y, i \in [m]\} \cup \{3i-1 : i \notin Y, i \in [m]\}$ which has size $[m]$. By construction every edge in G_Y is connected to a vertex in this set. Hence G_Y does not contain a matching of size $\geq m+1$ and $\mathrm{MATCH}(G_Y) = 0$. Conversely we observe that all of the edges in G_X form a matching and since there are precisely $m+1$ edges in G_X we have $\mathrm{MATCH}(G_X) = 1$.

It follows that G_X and G_Y are valid inputs in the monotone KW game for MATCH. Recall that in this case the players must communicate to determine which edge e is present in G_X but not in G_Y. If X and Y are disjoint then the only edge in G_X which is not in G_Y is $\{3m+1, 3m+2\}$. Conversely if X and Y intersect and their intersection has size $k > 0$ then G_X will contain exactly $k+1$ edges that G_Y doesn't contain: namely vertices of the form $\{3i, 3i-1\}$ for $i \in X \cap Y$ and the vertex $\{3m+1, 3m+2\}$.

In this case, as the vertices are permuted uniformly at random each of the $k+1$ edges are equally likely to result from the protocol for the KW game. Thus, the players can run the KW game protocol and output "disjoint" if the edge is $\{3m+1, 3m+2\}$ and "not disjoint" otherwise. The probability that $\{3m+1, 3m+2\}$ is output when the sets are not disjoint is $1/(k+1)$ which is at most $1/2$. Hence the players can repeat a constant number of times to achieve an answer with error at most $1/3$.

□

Information complexity

Preliminaries

We introduce some requisite notation and recall various concepts from information theory. The main goal of this section will be to overview the KRW conjecture and how information complexity can be used to make progress on it.

We denote the entropy of a random variable $\mathbf{X}$ as $H(\mathbf{X})$. Similarly the conditional entropy of $\mathbf{X}$ given $\mathbf{Y}$ is defined as

$$H(\mathbf{X} \mid \mathbf{Y}) := \mathbb{E}_{\mathbf{Y}}[H(\mathbf{X} \mid \mathbf{Y} = y)].$$

A key ingredient in our proofs is that $0 \leq H(\mathbf{X}) \leq \log|\chi|$ where χ is the support of $\mathbf{X}$ (so e.g. $\mathbf{X} : \Omega \to \chi$ and $P(X = x) \neq 0$ for all $x \in \chi$) and the upper bound $\log|\chi|$ is attained when $\mathbf{X}$ is uniform.

We denote $I(\mathbf{X} : \mathbf{Y})$ as the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ defined by

$$I(\mathbf{X} : \mathbf{Y}) := H(\mathbf{X}) - H(\mathbf{X} \mid \mathbf{Y}).$$

Similarly we write $I(\mathbf{X} : \mathbf{Y} \mid \mathbf{Z})$ to denote the mutual information of $\mathbf{X}$ and $\mathbf{Y}$ conditional on $\mathbf{Z}$. This is defined as

$$I(\mathbf{X} : \mathbf{Y} \mid \mathbf{Z}) := H(\mathbf{X} \mid \mathbf{Z}) - H(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}).$$

From these definitions one can derive the chain rule[5]

$$I(\mathbf{X}, \mathbf{Y} : \mathbf{Z}) = I(\mathbf{X} : \mathbf{Z}) + I(\mathbf{Y} : \mathbf{Z} \mid \mathbf{X})$$

which we also use below.

A basic result from these definitions is that if a variable is uniformly distributed, and conditioning on an event decreases the entropy much then the event must have small probability.

