The communication complexity of pointer chasing*

Stephen J. Ponzio[1] Jaikumar Radhakrishnan[2] S. Venkatesh[2]

Abstract

We study the k-round two-party communication complexity of the pointer chasing problem for fixed k. Damm, Jukna and Sgall [2] showed an upper bound of $O(n \log^{(k-1)} n)$ for this problem. We prove a matching lower bound; this improves the lower bound of $\Omega(n)$ shown by Nisan and Wigderson [11], and yields a corresponding improvement in the hierarchy results derived in [11, 7] for bounded-depth monotone circuits.

We consider the *bit version* of this problem, and show upper and lower bounds. This implies that there is an abrupt jump in complexity, from linear to superlinear, when the number of rounds is reduced to $k/2$ or less. We also consider the *s-paths version* (originally studied by Klauck [7]) and show an upper bound.

The lower bounds are based on arguments using entropy. One of the main contributions of this work is a *transfer lemma* for distributions with high entropy; this should be of independent interest.

1 Introduction

The following pointer chasing problem plays a central role in the study of bounded round communication complexity.

> There are two players A and B. There are two sets of vertices V_A and V_B of size n each. Player A is given a function $f_A : V_A \to V_B$ and player B is given a function $f_B : V_B \to V_A$. In the problem g_k, the players have to determine the vertex reached by applying f_A and f_B alternately, k times, starting with a fixed vertex $v_0 \in V_A$. That is, in g_1, they must determine $f_A(v_0)$, in g_2 they must determine $f_B(f_A(v_0))$, in g_3 they must determine $f_A(f_B(f_A(v_0)))$, and so on.

It is easy to see that there is a k-round protocol for g_k, where A starts and each player alternately sends $\log n$ bits, thus finishing the task in k rounds with a total of $k \log n$ bits of communication. It appears much harder, however, to solve the problem if only $k-1$ rounds of communication are allowed, or even if k rounds of communication are allowed, but B is required to send the first message. Indeed, it was shown in a series of papers (Papadimitriou and Sipser [13], Duris, Galil and Schnitger [3], and Nisan and Wigderson [11]) that the problem does become much harder

*A preliminary version of this paper appeared in the *proceedings of the Thirty-First Annual ACM symposium on Theory of Computing*, 1999 [14]. A one-page abstract of this paper also appeared in the *proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity*, 1999.

[1]Integrated Objects, Boston, USA, email:ponzio@erols.com.

[2]School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, email: {jaikumar,venkat}@tcs.tifr.res.in.

[3]Part of this work was done while Stephen Ponzio and Jaikumar Radhakrishnan were visiting The Hebrew University, Jerusalem.

in this case. To describe the results in those papers, we will use the following notation: $C^{A,k}(f)$ [$C^{B,k}(f)$] denotes the cost of the best k-round deterministic protocol for f in which player A [B] sends the first message.

1.1 Previous work

Papadimitriou and Sipser [13] conjectured that $C^{B,k}(g_k)$ is exponentially higher than $C^{A,k}(g_k)$; they proved this for $k = 2$. Later, Duris, Galil and Schnitger [3] solved the general case and showed that $C^{B,k}(g_k) = \Omega(n/k^2)$. Yao [15] and Halstenberg and Reischuk [4] studied the *rounds* problem in the probabilistic model. Halstenberg and Reischuk [4] showed the existence of functions with exponential gap between k and $k-1$ round protocols in the probabilistic setting. Yao [15] gave an explicit function for the case $k = 2$ with such a gap. Nisan and Wigderson [11] improved the lower bound of Duris, Galil and Schnitger [3] to $\Omega(n)$ for deterministic protocols. For ϵ-error randomized protocols, they proved a lower bound of $\Omega(n/k^2)$ and an upper bound of $O((n/k)\log n)$.

Damm, Jukna and Sgall [2] introduced the model of conservative one-way multi-party complexity. They showed tight upper and lower bounds of $\Theta(n \log^{(k-1)} n)$ for the pointer chasing problem in this model. (In this model, the pointer chasing problem has a directed layered graph with a starting node at layer 0, and k layers of n nodes each. Edges go from one layer to the next, and each node has exactly one edge leaving it. Each player communicates exactly once, and when it is player i's turn to speak, he knows the node reached by following the first $i-1$ edges and all the edges leaving layer j, for $j > i$.)

More recently, Klauck [7] considered a generalization of g_k in the two-party model, and used it to prove a round-communication hierarchy theorem in the presence of limited nondeterministic bits solving an open problem of Hromkovič and Schnitger [5]. In this generalization, the players have to find the sum (mod 2) of the nodes (which are viewed as $\log n$ bit vectors) reached by following edges from s different vertices in V_A. Klauck showed that any k-round nondeterministic protocol, using at most s nondeterministic bits and with player B sending the first message, needs $\Omega(n/(s^2k^2 \log n))$ bits of communication.

1.2 Results of this paper

All the problems considered below are in the two-party model.

1.2.1 The pointer game, g_k

Although the problem g_k was the first problem studied in bounded round communication complexity, the bounds for $C^{B,k}(g_k)$ are not tight. In fact, Nisan and Wigderson [11], while showing a lower bound of $\Omega(n)$ state that "it is not clear how to use less than $n \log n$ bits if only $k-1$ rounds are allowed or, in fact, if k rounds are allowed but player B starts." Damm, Jukna and Sgall [2] showed that $C^{B,k}(g_k) = O(n \log^{(k-1)} n)$ for any fixed k. In fact, they observed that the protocol they discovered for the multi-party model can be used even for the two-party model. A lower bound of $\Omega(n \log^{(k-1)} n)$ is shown in [2] for the multi-party model; however, the method used in that paper does not appear to be applicable to the two-party model (Damm and Sgall, personal communication). Thus, there is still a gap between the best lower bound ($\Omega(n)$) and upper bound ($O(n \log^{(k-1)} n)$) for $C^{B,k}(g_k)$. Our first result shows that the protocol of Damm, Jukna and Sgall [2] is optimal up to a constant factor.

Result 1

(a) $C^{B,k}(g_k) = \Omega(n \log^{(k-1)} n)$ for all fixed k.

(b) $C^{B,k}_{\frac{1}{3}}(g_k) = \Omega(n \log^{(k-1)} n)$ for all fixed k.

Here, $C^{B,k}_{\epsilon}(f)$ denotes the cost of the best k-round ϵ-error randomized protocol for f when player B sends the first message. This improves the $\Omega(n/k^2)$ lower bound of Nisan and Wigderson [11] for any constant k. The techniques used in obtaining these results are related to those of [11].

1.2.2 The bit game p_k

The problem g_k demands a $\log n$ bit answer; however, the lower bound of Nisan and Wigderson [11] holds even if only the most significant bit of the answer is required. (In [11], they consider the parity of the $\log n$ bits, but it is easy to see that their claim holds for the most significant bit as well.) In this case, however, it can be shown that the lower bound is tight: there does exist a $(k-1)$-round protocol that uses $O(n)$ bits. In fact, $O(n)$ bits suffice even if we have fewer than $k-1$ rounds. Let $p_k(f_A, f_B) = [g_k(f_A, f_B)]_o$, where b_o denotes the most significant bit of the Boolean vector b. We only consider even k. The case of odd k can be handled similarly.

Result 2

(a) For $r \leq \frac{k}{2} - 1$, $C^{B,k-r}(p_k) = O((k-r-2)\log n + (r+1)n)$.

(b) $C^{B,\frac{k}{2}}(p_k) = O(n \log^{(k/2-1)} n)$ for all fixed k.

Note that these upper bounds show an abrupt jump at $r = \frac{k}{2}$: while the Result 2(a) gives a linear upper bound up to $r = \frac{k}{2} + 1$, Result 2(b), which works with one less round, needs *superlinear* number of bits. We next show that such an abrupt jump is unavoidable.

Result 3 $C^{B,\frac{k}{2}}(p_k) = \Omega(n \log^{(k-2)} n)$ for all fixed k.

The techniques used in the proof of Result 1(a), as well as those employed in [11, 2], are essentially top down. This is not surprising, considering that the optimal protocols for those problems also work in a top-down fashion—by chasing the pointer from the top. However, the upper bounds for the problem p_k described in Results 2(a) and 2(b) use both the top down and bottom up approaches simultaneously, and previous techniques go only half-way in providing us a lower bound for p_k. For the other half, we need to use different techniques. In Section 3, we describe the technical difficulty we encounter, and describe how this can be surmounted by using a *transfer lemma*, based on ideas that connect entropy and sampling. In particular, to measure the amount of randomness in a source on $\{0,1\}^n$, we consider how well it can two-color a set of large size. We believe that the techniques developed here are an important contribution of this work, and that this method should find other applications.

1.2.3 The s-pointer game, g_k^s

We now consider the following generalization of the pointer chasing problem g_k: starting from fixed vertices $v_1, \ldots, v_s \in V_A$, players A and B need to find the s nodes reached by applying f_A and f_B alternately, k times. We call this function g_k^s. It is easy to see that there is a k-round protocol for g_k^s, when player A starts and each player alternately sends $s \log n$ bits, thus finishing the task in k rounds with a total of $sk \log n$ bits of communication.

This function is important in the investigation of limited nondeterminism in communication complexity. Hromkovič and Schnitger [5] proved a weak rounds-communication hierarchy for limited nondeterminism that holds only for at most a logarithmic number of guess bits. Klauck [7] used a function closely related to g_k^s and showed that such a hierarchy result holds also for superlogarithmic number of guess bits.

Define $h^{(k)}(n,s)$ by: $h^{(0)}(n,s) = \log n$, $h^{(k+1)}(n,s) = \log h^{(k)}(n,s) + \log s$.

By generalizing the protocol of Damm, Jukna and Sgall [2], we obtain the following upper bound.

Result 4 For all fixed k, $C^{B,k}(g_k^s) = O(nh^{(k-2)}(n,s))$.

1.2.4 Applications to circuit complexity.

The improved lower bounds for communication complexity of g_k (Result 1(a)) can be used to improve the monotone constant depth hierarchy theorems derived by Nisan and Wigderson [11] and Klauck [7].

- Let H_k be a boolean function defined by the formula that is a complete n-ary tree of depth k, alternating levels of $\vee$ and $\wedge$ gates (with $\vee$ at the root) and n^k distinct variables at the leaves.
- Let Q_k^s be the alternating $\wedge - \vee$ tree of depth $k+1$ where the top $\wedge$ gate has fan-in s and all other gates have fan-in n with sn^k distinct variables at the leaves.

Corollary 5

(a) Any monotone circuit of depth $k-1$ (unbounded fan-in) for H_k requires size $2^{\Omega(n \log^{(k-1)} n/k)}$.

(b) Any monotone circuit of depth k (unbounded fan-in) with $O(s)$ nondeterministic bits for Q_k^s has size $2^{\Omega(n \log^{(k)} n/k)}$.

1.3 Organization of the paper

The rest of this paper is organized as follows. In the next section, we give proofs of the upper and lower bounds for the pointer game. Section 4 gives an informal description of the transfer lemma and the proofs of the upper and lower bounds for the bit game. We presents the upper bound proof for the s-pointer game in Section 5 and discuss the applications to circuit complexity in Section 6 . In Section 7, we present the proof of the transfer lemma.

2 Preliminaries

We will need the following information theoretic definitions in this paper.

Definition 2.1 (Entropy) *Let X be a random variable with finite range,* range(X)*. Then we define its entropy by*

$$H[X] = -\sum_x \Pr[X = x] \log \Pr[X = x].$$

Definition 2.2 (Conditional Entropy) *If X and Y are random variables, then the conditional entropy of Y given X is defined by*

$$H[X|Y] = E_X[H[Y_X]]$$

where Y_x is the random variable such that, for all y,

$$\Pr[Y_x = y] = \Pr[Y = y \mid X = x].$$

Unless mentioned explicitly, all logarithms are to the base 2. The iterated logarithm $\log^{(i)} n$ is defined by $\log^{(0)} n = n$ and $\log^{(k+1)} n = \log \log^{(k)} n$. We define $\textsc{tower}(i, b, c)$ as follows: $\textsc{tower}(0, b, c) = c$, $\textsc{tower}(i + 1, b, c) = b^{\textsc{tower}(i,b,c)}$ for $i \geq 0$.

3 The pointer game

We start with the formal definition of the pointer game.

Definition 3.1 *Let V_A and V_B be two disjoint sets of n vertices each. Let $V = V_A \cup V_B$. Let $F_A = \{f_A | f_A : V_A \to V_B\}$ and $F_B = \{f_B | f_B : V_B \to V_A\}$. Let $f : V \to V$, be the function $f_A \cup f_B$; that is,*

$$f(v) = \begin{cases} f_A(v), & v \in V_A \\ f_B(v), & v \in V_B \end{cases}$$

Fix $v_0 \in V_A$. The function A and B wish to compute is $g_k : F_A \times F_B \to V$ defined by $g_k(f_A, f_B) = f^{(k)}(v_0)$, where $f^{(0)}(v) = v$, and $f^{(k)}(v) = f(f^{(k-1)}(v))$ for $k > 0$. That is, A is given a set of pointers f_A and B is given a set of pointers f_B, and they need to determine $f^{(k)}(v_0)$. We say that a vertex v is at a distance i from v_0 if $v = f^{(i)}(v_0)$.

Theorem 1 (Damm, Jukna and Sgall [2]) *$C^{B,k}(g_k) = O(n \log^{(k-1)} n)$ for all fixed k.*

Proof:

Round 1: B sends the last $\log^{(k-1)} n$ bits of $f_B(v)$ for every $v \in V_B$.

Round 2: A now knows the last $\log^{(k-1)} n$ bits of $f_B(v_1)$, where $v_1 = f_A(v_0)$. A sends v_1 and the last $\log^{(k-2)} n$ bits of $f_A(v)$ for every possible vertex v in A's list that v_1 could be pointing to.

Round 3: B now knows the last $\log^{(k-2)} n$ bits of $f_A(v_2)$, where $v_2 = f_B(v_1)$. B sends v_2 and the last $\log^{(k-3)} n$ bits of $f_B(v)$ for every possible vertex in B's list that v_2 can be pointing to.

⋮

Round k: If k is even, then A would know v_k by this time, and A sends this information over to B. If k is odd, B would know v_k by this time, and B sends this information over to A.

It is easy to see that this protocol uses $O(n \log^{(k-1)} n)$ bits of communication. ■

To obtain Results 1(a) and 1(b), we sharpen the analysis used by Nisan and Wigderson [11] to obtain their $\Omega(n)$ lower bound for $C^{B,k}(g_k)$; our proof, like theirs, uses entropy. We briefly recall their argument. Suppose at most ϵn bits of communication are allowed, for a k-round protocol, where player B starts the communication. Then, it can be shown that after $k-1$ rounds are over, the player whose turn it is to speak next will have less than 1 bit of information about the final answer. However, this player must already know the answer, because he is not going to receive any more messages—a contradiction.

We modify their proof as follows. Suppose only cn bits are allowed per round. Then, after $k-1$ rounds, the player whose turn it is to speak next knows only some $\alpha(c,k)$ bits of information about the answer. Thus, since this player must know the answer exactly, $\alpha(c,k)$ must be at least $\lfloor \log n \rfloor$. It turns out that $\alpha(c,k)$ looks like a tower of 2's of height $k-2$, with c on the top. This, implies a lower bound of $\Omega(\log^{(k-1)} n)$ for c, and hence a lower bound of $\Omega(n \log^{(k-1)} n)$ for $C^{B,k}(g_k)$.

Theorem 2 *$C^{B,k}(g_k) = \Omega(n \log^{(k-1)} n)$ for all fixed k.*

Proof: Fix a k-round protocol of g_k where player B sends the first message. To analyze the protocol, we consider inputs (f_A, f_B) chosen uniformly at random from the set of all possible inputs. For, $i = 1, 2, \ldots, k$, let V_i be the vertex at distance i from v_0. Let M_i be the message sent in the ith round of the protocol. Note that V_i's and M_i's are random variables, whose values are completely determined once f_A and f_B are chosen. The players have to communicate and determine V_k. We will show that there is at least one choice of (f_A, f_B) for which A and B exchange many bits in this protocol.

The proof will proceed in stages. In the first stage we will restrict attention to inputs where the first message M_1 is m_1 and the first vertex V_1 is v_1. In general, at the end of stage i, we will fix the first i messages $M_1, M_2, \ldots, M_i$ at $m_1, m_2, \ldots, m_i$, and the first i vertices $V_1, V_2, \ldots, V_i$ at $v_1, v_2, \ldots, v_i$. We use X_i to refer to the set of inputs compatible with the setting determined after stage i, and let (f_A^i, f_B^i) be a uniformly chosen random element of X_i. Note that if we consider the original random inputs (f_A, f_B) under the condition that they belong to X_i, then the resulting distribution is identical to the distribution of (f_A^i, f_B^i). Also, X_i is the Cartesian product of sets of inputs of A and B that are consistent with the setting. This is analogous to the fact (see [9, pages 7-9]) that the set of inputs consistent with the communication, after the first few messages have been fixed, forms a subrectangle of the communication matrix. Thus, f_A^i and f_B^i are independent of each other.

Now, suppose only cn bits are sent in each round. Suppose it is B's turn to speak last (that is after $k-1$ rounds). Then, after $k-1$ message have been sent, B knows the answer. We will show that unless c is large, $H[f_A^{k-1}(v_{k-1})]$ is non-zero. On the other hand, since B knows the answer at this stage, the value of V_k is the same for all inputs in X_{k-1}, that is, $H[f_A^{k-1}(v_{k-1})]$ is zero. We conclude from this that c must be large. To implement the plan outlined above, we will maintain the following induction hypothesis.

> There exist choices $m_1, m_2, \ldots, m_i$ and $v_1, v_2, \ldots, v_i$, such that X_i is non-empty, and the following three invariants hold.

(I1) $H[f_A^i] \;\geq\; n\log n - (2^i - 1)cn$.

(I2) $H[f_B^i] \;\geq\; n\log n - (2^i - 1)cn$.

(I3) $H[f^i(v_i)] \;\geq\; \log n - \delta(i)$, where the function δ is defined inductively by $\delta(0) = 0$, and for $j \geq 1$, $\delta(j) = (2^j - 1)2^{2(\delta(j-1)+1)}c$. Also, here f^i stands for f_A^i if i is even, and f_B^i if i is odd.

Base case: After Round 0 (that is, initially), (f_A^0, f_B^0) has the same distribution as (f_A, f_B). Therefore, $H[f_A^0], H[f_B^0] = n\log n$ and $H[f_A^0(v_0)] = \log n$. Clearly, the hypothesis is true.

Induction step: We will show the induction step for odd i; the proof for even i is identical. When i is odd, it is B's turn to speak in Round i. By the induction hypothesis, after Round $i-1$, $m_1, m_2, \ldots, m_{i-1}$ and $v_1, v_2, \ldots, v_{i-1}$ have been chosen, so that

$$\begin{aligned} H[f_A^{i-1}] &\geq n\log n - (2^{i-1} - 1)cn; \\ H[f_B^{i-1}] &\geq n\log n - (2^{i-1} - 1)cn; \\ H[f_A^{i-1}(v_{i-1})] &\geq \log n - \delta(i-1). \end{aligned}$$

We need to find values m_i for M_i and v_i for V_i, such that (I1), (I2) and (I3) continue to hold after Round i.