Proposition 0.14 (Fact 2.27 in [GMWW14]). *Let $\mathbf{X} : \Omega \to \chi$ be a uniformly distributed random variable and E, an event such that $H(\mathbf{X} \mid E) \leq \log|\chi| - t$ for some $t \geq 0$. Then we have*

$$\mathbb{P}[E] \leq \frac{1}{2^t}.$$

Proof. First we observe that for $x \in \chi$

$$\mathbb{P}[\mathbf{X} = x \mid E] = \frac{\mathbb{P}[\mathbf{X} = x \wedge E]}{\mathbb{P}[E]} \leq \frac{1/|\chi|}{\mathbb{P}[E]}$$

so that $1/\mathbb{P}[\mathbf{X} = x \mid E] \geq |\chi| \cdot \mathbb{P}[E]$. Now, we write

$$\begin{aligned} \log|\chi| - t &\geq H(\mathbf{X} \mid E) \\ &= \mathbb{E}_{x \sim \mathbf{X}|E}\left[\log \frac{1}{\mathbb{P}[\mathbf{X} = x \mid E]}\right] \\ &\geq \mathbb{E}_{x \sim \mathbf{X}|E}\left[\log(|\chi| \cdot \mathbb{P}[E])\right] \\ &= \log|\chi| + \log \mathbb{P}[E] \end{aligned}$$

which shows that $-t \geq \log \mathbb{P}[E]$ as desired. □

A good exposition of the information complexity basics can be found in [GMWW14, Section 2.6] but also in [BBCR13].

Let Π be a 2-party communication protocol viewed as a random variable of the inputs x, y which are sampled from a distribution μ (so $\Pi(x, y)$ is the communication transcript resulting from executing the protocol on inputs x, y). We denote the *external information cost* of Π over μ by

$$\mathsf{IC}_\mu(\Pi) := I(\Pi : \mathbf{x}, \mathbf{y})$$

which denotes how much information is revealed about the inputs when an external agent observes the transcript. In our case we only consider deterministic protocols and so $H(\Pi \mid \mathbf{x}, \mathbf{y}) = 0$ as Π is fixed once $\mathbf{x}, \mathbf{y}$ are known. Thus

$$\mathsf{IC}_\mu(\Pi) = I(\Pi : \mathbf{x}, \mathbf{y}) = H(\Pi) - H(\Pi \mid \mathbf{x}, \mathbf{y}) = H(\Pi).$$

[5] In its general form, the chain rule states that

$$I(\mathbf{X_1}, \ldots, \mathbf{X_n} : \mathbf{Y}) = \sum_{i=1}^{n} I(\mathbf{X_i} : \mathbf{Y} \mid \mathbf{X_1}, \ldots, \mathbf{X_{i-1}})$$

Similarly the *internal information cost* of Π over μ is denoted

$$\mathsf{IC}_\mu^{\mathsf{int}}(\Pi) = I(\Pi : \mathbf{x} \mid \mathbf{y}) = I(\Pi : \mathbf{y} \mid \mathbf{x})$$

which represents the amount of information Bob receives about Alice's input upon seeing the transcript and his input (or vice versa from Alice's perspective). A fundamental relationship between internal information cost, external information cost, and communication complexity is captured in the following proposition.

Proposition 0.15. *For a protocol Π and distribution μ, we have*

$$\mathsf{IC}_\mu^{int}(\Pi) \le \mathsf{IC}_\mu(\Pi) \le \mathsf{P}^{cc}(\Pi).$$

Note that in this case $\mathsf{P}^{\mathsf{cc}}(\Pi)$ denotes the depth of the communication protocol (viewed as a protocol tree) of Π.

Proof. We observe that viewed as a random variable, the support of Π is precisely the number of leaves in the protocol tree and hence the log of the support size is a lower bound on the depth, $\mathsf{P}^{\mathsf{cc}}(\Pi)$. This shows in particular, that

$$H(\Pi) \le \mathsf{P}^{\mathsf{cc}}(\Pi)$$

and so we have

$$\mathsf{IC}_\mu(\Pi) \le \mathsf{P}^{\mathsf{cc}}(\Pi).$$

Hence it remains to show that $\mathsf{IC}_\mu^{\mathsf{int}}(\Pi) \le \mathsf{IC}_\mu(\Pi)$. We prove this inequality by induction. Suppose Alice sends the first bit of a protocol on inputs $\mathbf{x}, \mathbf{y}$ and denote this first bit by Π_1. Since this bit is independent of Bob's input $\mathbf{y}$ we have $I(\Pi_1 : \mathbf{y} \mid \mathbf{x}) = 0$ and

$$\begin{aligned} I(\Pi_1 : \mathbf{x}, \mathbf{y}) &= I(\Pi_1 : \mathbf{y}) + I(\Pi_1 : \mathbf{x} \mid \mathbf{y}) \\ &\ge I(\Pi_1 : \mathbf{x} \mid \mathbf{y}) \\ &\ge I(\Pi_1 : \mathbf{x} \mid \mathbf{y}) + I(\Pi_1 : \mathbf{y} \mid \mathbf{x}). \end{aligned}$$