Choosing m_i: Since the values for first $i-1$ messages have been fixed, M_i is now a function of f_B^{i-1} alone. We pick the value m_i for M_i such that the resulting distribution f_B^i has the maximum entropy. (Since f_B^{i-1} is uniformly distributed, this just amounts to picking m_i such that $\Pr_{(f_A^{i-1}, f_B^{i-1})}[M_i = m_i]$ is maximum.)

Choosing v_i: We will use the following lemma.

Lemma 3.2 *Let X and Y be random variables taking values in sets of size n and m respectively. If $H[X] \geq \log n - k$ and $H[Y \mid X] \geq \log m - \ell$, then for at least $n/2^{2(k+1)}$ values x of X,*

$$H[Y_x] \geq \log m - 2\ell.$$

Here Y_x denotes the restriction of Y obtained by fixing X at the value x.

Proof: By definition, $H[Y \mid X] = E_X[H[Y_X]]$. Since $H[Y \mid X] \geq \log m - \ell$, by Markov's inequality, we get

$$\Pr[H[Y_X] < \log m - 2\ell] < \frac{1}{2}.$$

Therefore, we know that with high probability ($> \frac{1}{2}$), X satisfies the condition $H[Y_X] \geq \log m - 2\ell$. But what we need to show is that this high probability event includes a large number of values for X. For this, we will use the assumption $H[X] \geq \log n - k$. Let S be the set of values x of X for which the condition $H[Y_x] \geq \log m - 2\ell$ holds; let $p \stackrel{\text{def}}{=} \Pr[X \in S]$ (note $p \geq 1/2$). Let $\chi_S = 1$ when $X \in S$, and $\chi_S = 0$ otherwise. We have

$$\begin{aligned} \log n - k \;\leq\; H[X] &= H[X, \chi_S] \\ &= H[\chi_S] + H[X \mid \chi_S] \\ &\leq 1 + p\log|S| + (1-p)\log n. \end{aligned}$$

This gives $\log |S| \geq \log n - (k+1)/p \geq \log n - 2(k+1)$; that is, $|S| \geq n/2^{2(k+1)}$. □

We now return to the induction step above, and apply this lemma with $X = f_A^{i-1}(v_{i-1})$ and $Y = f_A^{i-1}$. By condition (I3) of the induction hypothesis, $H[X] \geq \log n - \delta(i-1)$, and by (I1), we have

$$\begin{aligned} H[Y \mid X] &\geq n \log n - (2^{i-1} - 1)cn - H[X] \\ &\geq n \log n - (2^{i-1} - 1)cn - \log n. \end{aligned}$$

The lemma, then gives us a set T, with size $t \geq n/2^{2(\delta(i-1)+1)}$, of values x for X, such that

$$H[Y_x] \geq n \log n - 2((2^{i-1} - 1)cn + \log n).$$

Recall that X stands for $f_A^{i-1}(v_{i-1})$, so these value x are actually values for $V_i = f_A^{i-1}(v_{i-1})$. We now pick the value v_i among these possibilities such that $H[f_B^i(v_i)]$ is maximum. This completes the choice of v_i.

We will now show that if m_i and v_i are so chosen, then (I1), (I2) and (I3) hold. (I1) and (I2) are straightforward to establish:

$$\begin{aligned} H[f_A^i] &\geq n \log n - 2((2^{i-1} - 1)cn + \log n) \\ &\geq n \log n - (2^i - 1)cn; \\ H[f_B^i] &\geq n \log n - (2^{i-1} - 1)cn - cn \\ &\geq n \log n - (2^i - 1)cn. \end{aligned}$$

It remains to verify (I3). We know $H[f_B^i] \geq n \log n - (2^i - 1)cn$. We have

$$\sum_{v \notin T} \log n + \sum_{v \in T} H[f_B^i(v)] \;\geq\; H[f_B^i] \;\geq\; n \log n - (2^i - 1)cn.$$

Since v_i was chosen to be that vertex in T for which $H[f_B^i(v)]$ was maximum, we have

$$\begin{aligned} H[f_B^i(v_i)] &\geq (n \log n - (2^i - 1)cn - (n-t)\log n)/t \\ &\geq \log n - (2^i - 1)cn/t \\ &\geq \log n - \delta(i). \end{aligned}$$

For the last inequality, we used $t \geq n/2^{2(\delta(i-1)+1)}$ and the definition of $\delta(i)$. This establishes (I3), and completes the induction.

Claim 3.3 *$\forall k$, $\forall n$ large enough, $c = \Omega(\log_2^{(k-1)} n)$.*

Proof: We first show by induction that for $j \geq 1$,

$$\delta(j) \leq \textsc{tower}(j-1, 16, 4c).$$

For the base case, we have $\delta(1) = 4c = \textsc{tower}(0, 16, 4c)$. For the induction step, we have

$$\begin{aligned} \delta(j) &= (2^j - 1)2^{2(\delta(j-1)+1)}c \\ &\leq (2^{j+2})c \times 4^{\textsc{tower}(j-2,16,4c)} \\ &\leq 4^{\textsc{tower}(j-2,16,4c)} \times 4^{\textsc{tower}(j-2,16,4c)} \\ &\leq \textsc{tower}(j-1, 16, 4c). \end{aligned}$$

Now, if the protocol is correct, then after $k-1$ rounds, the player whose turn it is to speak next should already know the answer V_k. This implies that the entropy of V_k is zero, that is, $\delta(k-1) \geq \log_2 n$. Thus,

$$4c \geq \log_{16}^{(k-2)} \log_2 n \geq \log_{16}^{(k-1)} n.$$

We will show that for all large n (such that $\log_2^{(k-1)} n \geq 15$),

$$\log_{16}^{(k-1)} n \geq \frac{1}{5} \log_2^{(k-1)} n. \tag{1}$$

Then, we have the bound $c \geq \frac{1}{20} \log_2^{(k-1)} n$, justifying the claim.

We use induction to show (1). For $j = 1$, we have

$$\log_{16}^{(j-1)} n = n > \frac{n}{5} = \frac{1}{5} \log_2^{(j-1)} n.$$

For $1 < j \leq k-1$, using the induction hypothesis, we have

$$\begin{aligned} \log_{16}^{(j-1)} n &\geq \log_{16}(\frac{1}{5} \log_2^{(j-2)} n) \\ &\geq \frac{1}{4} \log_2(\frac{1}{5} \log_2^{(j-2)} n) \\ &\geq \frac{1}{5} \log_2^{(j-1)} n \times (\frac{5}{4} - \frac{5 \log 5}{4 \log_2^{(j-2)} n}) \\ &\geq \frac{1}{5} \log_2^{(j-1)} n. \end{aligned}$$

For the last inequality, we assumed that $\log_2^{(j-1)} n \geq 15$. $\square$

The theorem follows from the above claim. $\blacksquare$

Next, we show that the lower bound holds even for randomized protocols.

Theorem 3 $C_{\frac{1}{3}}^{B,k}(g_k) = \Omega(n \log^{(k-1)} n)$ *for all fixed* k.

Proof: We show that no deterministic k-round protocol, with B starting the communication, that uses significantly less than $n \log^{(k-1)} n$ bits can solve the problem correctly for a non-negligible fraction of the inputs when the inputs are chosen according to the uniform distribution. The theorem then follows by an application of Yao's lemma [15].

Fix a k-round deterministic $\frac{1}{3}$-error protocol for g_k, where player B begins the communication. The method used in the proof of this theorem is almost identical to that used in the proof of the previous theorem. There, we showed that if the protocol used a small amount of communication, then there was at least one input on which it erred; here we strengthen this and show that if the protocol uses a small amount of communication, then it must give wrong answers for a large fraction of the inputs.

As before, we consider inputs of the form (f_A, f_B) chosen uniformly at random from the set of all inputs, and let $M_1, M_2, \ldots, M_k$ and $V_1, V_2, \ldots, V_k$ be the random variables representing the first k messages and the first k vertices. Suppose the protocol uses cn bits of communication on the average. By stopping this protocol, as soon as $100cn$ bits have been exchanged, we get another protocol for g_k, which never uses more than Cn ($C = 100c$) bits of communication, and yet gives wrong answers with probability at most $\frac{1}{3} + \frac{1}{100}$ over choices of (f_A, f_B). In the rest of the proof we shall work with this modified protocol, and show that C must be large.

Pick a random input from the input distribution defined above, and consider the situation after i rounds are over. Let (f_A^i, f_B^i) denote the induced distribution on the set of inputs X_i that are consistent with the i messages $M_1, M_2, \ldots, M_i$ sent so far and give rise to the same values for $V_1, V_2, \ldots, V_i$ as the present input. Note that this is, in fact, uniform distribution on the set of such inputs; furthermore, f_A^i and f_B^i are independent. We will show by induction that the following hypothesis holds.

> For $i = 0, 1, 2, \ldots, k-1$, with probability at least $\prod_{j=1}^{i}(1 - 1/2^{j+1})$ over choices of inputs for the two players, the following three invariants hold.
>
> (I1) $H[f_A^i] \geq n \log n - 2^{10i^2}Cn$.
>
> (I2) $H[f_B^i] \geq n \log n - 2^{10i^2}Cn$.
>
> (I3) $H[f^i(V_i)] \geq \log n - \delta(i)$, where the function δ is defined inductively by $\delta(0) = 0$ and $\delta(i) = 2^{2^{i+3}(\delta(i-1)+1)}2^{10i^2}C$. Also, here f^i stands for f_A^i if i is even, and f_B^i if i is odd.

Base case: Initially, with probability 1, (f_A^0, f_B^0) is the uniform distribution on the set of all inputs, and the claim follows immediately.

Induction step: We will assume that i is odd; the proof for even i is identical. It is B's turn to speak in Round i. By the induction hypothesis, after Round $i-1$, with probability $\prod_{j=1}^{i-1}(1 - 1/2^{j+1})$, we have:

$$\begin{aligned} H[f_A^{i-1}] &\geq n \log n - 2^{10(i-1)^2}Cn. \\ H[f_B^{i-1}] &\geq n \log n - 2^{10(i-1)^2}Cn. \\ H[f_A^{i-1}(V_{i-1})] &\geq \log n - \delta(i-1). \end{aligned}$$

Note, that all inputs in f_A^{i-1} and f_B^{i-1} have the same value for $V_1, \ldots, V_{i-1}$; so the variable V_{i-1} in the third inequality above is actually constant over the set of inputs X_i. To show that the induction hypothesis holds after Round i, we first fix an input (f_A, f_B) such that the hypothesis holds after Round $i-1$, and show that more than a fraction $1 - \frac{1}{2^{i+1}}$ of the inputs in X_{i-1} satisfy the induction hypothesis after Round i.

(I1) *holds with high probability:* Since $H[V_i] \leq \log n$, fixing the value of V_1 can cause an average reduction of at most $\log n$ in $H[f_A^{i-1}]$. Thus, by Markov's inequality, with probability at least $1 - \frac{1}{2^{i+3}}$ over choices of inputs from X_i, we have

$$H[f_A^i] \geq n \log n - 2^{i+3}(2^{10(i-1)^2}Cn + \log n) \geq n \log n - 2^{10i^2}Cn.$$

(I2) *holds with high probability:* Since messages are only Cn bits long, the average reduction in entropy of f_B^{i-1} after B's message is at most Cn. Then, by Markov's inequality, with probability at least $1 - \frac{1}{2^{i+2}}$, B sends a message m such that the remaining entropy of f_B^{i-1} is at least

$$H[f_B^i] \geq n \log n - 2^{10(i-1)^2}Cn - 2^{i+2}Cn \geq n \log n - 2^{10i^2}Cn.$$

(I3) *holds with high probability given* (I2)*:* Assume that (I2) holds. Note that (I2) depends only on f_B^{i-1}, which is independent of f_A^{i-1}. We shall show that with probability at least $1 - \frac{1}{2^{i+3}}$,

$$\Pr[H(f_B^i(V_i)) \geq \log n - \delta(i)] \geq 1 - \frac{1}{2^{i+3}}.$$

For otherwise,

$$\Pr[H[f_B^i(V_i)] < \log n - \delta(i)] > \frac{1}{2^{i+3}}.$$

Let S be the set of vertices in V_B such that $H[f_B^i(v)] < \log n - \delta(i)$. By the above inequality, we know that $\Pr[V_i \in S] = p \geq \frac{1}{2^{i+3}}$. On the other hand, we know by (I3) of the induction hypothesis that $H[V_i] \geq \log n - \delta(i-1)$. By a calculation similar to the one used at the end of the proof of Lemma 3.2, we see that $|S| \geq n/2^{2^{i+3}(\delta(i-1)+1)}$. But then, we have

$$\begin{aligned} H[f_B^i] &\leq \sum_{v \in S} H[f_B(v)] + \sum_{v \notin S} H[f_B(v)] \\ &< \frac{n}{2^{2^{i+3}(\delta(i-1)+1)}}(\log n - \delta(i)) + (n - \frac{n}{2^{2^{i+3}(\delta(i-1)+1)}}) \log n \\ &= n \log n - 2^{10i^2} Cn, \end{aligned}$$

where the last inequality follows from the definition of $\delta(i)$. But this contradicts our assumption that (I2) holds.

To ensure that with high probability the three invariants continue to hold after Round i, we observe that

$$\begin{aligned} \Pr[(\mathrm{I1}) \wedge (\mathrm{I2}) \wedge (\mathrm{I3})] &\geq \Pr[(\mathrm{I2})] \times \Pr[(\mathrm{I3}) \mid (\mathrm{I2})] - \Pr[\neg(\mathrm{I1})] \\ &\geq (1 - \frac{1}{2^{i+2}})(1 - \frac{1}{2^{i+3}}) - \frac{1}{2^{i+3}} \\ &\geq 1 - \frac{1}{2^{i+1}}. \end{aligned}$$

This completes the induction.

This implies that with probability at least $\prod_{j=1}^{k-1}(1 - 1/2^{j+1}) \geq \frac{1}{2}$, at the end of Round $k-1$, the condition (I3) above holds. If $\delta(k-1) < \frac{\log n}{100}$, $H[f^{k-1}(V_{k-1})] \geq \frac{99}{100} \log n$, and hence V_k cannot take any fixed value with probability more than $\frac{1}{50}$. But then, the protocol makes an error with probability at least $\frac{1}{2} \times \frac{49}{50} > \frac{1}{3} + \frac{1}{100}$, which is more than it is allowed. Hence, $\delta(k-1) \geq \frac{\log_2 n}{100}$.

Claim 3.4 *$\forall k$, $\forall n$ large enough, $C = \Omega(\log_2^{(k-1)} n)$.*

Proof: Let $C' = 2^{26}C$. It can be easily checked that $\delta(1) = C'$. We show by induction that for $j \geq 2$,

$$\delta(j) \leq 8^{\mathrm{TOWER}(j-2,16,16C')}.$$

For the base case,

$$\begin{aligned} \delta(2) &= 2^{2^5\delta(1)} \times 2^{2^5+40}C \\ &\leq 2^{2^5 C'} \times 2^{C'} \\ &\leq 8^{16C'} \\ &= \mathrm{TOWER}(0, 16, 16C'). \end{aligned}$$

For the induction step, we have

$$\begin{aligned} \delta(j) &= 2^{2^{j+3}\delta(j-1)} \times 2^{2^{j+3}} \times 2^{10i^2} C \\ &\leq 2^{2^{\mathrm{TOWER}(j-3,16,16C')} \times 8^{\mathrm{TOWER}(j-3,16,16C')}} \times 2^{\mathrm{TOWER}(j-2,16,16C')} \times 2^{\mathrm{TOWER}(j-2,16,16C')} \\ &= 8^{\mathrm{TOWER}(j-2,16,16C')}. \end{aligned}$$

Now, using the fact that $\delta(k-1) \geq \frac{\log_2 n}{100}$, we get that $8^{\mathrm{TOWER}(k-2,16,16C')} \geq \frac{\log_2 n}{100}$. This implies that $C = \Omega(\log_2^{(k-1)} n)$ using calculations similar to those in Claim 3.3. □
The theorem now follows from this claim. ■

4 The bit game

In this section, we show upper and lower bounds for the bit game. We start with a formal definition of the bit game.

Definition 4.1 *Let V_A and V_B be two disjoint sets of n vertices each. Fix $v_0 \in V_A$. Let $F_A = \{f_A | f_A : V_A \to V_B\}$ and $F_B = \{f_B | f_B : V_B \to V_A\}$. Then, $p_k : F_A \times F_B \to \{0,1\}$ is defined by $p_k(f_A, f_B) = [f^{(k)}(v_0)]_0$, where f is as in Definition 3.1 and $[b]_0$ denotes the most significant bit of a Boolean vector b. Let v_i denote the vertex $f^{(i)}(v_0)$ for $i \geq 0$.*

Our next result shows that for any constant k, p_k has a protocol with communication complexity $O(n)$ if more than $k/2$ rounds of communication are allowed. The protocol is a combination of two protocols: one uses the standard pointer chasing protocol for $(k-r)$ steps, the other determines the answer for the $(k-r)$-th vertex by percolating the answer backwards. Thus, the protocol uses a top-down and bottom-up approach simultaneously.

Theorem 4 *$C^{B,k-r}(p_k) = O((k-r-2)\log n + (r+1)n)$ for all fixed $k \geq 2$ and for $r \leq \frac{k}{2} - 1$.*

Proof: In Round 1, B sends the $[f_B(v)]_0$ for every vertex $v \in V_B$. In Round 2, A sends v_1 and $[f_B(f_A(v))]_0$ for every vertex $v \in V_A$. B sends v_2 and $[f_B(f_A(f_B(v)))]_0$ for every vertex $v \in V_B$ in Round 3. The two players A and B repeat this till Round $r+1$, that is, they chase pointers from the top starting at v_0 to reach v_r and also compute the answer for v_{k-r-1} bottom-up.

1. If r is odd, then at the end of Round $r+1$, B knows v_r and $[f^{(r+1)}(v)]_0$ for every vertex v in V_A.

2. If r is even, then at the end of Round $r+1$, A knows v_r and $[f^{(r+1)}(v)]_0$ for every vertex v in V_B.

After Round $r+1$, A and B only follow pointers from the top by repeated application of f to the vertex v_r to determine v_{k-r-1}. Then, at the end of Round $k-r-1$,

1. If r is odd, B knows the vertex v_{k-r-1} and $[f^{(r+1)}(v)]_0$ for every vertex v in V_A. Hence, he knows $p_k[f_A, f_B]$. B sends this answer to A in Round $k-r$.

2. If r is even, A knows the vertex v_{k-r-1} and $[f^{(r+1)}(v)]_0$ for every vertex v in V_B. Hence, he knows $p_k[f_A, f_B]$. A sends this answer to B in Round $k-r$.

It is easy to see that this protocol uses $O((k-r-2)\log n + (r+1)n)$ bits of communication. ■

We now show a $k/2$-round protocol for p_k that uses $O(n \log^{(k/2-1)} n)$ bits of communication. This protocol is again a combination of two protocols: instead of the standard pointer chasing protocol used in the previous result, we now employ the protocol of Damm, Jukna and Sgall.