Inductively, we derive $\mathsf{IC}_\mu^{\mathsf{int}}(\Pi) \le \mathsf{IC}_\mu(\Pi)$. □

KRW Conjecture

Let $\mathrm{D}(f)$ denote the depth complexity of f which is the depth of the smallest depth NC circuit computing f. The KRW conjecture[6] [KRW95] is a statement about the depth complexity of the composition of two functions. In particular, given $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, their composition $g \diamond f : (\{0,1\}^n)^m \to \{0,1\}$ is defined by

$$g \diamond f(x_1, \ldots, x_m) = g(f(x_1), \ldots, f(x_m)).$$

The KRW conjecture essentially states that depth complexity of the composition of two functions f and g is roughly the sum of the depth complexities of f and g.

[6] Named after its proposers: Karchmer, Raz, and Wigderson

Conjecture 0.16 (KRW Conjecture)**.** $\mathrm{D}(g \diamond f) \approx \mathrm{D}(f) + \mathrm{D}(g)$

The precise meaning of $\approx$ is left ambiguous on purpose. There are various interpretations which would be of interest. For example, one variant of interest is a proof that $\mathrm{D}(g \diamond f) \geq \varepsilon \mathrm{D}(f) + \mathrm{D}(g)$ for any $\varepsilon > 0$.

A known corollary of the KRW conjecture is a separation of NC^1 from P.

Corollary 0.17 (Consequence of Conjecture 0.16)**.** $\mathrm{P} \not\subseteq \mathrm{NC}^1$.

There has been some progress on proving Conjecture 0.16 via information complexity.

In particular, the authors in [GMWW14] make progress towards the KRW conjecture by proving the following.

Theorem 0.18.

$$P^{cc}(g \diamond U_n) \geq \log \mathrm{L}(g) + n - O\left(1 + \frac{m}{n}\right) \log m$$

where $\mathrm{L}(g)$ denotes the formula size of g and U_n is *the universal relation* on n bits.

The proof of this result proceeds by fixing an arbitrary protocol Π and constructing a distribution μ such that the *external information cost*, $I_\mu(\Pi)$ is lower bounded by $\log \mathrm{L}(g) + n - O\left(1 + \frac{m}{n}\right) \log m$.

They also prove a stronger lower bound in the specific case where $g = \oplus_m$ is the parity on m input bits.

Theorem 0.19. $P^{cc}(\oplus_m \diamond U_n) \geq 2 \log m + n - O(\log \log m)$.

We prove a generalization of this theorem. Let $\mathcal{G}$ be the set of Boolean functions $g : \{0,1\}^m \to \{0,1\}$ such that there is a subset $\mathcal{X} \subseteq g^{-1}(0)$ and $Y \subseteq g^{-1}(1)$ and an index set $I \subseteq [m]$ where for every fixed $x \in \mathcal{X}$ we have

$$|\{y \in \mathcal{Y} : d(x,y) = 1, x_i \neq y_i, i \in I\}| = |I|$$

and the analogous statement holds for a fixed $y \in Y$ (and d here is the Hamming distance). Then we have the following theorem.

Theorem 0.20. *Let $g \in \mathcal{G}$ and I be the associated index set. Then*

$$P^{cc}(g \diamond U_n) \geq 2 \log |I| + n - O(\log \log |I|).$$

Define a distribution μ in the following way

- Choose a matrix $\mathbf{X} \in \{0,1\}^{m \times n}$ uniformly at random
- Choose a pair $(a,b) \in \mathcal{X} \times \mathcal{Y}$ uniformly at random. Note (x,y) differ on a unique coordinate $\mathbf{j}$ which is uniformly distributed over I.