Theorem 5 *For all fixed k, $C^{B,\frac{k}{2}}(p_k) = O(n \log^{(k/2-1)} n)$.*

Proof: In Round 1, B sends the most significant $\log^{(\frac{k}{2}-1)} n$ bits of $f_B(v)$ for every vertex $v \in V_B$. In Round 2, A sends v_1 and the most significant $\log^{(\frac{k}{2}-2)} n$ bits of $f_A(v)$ for every possible vertex v in A's list that v_1 could be pointing to. A also sends $[f_B(f_A(v))]_0$ for every vertex $v \in V_A$. In Round 3, B sends v_2 and the most significant $\log^{(\frac{k}{2}-3)} n$ bits of $f_B(v)$ for every possible vertex v in B's list that v_2 could be pointing to. B also sends $[f_B(f_A(f_B(v)))]_0$ for every vertex $v \in v_B$. They repeat this for $\frac{k}{2}-1$ rounds, that is, they follow the protocol of Damm, Jukna and Sgall to chase pointers from the top starting at v_0 to reach $v_{\frac{k}{2}}$. Simultaneously, they also compute the answer for vertex $v_{\frac{k}{2}}$ bottom-up. Thus at the end of Round $\frac{k}{2}-1$,

1. If $\frac{k}{2}$ is odd, then B knows the answer. B sends it to A in Round $\frac{k}{2}$.
2. If $\frac{k}{2}$ is even, then A knows the answer. A sends it to B in Round $\frac{k}{2}$.

We can easily check that the protocol uses $O(n \log^{(\frac{k}{2}-1)} n)$ bits of communication. ∎

The previous two results show an abrupt jump in communication complexity from linear to superlinear number of bits of communication when $k/2$ rounds or less are available. We now show that this abrupt jump is unavoidable. Proving a lower bound for the bit version of the problem with fewer than $k/2$ rounds needs new ideas. In the rest of this section, we point out why the ideas already used do not completely solve the problem, and describe the new ideas that help us complete the proof.

Why is the entropy method not sufficient? Let $k = 2r$, r even. We wish to show that no r-round protocol with cn bits per round can solve the problem for constant c. Suppose it is A's turn to communicate in Round r. What can the entropy method described above tell us about the state of the players in the beginning of Round r? Using the entropy method, we can ensure the following: The vertex $v_{r-1} \in V_B$ (at distance $r-1$ from v_0) has already been fixed. The inputs of A and B, that is f_A and f_B, still have large entropy ($n \log n - O(n)$). In particular, for most choices of f_A, the pointer that leaves vertex v_{r-1} still has several possibilities. This tells us that as we vary the B's input (consistently with whatever choices made so far) then the vertex v_r has many possibilities ($\Omega(n)$, the constant depends on k). Since A does not have access to B's input, A does not know which of these vertices is the right one. The entropy method gets us this far. To complete the proof, we now need to show that there are two choices for vertex v_r so that the vertex $f^{(r)}(v_r)$ is in different halves. For this, we need the transfer lemma.

What is the transfer lemma? Roughly speaking, the transfer lemma states the following: suppose χ is a random function from $[n] \to \{0,1\}$. Suppose f is a function picked at random from a probability distribution with entropy $n \log n - O(n)$ on the set of all functions from $[n]$ to $[n]$ (f and χ are independent). Suppose χ has the property that with high probability it can two-color any set of size at least 2^k (say $k = \log n - a$). Then, for most of the values of χ, if we fix χ at that value then the composed function $\chi \circ f$ can two-color any set of size $2^{k'}$, where $k' = \log n - \log a$. (This is a very rough description. For details, see the statement of the lemma below.)

Thus, if initially we have a distribution that can two-color sets of size $\sqrt{n}$, then after one composition, we get a distribution that can two-color sets of size at least $n/\log n$, after another composition (with a 'fresh' function), it can two-color sets of size $n/\log\log n$, and next time $n/\log\log\log n$, and so on.

We now make a few definitions and state the transfer lemma. The proof of the transfer lemma will be given in Section 7.

Definition 4.2 (Distance between distributions) *The variational distance (see Nisan and Zuckerman [12]) between two distributions D and D' on the same underlying set X is*

$$d(D, D') \stackrel{\text{def}}{=} \frac{1}{2} \sum_{x \in X} |D(x) - D'(x)|.$$

We say that a distribution D_1 is within distance d of another distribution D_2 if $d(D_1, D_2) \leq d$.

The following properties follow easily from the definition above.

Proposition 4.3

1. *$d(D, D') = \max_{Y \subseteq X} |\sum_{y \in Y} D(y) - \sum_{y \in Y} D'(y)|$.*
2. *$d(D_1, D_2) \leq d(D_1, D_3) + d(D_3, D_2)$.*
3. *If $f : X \to Y$, and D_1 and D_2 are distributions on X, then $d(f(D_1), f(D_2)) \leq d(D_1, D_2)$.*

Definition 4.4 (Min-entropy) *We say that a distribution D on $[n]$ has min-entropy k if $D(i) \leq 1/2^k$, for all i.*

Definition 4.5 (Coloring distributions) *Let $\chi : [n] \to \{0, 1\}$. Let D be a distribution on $[n]$. We say that χ ϵ-well-colors D, if*

$$|\sum_{i=1}^{n} D(i)\chi(i) - \frac{1}{2}| \leq \epsilon.$$

If χ is a random coloring, then we say that it is a (k, ϵ, δ)-coloring, if for all distributions D with min-entropy k,

$$\Pr_{\chi}[\chi \text{ } \epsilon\text{-well-colors } D] \geq 1 - \delta.$$

We say that χ is a (k, ϵ)-coloring if for all distributions D with min-entropy k and for all $t \geq 0$,

$$\Pr_{\chi}[\chi \text{ } (\epsilon + t)\text{-well-colors } D] \geq 1 - \exp(-\frac{t^2 2^{2k}}{2n}).$$

We can now give a formal statement of the transfer lemma.

Lemma 4.6 (The transfer lemma) *Suppose χ is a random $(\log n - A, \epsilon, \delta)$-coloring, where $\epsilon A \geq 2$. Assume f is a random function from $[n]$ to $[n]$ with entropy at least $n \log n - Cn$, where $C \geq 1$. Then, for all $\alpha > 0$ with probability at least $1 - (n+1)\sqrt{\delta}$ over the choices of χ, $\chi \circ f$ is within $\alpha + \sqrt{\delta}$ of a (k', ϵ')-coloring with*

$$\begin{aligned} k' &= \log n - \log \frac{\alpha \epsilon^2 A}{2C}; \\ \epsilon' &= 3\epsilon. \end{aligned}$$

4.1 Lower bound for the bit game

In this section, we show the main result of this paper.

Theorem 6 *For all fixed k, $C^{B,\frac{k}{2}}(p_k) = \Omega(n \log^{(k-2)} n)$.*

Proof: Fix $k = 2t$. Let $c = \frac{1}{320} \log^{(k-2)} n$. We wish to show that there is no t-round protocol for p_{2t} where player B starts and the players send at most cn bits in each round. The proof uses the method of round elimination. We now define a slight generalization of the problem p_{2t}, $\tilde{p}_{2t-1}$, and prove a round elimination lemma for $\tilde{p}_{2t-1}$. The lower bound for p_{2t} then follows from a reduction of p_{2t} to $\tilde{p}_{2t-1}$.

Definition 4.7 (The problem $\tilde{p}_{2t-1}$) *Players A and B are given inputs f_A and f_B as before; in addition, B is now given a coloring $\chi : V_B \to \{0,1\}$. The players need to determine the color of the $(2t-1)$-st vertex reached by chasing pointers starting from the distinguished vertex $v_0 \in V_A$; that is, they must determine $\chi(f^{(2t-1)}(v_0))$.*

In the proof below, the inputs f_A and $\langle f_B, \chi \rangle$ given to A and B respectively will be randomly chosen according to some distribution. Let

$$I[f_A] \stackrel{\text{def}}{=} n \log n - H[f_A] \quad \text{and} \quad I[f_B] \stackrel{\text{def}}{=} n \log n - H[f_B].$$

Similarly, let $I[v_1] \stackrel{\text{def}}{=} I[f_A(v_0)] = \log n - H[f_A(v_0)]$. Here $I[f_A], I[f_B]$ and $I[v_1]$ measure the amount of information the protocol has about these input variables: the smaller these quantities, the harder the task. Thus, if f_A is chosen uniformly from the set of all possible values, then $I[f_A] = 0$. We also use the conditional version of this definition: $I[f_A \mid X] \stackrel{\text{def}}{=} n \log n - H[f_A \mid X]$. We have already seen how to quantify the randomness in a random coloring χ using the notion of (k, ϵ)-coloring. Recall that there k gives the min-entropy of the distribution that χ can color well. Just as we did for the other input random variables, we will replace the parameter k by $\log n - k$ while measuring the quality of χ. Thus, if we say χ has quality $\langle q, \epsilon \rangle$, then we mean χ is within ϵ of some $(\log n - q, \epsilon, \frac{1}{n^6})$-coloring. Note, however, that this time, the larger the parameter q, the harder the task will be for the protocol.

We shall show that if f_A and $\langle f_B, \chi \rangle$ are drawn from sufficiently 'rich' distributions, then there can be no protocol for $\tilde{p}_{2t-1}$ that uses at most cn bits per round, where B starts and both players know the answer after t rounds of communication.

Definition 4.8 (The predicate ϕ) *We say that $\phi_t(I, i, \langle q, \epsilon \rangle)$ holds if there exist distributions for f_A and $\langle f_B, \chi \rangle$ such that*

1. *$I[f_A], I[f_B] \le I, I[f_A(v_0)] \le i$ and χ has quality $\langle q, \epsilon \rangle$.*
2. *There is a protocol for $\tilde{p}_{2t-1}$ that uses at most cn bits per round, where B starts the communication and both players know the answer after t rounds of communication.*

Lemma 4.9 *(a) $\phi_1(I, i, \langle q, \epsilon \rangle)$ is false whenever $\epsilon < \frac{1}{4}$ and $\frac{i+1}{\epsilon} \le q$.*

(b) $\phi_t(I, i, \langle q, \epsilon\rangle) \Rightarrow \phi_{t-1}(I', i', \langle q', \epsilon'\rangle)$, *where*

$$\begin{aligned} I' &= 2I + 2cn; \\ i' &= \frac{I' 2^{2(i+1)}}{n}; \\ q' &= \log\left(\frac{\epsilon^3 qn}{4(I+n)}\right); \\ \epsilon' &= 6\epsilon. \end{aligned}$$

Proof:

Part (a). Suppose $\phi_1(I, i, \langle q, \epsilon\rangle)$ is true. In $\tilde{p}_1$, the player B knows the answer before any messages are exchanged. By the definition of ϕ_1, $I[f_A(v_0)] \le i$ and hence $H[f_A(v_0)] \ge \log n - i$. Let D denote the distribution of the random variable $f_A(v_0)$. This distribution D is within distance ϵ of distribution D' with min-entropy $\log n - \frac{i+1}{\epsilon}$ (see Lemma 7.2 below). We know that the coloring χ has quality $\langle q, \epsilon\rangle$. Suppose $\frac{i+1}{\epsilon} \le q$. Then,

$$\Pr_{\chi}[\chi \ \epsilon\text{-well-colors } D'] \ge 1 - \frac{1}{n^6} - \epsilon.$$

This implies that, for all large n,

$$\Pr_{\chi}[\chi \ 2\epsilon\text{-well-colors } D] \ge 1 - \frac{1}{n^6} - \epsilon > 0.$$

If we choose a χ that 2ϵ-well-colors D, then the neighborhood of v_0 will be two-colored and B cannot know the answer. This is a contradiction.

Part (b). Suppose $\phi_t(I, i, \langle q, \epsilon\rangle)$ is true. Fix input random variables f_A and $\langle f_B, \chi\rangle$ satisfying the first condition in the definition of ϕ_t. We want, now, to fix a value v_1 for $f_A(v_0)$, a value for χ and the first message m_1 of B in such a way that

(i) for the restricted random variables f'_A and f'_B, we have $I[f'_A], I[f'_B] \le I'$ and $I[f'_B(v_1)] \le i'$.

(ii) $\chi' = \chi \circ f'_A$, which is a 0-1 coloring of V_A has quality $\langle q', \epsilon'\rangle$.

Suppose, we succeed in this. Then,

$$\chi(f^{(2t-1)}(v_0)) = (\chi \circ f'_A)(f'^{(2t-3)}(v_1)) = \chi'(f'^{(2t-3)}(v_1))$$

where $f' = f'_A \cup f'_B$. Thus, if we use the same protocol assuming the first message is m_1 and v_1 is the starting vertex, then we a get a protocol for $\tilde{p}_{2t-3}$ for inputs f'_A, f'_B, χ' (with the roles of payers A and B are interchanged) with parameters $(I', i', \langle q'\epsilon'\rangle)$, thereby establishing part (b) of the lemma.

We now describe how v_1, χ and m_1 are chosen.

Finding good values for v_1**:** Since $I[f_A] \le I$, we have $I[f_A \mid f_A(v_0)] \le I + \log n \le I + cn$. We also know that $I[f_A(v_0)] \le i$. By Lemma 3.2, for at least $n/2^{2(i+1)}$ values v of $v_1 = f_A(v_0)$, the restriction $f_{A,v}$ (of f_A obtained by conditioning on the event $v_1 = v$) has the property

$I[f_{A,v}] \leq 2(I+cn)$. Thus, the set of good values of v_1, $G = \{v \in V_B : I[f_{A,v}] \leq 2(I+cn)\}$ has size at least $n/2^{2(i+1)}$. Later we will choose a value for v_1 from G. This will ensure that $I[f'_A] \leq 2(I+cn) = I'$.

Finding good values for χ: Note that χ is within distance ϵ of a $(\log n - q, \epsilon, \frac{1}{n^6})$-coloring. Thus, by Transfer Lemma, for any $v \in G$, with probability $1 - \frac{(n+1)}{n^3}$ over choices of χ, $\chi \circ f_{A,v}$ is within distance $2\epsilon + \frac{1}{n^3}$ of a $(\log n - q', 3\epsilon)$-coloring. Since, $|G| \leq n$, with probability $1 - \frac{n(n+1)}{n^3}$ over choices of χ, for *all* $v \in G$, $\chi' = \chi \circ f_{A,v}$ is within distance $2\epsilon + \frac{1}{n^3}$ of a $(\log n - q', 3\epsilon)$-coloring; call such a χ *nice*. Thus,

$$\Pr[\chi \text{ is nice}] \geq 1 - \frac{n(n+1)}{n^3} > \frac{1}{2}. \tag{2}$$

Claim 4.10 *If χ is nice, χ' has quality $\langle q', \epsilon' \rangle$.*

Proof: To show that χ' has quality $\langle q', \epsilon' \rangle$, we need to show that χ' is within distance ϵ' of a $(\log n - q', \epsilon', \frac{1}{n^6})$-coloring. As observed previously, χ' is within distance $2\epsilon + \frac{1}{n^3}$ of a $(\log n - q', 3\epsilon)$-coloring. For n large, χ' is within $\epsilon' = 6\epsilon$ of a $(\log n - q', 3\epsilon)$-coloring. Hence it suffices to show that any $(\log n - q', 3\epsilon)$-coloring is also a $(\log n - q', 6\epsilon, \frac{1}{n^6})$-coloring. By definition, if χ^* is a $(\log n - q', 3\epsilon)$-coloring, then for any distribution D with min-entropy $\log n - q'$,

$$\Pr_{\chi^*}[\chi^* \ (3\epsilon + t)\text{-well-colors } D] \geq 1 - \exp\left(\frac{-t^2 n}{2^{2q'+1}}\right).$$

Therefore,

$$\Pr_{\chi^*}[\chi^* \ 6\epsilon\text{-well-colors } D] \geq 1 - \exp\left(\frac{-9\epsilon^2 n}{2^{2q'+1}}\right).$$

Using the fact that $q' < q < \frac{1}{4}\log n$, for large n, we have,

$$\Pr_{\chi^*}[\chi^* \ \epsilon'\text{-well-colors } D] \geq 1 - \frac{1}{n^6}.$$

This proves the claim. $\square$

Fixing χ and m_1: Since $I[f_B] \leq I$, we have $I[f_B \mid \chi] \leq I + n$; thus for at least half the values of χ, the restricted input $f_{B,\chi}$ satisfies

$$I[f_{B,\chi}] \leq 2(I+n).$$

By (2), there exists a nice χ with this property. Fix one such χ. Since the first message M sent by B contains at most cn bits, $I[f_{B,\chi} \mid M] \leq 2I + 2n + cn$. Thus, we can fix a value m_1 for M so that for the resulting restriction on B's input, denoted f'_B,

$$I[f'_B] \leq 2I + 2cn = I'.$$

Fixing v_1: Finally, we want to choose a value of v for v_1 from the set of good values G, so that $I[f'_B(v)] \leq i'$. Now,

$$\sum_{v \in G} I[f'_B(v)] \leq \sum_{v \in V_B} I[f'_B(v)] \leq I'.$$

So, we can fix a value $v_1 \in G$ for $f_A(v_0)$ such that

$$I[f'_B(v)] \leq \frac{I'}{|G|} = \frac{I' 2^{2(i+1)}}{n} = i'.$$

This completes the proof of the lemma. $\square$

Claim 4.11 *If n is large enough and if $\phi_t(n, 0, \langle \frac{1}{4}\log n, 6^{-2t}\rangle)$ holds initially, then for all $j = 1, 2, \ldots, t-1$, $\phi_j(I_j, i_j, \langle q_j, \epsilon_j\rangle)$ holds where*

$$\begin{aligned} I_j &\leq (2^{t-j+1} - 1)2cn \\ i_j &\leq \text{TOWER}(t-j-1, 16, 16c) \\ q_j &= \frac{1}{4}\log^{(t-j+1)} n \\ \epsilon_j &= 1/6^{t+j}. \end{aligned}$$

Proof: We prove the claim by a backward induction on j.

Base case: $j = t-1$

$$\begin{aligned} I_{t-1} &= 2I_t + cn = 2n + cn \leq 6cn. \\ i_{t-1} &= 8(c+1) \leq \text{TOWER}(0, 16, 16c) = 16c. \\ q_{t-1} &= \log\left(\frac{\epsilon^3 q_t n}{8n}\right) \leq \frac{1}{4}\log^{(2)} n. \\ \epsilon_{t-1} &= 6\epsilon_t = \frac{1}{6^{2t-1}}. \end{aligned}$$

Induction step: We will show that if $\phi_j(I_j, i_j, \langle q_j, \epsilon_j\rangle)$ holds, then $\phi_{j-1}(I_{j-1}, i_{j-1}, \langle q_{j-1}, \epsilon_{j-1}\rangle)$ holds. Using Lemma 4.9,

$$\begin{aligned} I_{j-1} &= 2I_j + 2cn \\ &\leq 2(2^{t-j+1} - 1)2cn + 2cn \\ &= (2^{t-(j-1)+1} - 1)2cn. \\ i_{j-1} &= \frac{I_{j-1} 2^{2(i_j+1)}}{n} \\ &\leq (2^{t-(j-1)+1} - 1)2c2^{2(i_j+1)} \\ &\leq (2^{t-(j-1)+1} 4c)4^{i_j} \\ &\leq (4^{\text{TOWER}(t-j-1,16,16c)})4^{\text{TOWER}(t-j-1,16,16c)} \\ &= \text{TOWER}(t-j, 16, 16c). \\ q_{j-1} &= \log\left(\frac{\epsilon^3 q_j n}{4(I_j + cn)}\right) \\ &\leq \log\left(\frac{\epsilon^3 q_j}{4}\right) \\ &\leq \frac{1}{4}\log^{(t-j+1} n. \end{aligned}$$

$$\begin{aligned} \epsilon_{j-1} &= 6\epsilon_j \\ &= 6\frac{1}{6^{t+j}} \\ &= \frac{1}{6^{t+j-1}}. \end{aligned}$$

This completes the induction step. □

To prove the lower bound for p_{2t} we reduce it to the problem $\tilde{p}_{2t-1}$.