The input $(\mathbf{X}, \mathbf{a})$ is given to Alice and $(\mathbf{X}, \mathbf{b})$ to Bob.

To start we write

$$\begin{aligned}\mathsf{P}^{\mathsf{cc}}(R_{g\diamond U_n}) &\geq \log \mathsf{L}(R_{g\diamond U_n})\\ &\geq \mathsf{IC}_\mu(\Pi)\\ &= I(\Pi : \mathbf{X}, \mathbf{a}, \mathbf{b})\\ &= I(\Pi : \mathbf{X}) + I(\Pi : \mathbf{a}, \mathbf{b} \mid \mathbf{X}).\end{aligned}$$

Thus it is sufficient to lower bound the last two quantities. We give these bounds in two lemmas.

Lemma 0.21. $I(\Pi : \mathbf{X}) \geq n$

Proof. See [GMWW14, Proof of lemma 4.3]. □

Lemma 0.22. $I(\Pi : \mathbf{a}, \mathbf{b} \mid \mathbf{X}) \geq 2\log|I| - O(t)$.

Proof. Fix a protocol Π.

We observe

$$\begin{aligned}I(\Pi : \mathbf{a}, \mathbf{b} \mid \mathbf{X}) &\geq \mathsf{IC}^{\mathrm{int}}_\mu(\Pi)\\ &= I(\Pi : \mathbf{a} \mid \mathbf{b}, \mathbf{X}) + I(\Pi : \mathbf{b} \mid \mathbf{a}, \mathbf{X})\\ &= I(\Pi : \mathbf{j} \mid \mathbf{b}, \mathbf{X}) + I(\Pi : \mathbf{j} \mid \mathbf{a}, \mathbf{X})\\ &= H(\mathbf{j} \mid \mathbf{b}, \mathbf{X}) - H(\mathbf{j} \mid \mathbf{b}, \mathbf{X}, \Pi) + H(\mathbf{j} \mid \mathbf{a}, \mathbf{X}) - H(\mathbf{j} \mid \mathbf{a}, \mathbf{X}, \Pi)\\ &= 2H(\mathbf{j}) - H(\mathbf{j} \mid \mathbf{b}, \mathbf{X}, \Pi) - H(\mathbf{j} \mid \mathbf{a}, \mathbf{X}, \Pi)\\ &= 2\log|I| - H(\mathbf{j} \mid \mathbf{b}, \mathbf{X}, \Pi) - H(\mathbf{j} \mid \mathbf{a}, \mathbf{X}, \Pi)\\ &\geq 2\log|I| - H(\mathbf{j} \mid \mathbf{b}, \Pi) - H(\mathbf{j} \mid \mathbf{a}, \Pi)\end{aligned}$$

Note given $\mathbf{b}$ the random variable $\mathbf{j}$ is independent of $\mathbf{b}$ (it's chosen independently at random from I) and of $\mathbf{X}$ and hence $H(\mathbf{j}) = H(\mathbf{j} \mid \mathbf{b}, \mathbf{X})$ and likewise for $\mathbf{a}$. We say a transcript is *good* if

$$H(\mathbf{j} \mid \mathbf{b}, \Pi) \leq t \qquad \text{and} \qquad H(\mathbf{j} \mid \mathbf{a}, \Pi) \leq t.$$

Otherwise a transcript is *bad*. Let $\mathcal{B}$ denote the set of *bad* transcripts.

Fix a bad transcript π such that wlog $H(\mathbf{j} \mid \mathbf{a}, \Pi = \pi) > t$. Fix some a in the support of $\mathbf{a}$. Let J be the support of the random variable $\mathbf{j} \mid \mathbf{a} = a, \Pi = \pi$. Let T be the set of matrices in the support of Π. We show that all matrices in T agree on every row j for $j \in J$. Fix some $j \in J$. Let $\mathcal{X}_\pi \times \mathcal{Y}_\pi$ be the rectangle associated with π. Fix some $j \in J$. Since j is in the support of $\mathbf{j} \mid \mathbf{a} = a, \Pi = \pi$ there must be some $b \in \mathcal{Y}$ and a matrix Y such that $((a, Y), (b, Y)) \in \mathcal{X}_\pi \times \mathcal{Y}_\pi$. Hence it is sufficient to show that Y agrees with every jth row in every $X \in T$. Note that if Bob gets (b, X) and Alice gets (a, Y) then the resulting transcript is π by definition. Also note that since the transcript π comes from the distribution μ the output must be $\perp$ and