Lemma 4.12 *If there is a t-round protocol for p_{2t} that uses at most cn bits per round, then $\phi_t(0, 0, \langle \frac{1}{4} \log n, 6^{-2t} \rangle)$ holds.*

Proof: Initially, f_A, f_B and $f_A(v_0)$ have full entropy and hence zero information. The coloring χ has uniform distribution on $\{0,1\}^n$. It follows from the method of bounded differences (see Lemma 1.2 of McDiarmid [10]) that

$$\Pr[|\sum_{i=1}^{n} D(i)\chi(i) - \frac{1}{2}| \geq \epsilon] \leq 2e^{\frac{-2\epsilon^2}{\max_i D(i)}},$$

where D is any probability distribution on $[n]$. In particular, if D has min-entropy $\frac{3}{4}\log n$, then $D(i) \leq n^{-\frac{3}{4}}$ for all i. Hence, for all large n,

$$\Pr[|\sum_{i=1}^{n} D(i)\chi(i) - \frac{1}{2}| \geq \frac{1}{6^{2t}}] \leq 2e^{\frac{-2n^{\frac{3}{4}}}{6^{4t}}} \leq \frac{1}{n^6}.$$

Therefore, χ has quality $\langle \frac{1}{4}\log n, 6^{-2t} \rangle$. □

Claim 4.13 *$\forall k$, $\forall n$ large enough, $c = \Omega(\log_2^{(k-2)} n)$.*

Proof:
$$\phi_1((2^t - 1)2cn, \textsc{tower}(t-2, 16, 16c), \langle \frac{1}{4}\log^{(t)} n, 1/6^t \rangle) \tag{3}$$

holds as a consequence of Claim 4.11 and Lemma 4.12. Now, by part (a) of Lemma 4.9, $\frac{1}{4}\log^{(t)} n \leq \textsc{tower}(t-2, 16, 16c)$. We now show that this is not possible if $c = \frac{1}{320}\log^{(k-2)} n$. As shown in (1), for large n (such that $\log_2^{(k-2)} n \geq 15$),

$$\log_{16}^{(k-2)} n \geq \frac{1}{5}\log_2^{(k-2)} n.$$

This implies that

$$\frac{1}{64}\log_{16}^{(k-2)} n \geq \frac{1}{320}\log_2^{(k-2)} n.$$

But, (3) does not hold if $c = \frac{1}{64}\log_{16}^{(k-2)} n = \frac{1}{64}\log_{16}^{(2t-2)} n$. This proves the claim. □

The theorem now follows from the claim. ■

5 The s-pointer game g_k^s

In this section, we consider the generalized pointer game where we follow s paths in parallel instead of one.

Definition 5.1 *Let $v_{01}, \ldots, v_{0s} \in V_A$. The function we will be interested in computing is $g_k^s : F_A \times F_B \to V^s$ defined by $g_k^s(f_A, f_B) = (f^{(k)}(v_{01}), \ldots, f^{(k)}(v_{0s}))$ where $f^{(k)}$ is as in Definition 3.1.*

Given n and s, let us define the following function: $h^{(0)}(n,s) = \log n$. $h^{(k+1)}(n,s) = \log h^{(k)}(n,s) + \log s$.

Theorem 7 *For all fixed k, $C^{B,k}(g_k^s) = O(nh^{(k-2)}(n,s))$.*

Proof:

Round 1: B sends the last $h^{(k-2)}(n,s)$ bits of $f_B(v)$ for every $v \in V_B$.

Round 2: A now knows the last $h^{(k-2)}(n,s)$ bits of $f_B(v_{11}), \ldots, f_B(v_{1s})$, where $v_{11} = f_A(v_{01}), \ldots, v_{1s} = f_A(v_{0s})$. A sends $v_{11}, \ldots, v_{1s}$ and the last $h^{(k-3)}(n,s)$ bits of $f_A(v)$ for every possible vertex v in A's list that $v_{11}, \ldots, v_{1s}$ could be pointing to.

Round 3: B now knows the last $h^{(k-3)}(n,s)$ bits of $f_A(v_{21}), \ldots, f_A(v_{2s})$, where $v_{21} = f_B(v_{11}, \ldots, v_{2s} = f_B(v_{1s})$. B sends $v_{21}, \ldots, v_{2s}$ and the last $h^{(k-4)}(n,s)$ bits of $f_B(v)$ for every possible vertex v in B's list that $v_{21}, \ldots, v_{2s}$ could be pointing to.

⋮

Round k: If k is even, A will now know $v_{k1}, \ldots, v_{ks}$ by this time and A sends this information over to B. If k is odd, B would know $v_{k1}, \ldots, v_{ks}$ by this time, and B sends this information over to A.

It is easy to see that this protocol uses $O(nh^{(k-2)}(n,s))$ bits of communication. ∎

6 Monotone constant depth hierarchy

Let $L^{k-1}(H_k)$ be the minimum size of a monotone formula of depth $k-1$ (unbounded fan-in) that computes H_k.

Theorem 8 (Nisan and Wigderson [11]) $C^{B,k}(g_k) \le \log L^{k-1}(H_k)$.

Now using the fact that $C^{B,k}(g_k) = \Omega(n \log^{(k-1)} n)$, we get the following corollary:

Corollary 6.1 *Any monotone circuit of depth $k-1$ (unbounded fan-in) for H_k requires size $2^{\Omega(n \log^{(k-1)} n/k)}$.*

Let $L_s^k(Q_k^s)$ be the minimum size of a monotone formula of depth k (unbounded fan-in) with $O(s)$ nondeterministic bits that computes Q_k^s.

Theorem 9 (Klauck [7]) $C^{B,k+1}(g_{k+1}) - O(n) \le \log L_s^k(Q_k^s)$.

So, we have:

Corollary 6.2 *Any monotone circuit of depth k (unbounded fan-in) with $O(s)$ nondeterministic bits for Q_k^s requires size $2^{\Omega(n \log^{(k)} n/k)}$.*

7 The Transfer Lemma

In this section, we give a detailed proof of the transfer lemma. We begin by recalling the statement of the transfer lemma (see Lemma 4.6).

The transfer lemma: Suppose χ is a random $(\log n - A, \epsilon, \delta)$-coloring, where $\epsilon A \geq 2$. Assume f is a random function from $[n]$ to $[n]$ with entropy at least $n \log n - Cn$, where $C \geq 1$. Then, for all $\alpha > 0$ with probability at least $1 - (n+1)\sqrt{\delta}$ over the choices of χ, $\chi \circ f$ is within $\alpha + \sqrt{\delta}$ of a (k', ϵ')-coloring with

$$\begin{aligned} k' &= \log n - \log \frac{\alpha \epsilon^2 A}{2C}; \\ \epsilon' &= 3\epsilon. \end{aligned}$$

How do we prove the transfer lemma? Fix χ and f as in the description of the transfer lemma. We think of f as an element of $[n]^n$. Now the distribution of f gives rise to an n-ary *probability tree.* The n edges coming out of root of the tree correspond to the n possible values for $f(1)$. Each edge is labeled with the corresponding probability. In general, the n edges coming out of a node of the tree at level i, correspond to the n different possibilities for $f(i)$, and labeled with the appropriate conditional probabilities (see below for the precise definition). This defines a probability distribution on the leaves of the tree which is precisely the probability distribution of f.

We first observe that on most paths in this tree, there must be lots of nodes of high entropy (by the entropy of a node, we mean the entropy of the distribution at that node). We now use χ to color the underlying set $[n]$, so that each of the nodes of high entropy gets a balanced coloring (that is the choices out of this node that get colored 0 are roughly half).

Once, we fix the coloring χ, the probability tree now induces a distribution on $\{0,1\}^n$. We need to show that this distribution is capable of coloring sets of size $2^{k'}$, where $k' = \log n - \log a$ with high probability. For this we use the *method of bounded differences.* Unfortunately, this application is not straightforward. The difficulty arises because the good nodes of the probability tree are not always on the same level (as we range over the paths of the tree).

There is one more difficulty that arises from the use of the probability tree. Although we state that there are many good nodes on most paths of the tree, we cannot just absorb the bad paths in the probability of error. This is because, we will need to color several sets simultaneously and if we incur this error for each of them, we will be left with nothing in the end. So, we need to observe that these bad paths constitute bad events that are common to all the sets that we wish to color. This is formalized as follows. We show that the probability tree is close (here *distance is measured as defined in Definition 4.2*) to another tree, all of whose paths have many good nodes.

Definition 7.1 (The probability tree) *Suppose D is a distribution on A^m, where $|A| = n$ and m is a positive integer. Let X be a random variable with distribution D. We associate an n-ary tree τ of height m with the distribution D. The root r of τ will have n children, one for each possibility of X_1. The n outgoing edges will be labeled by the corresponding probabilities $p_1(r), \ldots, p_n(r)$. Suppose w is the child of the root corresponding to the possibility $X_1 = a \in A$. Then, the subtree rooted at w is the probability tree for distribution of $X_2, \ldots, X_n$ subject to the condition $X_1 = a$. The leaves of the tree τ will be denoted by $L(\tau)$. Note that the probabilities on the the edges impose a distribution on the leaves of the tree, and this is exactly the same as*

D. For $i \geq 1$, let τ_i be the probability tree corresponding to uniform distribution on $[n]^i$. The distance between two trees is the distance between the corresponding distributions.