Being in the support of π for a matrix means being in support of the leaf induced by π which means if there are inputs a, b such that when Alice is given (a, X) and Bob is given (b, X) the execution of the protocol makes the transcript π. A similar definition applies to strings.

hence because j is the *unique* coordinate on which a, b differ we must have $X_j = Y_j$. It follows that since all matrices agree on all rows in J we must have $|T| \leq 2^{(m-|J|)n}$. Note also by the definition of J, we have

$$\log |J| \geq H(\mathbf{j} \mid \mathbf{a} = a, \Pi = \pi)$$

and hence $|J| \geq 2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}$. Hence we can bound $|T|$ by

For a random variable X with support size $|X|$ the entropy $H(X)$ is maximized when X is uniform in which case $H(X) = \log n$.

$$\begin{aligned}\log |T| &\leq (m - |J|)n \\ &\leq mn - n2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}.\end{aligned}$$

In particular since the support of the random variable $\mathbf{X} \mid \mathbf{a} = a, \Pi = \pi$ is T we have

$$H(\mathbf{X} \mid \mathbf{a} = a, \Pi = \pi) \leq \log |T| \leq mn - n2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}.$$

Now for an arbitrary $\mathbf{a}$ we have

$$\begin{aligned}H(\mathbf{X} \mid \mathbf{a}, \Pi = \pi) &= \mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}[\mathbf{X} \mid \mathbf{a} = a, \Pi = \pi] \\ &\leq \mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}[mn - n2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}] \\ &= mn - n\mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}[2^{H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)}] \\ &\leq mn - n2^{\mathbb{E}_{a\sim\mathbf{a}|\Pi=\pi}[H(\mathbf{j}|\mathbf{a}=a,\Pi=\pi)]} \\ &= mn - n2^{H(\mathbf{j}|\mathbf{a},\Pi=\pi)}.\end{aligned}$$

We use the convexity of 2^x here and Jensen's inequality which says that if f is convex then $E(f(X)) \geq f(E(x))$.

as desired.

Now we use this fact to bound the probability of a bad transcript. We have

$$\begin{aligned}H(\mathbf{X}, \mathbf{a} \mid \Pi = \pi) &= H(\mathbf{a} \mid \Pi = \pi) + H(\mathbf{X} \mid \mathbf{a}, \Pi = \pi) \\ &\leq |\mathcal{X}| + mn - n2^{H(\mathbf{j}|\mathbf{a},\Pi=\pi)} \\ &< |\mathcal{X}| + mn - n2^t\end{aligned}$$

Note that $(a, \mathbf{X})$ is uniformly distributed over $\mathcal{X} \times \{0,1\}^{m \times n}$ and so

$$\log |\mathcal{X} \times \{0,1\}^{m\times n}| = \log(2^{|\mathcal{X}|}2^{mn}) = |\mathcal{X}| + mn$$

showing that

$$H(\mathbf{X}, \mathbf{a} \mid \Pi = \pi) < \log |\mathcal{X} \times \{0,1\}^{m\times n}| - n2^t.$$

Hence by Fact 2.27 in [GMWW14] we have

$$\mathbb{P}[\Pi = \pi] \leq 2^{-n2^t}.$$

Applying a union bound over all transcripts which can possibly be bad we derive

$$\mathbb{P}[\Pi \in \mathcal{B}] \leq \mathsf{L}(\Pi) \cdot 2^{-n2^t}.$$

We write

$$\begin{aligned}H(\mathbf{j} \mid \mathbf{b}, \Pi) &= \mathbb{E}_{b\sim\mathbf{b}}[H(\mathbf{j} \mid b, \Pi)] \\ &= \mathbb{P}[\Pi \notin \mathcal{B}] \cdot \mathbb{E}_{b\sim\mathbf{b}}[H(\mathbf{j} \mid b, \Pi) \mid \Pi \notin \mathcal{B}] + \\ &\quad \mathbb{P}[\Pi \in \mathcal{B}] \cdot \mathbb{E}_{b\sim\mathbf{b}}[H(\mathbf{j} \mid b, \Pi) \mid \Pi \in \mathcal{B}] \\ &\le \mathbb{P}[\Pi \notin \mathcal{B}] \cdot t + \mathbb{P}[\Pi \in \mathcal{B}] \cdot \log |I| \\ &\le t + \mathsf{L}(\Pi) \cdot 2^{-n2^t} \cdot \log |I|.\end{aligned}$$

An analogous derivation for $H(\mathbf{j} \mid \mathbf{a}, \Pi)$ yields the same bound and so

$$H(\mathbf{j} \mid \mathbf{b}, \Pi) + H(\mathbf{j} \mid \mathbf{a}, \Pi) \le 2t + \mathsf{L}(\Pi) \cdot 2^{1-n2^t} \cdot \log |I|.$$

It follows that we have

$$I(\Pi : \mathbf{a}, \mathbf{b} \mid \mathbf{X}) \ge 2\log |I| - 2t - \mathsf{L}(\Pi) \cdot 2^{1-n2^t} \cdot \log |I|$$

Note that if $\mathsf{L}(\Pi) \ge 2^{2\log |I| + n}$ then the theorem statement follows immediately. So we can safely assume $\mathsf{L}(\Pi) < 2^{2\log |I| + n} = |I|^2 2^n$. Now using $t = \log\log |I|$ we have

$$\begin{aligned}\mathsf{L}(\Pi) 2^{1-n2^t} &< |I|^2 2^n 2^{1-n2^t} \\ &= |I|^2 2^n 2^{1-n\log |I|} \\ &= |I|^2 2^{n+1} |I|^{-n} \\ &= 2^{n+1} |I|^{2-n} \\ &= \frac{2^{n+1}}{|I|^{n-2}} \\ &< \frac{1}{|I|}\end{aligned}$$

(assuming $|I| > 3$ and using $n > 2$). Thus we have

$$\begin{aligned}I(\Pi : \mathbf{a}, \mathbf{b} \mid \mathbf{X}) &\ge 2\log |I| - 2t - \mathsf{L}(\Pi) \cdot 2^{1-n2^t} \cdot \log |I| \\ &> 2\log |I| - 2t - \log |I| / |I| \\ &> 2\log |I| - O(t)\end{aligned}$$

as desired.

□

Corollary 0.23 (Theorem 1.11 in [GMWW14]). *Let $\oplus_m$ be the parity of m input bits. Then*

$$\mathsf{P}^{cc}(R_{\oplus_m \diamond U_n}) \ge 2\log m + n - O(\log\log m)$$

Proof. In this case $|I| = m$ and so the result is immediate. □

Corollary 0.24. *Let MAJ_m be the majority on m input bits when m is odd. Then*

$$\mathsf{P}^{cc}(R_{\mathrm{MAJ}_m \diamond U_n}) \ge 2\log m + n - O(\log\log m)$$

Proof. In this case $|I| = (m-1)/2$. Thus we have by the theorem

$$\begin{aligned}\mathsf{P}^{\mathsf{cc}}(R_{\textsc{maj}_m \diamond U_n}) &\geq 2\log(m/2) + n - O(\log\log(m/2)) \\ &= 2\log m + n - O(\log\log m).\end{aligned}$$

□

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[GMWW14] Dmitry Gavinsky, Or Meir, Omri Weinstein, and Avi Wigderson. Toward better formula lower bounds: an information complexity approach to the krw composition conjecture. In *Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing*, pages 213–222. ACM, 2014.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Computational Complexity*, 5(3-4):191–204, 1995.

[RW92] Ran Raz and Avi Wigderson. Monotone circuits for matching require linear depth. *Journal of the ACM (JACM)*, 39(3):736–744, 1992.

[RY19] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2019.