We refer to the internal vertices of this tree as nodes. *For a node v, let D_v be the distribution on the set A generated by the probabilities labeling the edges leaving v. We say that a node v of τ is ϵ-well-colored by the function $\chi : [n] \to \{0,1\}$, if D_v is ϵ-well-colored by χ.*

7.1 Approximating distributions

Lemma 7.2 (Entropy vs. min-entropy) *Let D be a distribution on $[n]$ with entropy at least $\log n - A$. Then, for all $\epsilon > 0$, D is within ϵ of a distribution with min-entropy $\log n - \frac{A+1}{\epsilon}$.*

Proof: Let $B = (A+1)/\epsilon$, and let $S = \{i : D(i) > \frac{2^B}{n}\}$. Let $q = \Pr[S]$. We will show that

$$q \leq \epsilon. \tag{4}$$

Then, we can obtain a distribution D' from D by redistributing the excess probability on S among other elements, so that

1. $D'(i) \leq 2^B/n$ for all $i \in [n]$, that is, D' has min-entropy $\log n - \frac{A+1}{\epsilon}$; and
2. $D(i) > D'(i) \Rightarrow i \in S$. That is, D' was obtained from D by moving the excess probability out of S. We now show that this can always be done. Using the fact that $B \geq 1$ and D is a probability distribution, we have

$$\sum_i \left(D(i) - \frac{2^B}{n} \right) \leq 0.$$

and hence

$$\sum_{i \in S} \left(D(i) - \frac{2^B}{n} \right) \leq \sum_{i \notin S} \left(\frac{2^B}{n} - D(i) \right).$$

In particular, we get $d(D, D') \leq \Pr[S] < \epsilon$.

We only need to show (4). If $q = 0$, we are done. Otherwise, we have,

$$H[D] \;=\; \sum_{i \in S} D(i) \log \frac{1}{D(i)} + \sum_{i \in [n]-S} D(i) \log \frac{1}{D(i)} \;\geq\; \log n - A.$$

Since, $D(i) > 2^B/n$ for $i \in S$, we have

$$q(\log n - B) + (1-q) \sum_{i \in [n]-S} \frac{D(i)}{1-q} \log \frac{1}{D(i)} \geq \log n - A.$$

[Note that $q < 1$, because otherwise we would have $H[D] < \log n - A$, contradicting the assumption.] Using Jensen's inequality for the second term on the left, we get

$$q(\log n - B) + (1-q) \log \frac{n - |S|}{1-q} > \log n - A.,$$

This implies that,

$$q(\log n - B) + (1-q)[\log n - \log(1-q)] > \log n - A.$$

The inequality (4) follows easily from this. ∎

Lemma 7.3 *Let X and X' be random variables taking values in $[n]$, with respective distributions D and D'. Suppose, $d(D, D') \le \epsilon$. Let $f : [n] \to \{0,1\}$. Then,*

$$|\,\mathrm{E}[f(X)] - \mathrm{E}[f(X')]| \le \epsilon.$$

Proof: Now,

$$|\,\mathrm{E}[f(X)] - \mathrm{E}[f(X')]| = |\Pr[f(X) = 1] - \Pr[f(X') = 1]|.$$

The claim, then, follows from the property of the distance functions (see Proposition 4.3). ■

Corollary 7.1 *Let X and X' be random variables taking values in $[n]$, with respective distributions D and D', such that $d(D, D') \le \epsilon$. For all distributions $\mathcal{F}$ on $\{0,1\}^n$, and all $a, t \ge 0$,*

$$\Pr_{f\in\mathcal{F}}[|E_X[f(X)] - a| \ge t + \epsilon] \;\le\; \Pr_{f\in\mathcal{F}}[|E_{X'}[f(X')] - a| \ge t].$$

Lemma 7.4 *Let P be a probability distribution on $[n]^n$ such that $H[P] \ge n \log n - Cn$. Consider the probability tree τ corresponding to P. We say that an internal node v of τ is* good *if $H[v] \ge \log n - T$; an internal node is* bad *if it is not good. Then,*

$$\Pr_{\ell\in L(\tau)}[\textit{The number of bad nodes on the path to } \ell \ge \frac{n}{S}] \le \frac{CS}{T}.$$

Proof: Let $p = \Pr_{\ell\in L(\tau)}[\text{The number of bad nodes on the path to } \ell \ge \frac{n}{S}]$. Now,

$$E_{\ell\in L(\tau)}[\sum_{v\in\mathsf{path}(\ell)} H[v]] = H[P].$$

Thus,

$$p[(n - \frac{n}{s})\log n + \frac{n}{s}(\log n - T)] + (1-p) n \log n \ge n \log n - Cn.$$

On simplification, this gives $p \le CS/T$. ■

Lemma 7.5 *Let τ be a probability tree corresponding to a distribution on $[n]^m$. Let $\mathcal{G}$ be a set of distributions on $[n]$; assume that the uniform distribution on $[n]$ is in $\mathcal{G}$. A node v of τ is* good *if $D_v \in \mathcal{G}$. (In the application, $\mathcal{G}$ will be the set of 'sufficiently random' distributions, and v will be a a node where the choices are sufficiently random.)*

Suppose for some $k (1 \le k \le m)$ and $\epsilon \ge 0$,

$$\Pr_{\ell\in L(\tau)}[\mathsf{path}(\ell) \textit{ has at least } k \textit{ good nodes}] \;\ge\; 1 - \epsilon.$$

Then, there is a probability tree τ' such that

(a) $\Pr_{\ell\in L(\tau')}[\mathsf{path}(\ell)$ *has at least k good nodes*$] = 1$*;*

(b) $d(T, T') \le \epsilon$.

Proof: We say that a leaf is good if the path leading to it has at least k good nodes. If for some node v of τ, the subtree rooted at v has no good leaves, then we replace the subtree by an appropriate copy of τ_i. After this has been done for all nodes v of the tree τ, the resulting tree τ' will have the required properties. The formal proof proceeds by induction on m.

For $m = 1$, we have two cases. If $\epsilon = 1$, we can take τ' to be the tree τ_1, corresponding to the uniform distribution on $[n]$. Otherwise, the root of τ must be a good node, then we take τ' to be τ.

If $m > 1$, then again we have two cases. If $\epsilon = 1$, that is no leaf of τ is good, then we take $\tau' = \tau_m$. On the other hand, if $\epsilon < 1$, then consider the root r of T. Let $v_1, \ldots, v_n$ be the n children of r, with $p_1, \ldots, p_n$ the respective probabilities. For $i = 1, \ldots, n$, let τ_{v_i} be the subtree rooted at v_i. Let

$$k' \stackrel{\text{def}}{=} \begin{cases} k-1 & \text{if the root is good} \\ k & \text{otherwise.} \end{cases}$$

Note that if $\ell \in L(\tau)$ is good, then the path leading to it must have at least k' good nodes after the root. Define ϵ_i as follows

$$\epsilon_i \stackrel{\text{def}}{=} 1 - \Pr_{\ell \in L(\tau_{v_i})}[\ell \text{ has at least } k' \text{ good nodes on its path to } v_i.].$$

Note that, $\sum_{i=1}^{n} p_i \epsilon_i \leq \epsilon$. We, now apply the induction hypothesis to each τ_{v_i} with (k', ϵ_i) instead (k, ϵ) and obtain τ'_{v_i} such that

$$\begin{aligned} \Pr_{\ell \in L(\tau'_{v_i})}[\ell \text{ has at least } k' \text{ good nodes on its path to } v_i] &= 1; \\ d(\tau_{v_i}, \tau'_{v_i}) &\leq \epsilon_i. \end{aligned}$$

Now, to obtain our tree τ', we replace each τ_{v_i} by τ'_{v_i}. Then, clearly,

$$\Pr_{\ell \in L(\tau')}[\text{ there are at least } k \text{ good nodes on the path to } \ell] = 1.$$

Furthermore,

$$d(\tau, \tau') \leq \sum_{i=1}^{n} p_i d(\tau_{v_i}, \tau'_{v_i}) \leq \sum_{i=1}^{n} p_i \epsilon_i \leq \epsilon.$$

■

7.2 Proof of the main lemma

Proof: In the following,

$$T = \epsilon A - 1; \qquad \text{and} \qquad S = \frac{\alpha T}{C} = \frac{\alpha(\epsilon A - 1)}{C}.$$

Let Γ be the probability tree corresponding to f.

Step 1. We say that a node v of Γ is good if $H[D_v] \geq \log n - T$. If v is a good node, then by Lemma 7.2, D_v is within ϵ of a distribution with min-entropy $\log n - \frac{T+1}{\epsilon} = \log n - A$. For a distribution D on $[n]$ we say that 'D is well-colored by χ' if

$$|\sum_{i=1}^{n} D_v(i)\chi(i) - \frac{1}{2}| \leq 2\epsilon.$$

Then, by Lemma 7.3, since χ is a $(\log n - A, \epsilon, \delta)$-coloring, we have

$$\Pr_{\chi}[D_v \text{ is not well-colored by } \chi] \leq \delta. \tag{5}$$

Step 2. For $\ell \in L(\Gamma)$, let $\mathcal{E}(\ell, \chi)$ denote the following event: *some good node of* $\mathsf{path}(\ell)$ *is not well-colored by* χ. Then (5) implies that

$$\mathop{\mathrm{E}}_{\chi}[\Pr_{\ell \in L(\Gamma)}[\mathcal{E}(\ell, \chi)]] \leq n\delta.$$

By Markov's inequality,

$$\Pr_{\chi}[\Pr_{\ell \in L(\Gamma)}[\mathcal{E}(\ell, \chi)] \geq \sqrt{\delta}] \leq n\sqrt{\delta}.$$

Let $\mathcal{E}_0(\chi)$ be the event $\Pr_{\ell \in L(\Gamma)}[\mathcal{E}(\ell, \chi)] \leq \sqrt{\delta}$. Then,

$$\Pr_{\chi}[\mathcal{E}_0(\chi)] \geq 1 - n\sqrt{\delta}.$$

Let $\mathcal{E}_1(\chi)$ denote the event 'χ well-colors the uniform distribution on $[n]$'. Since the uniform distribution has min-entropy $\log n \geq k$, we have

$$\Pr_{\chi}[\mathcal{E}_1(\chi)] \geq 1 - \delta.$$

Thus, with probability $1 - n\sqrt{\delta} - \delta \geq 1 - (n+1)\sqrt{\delta}$, over choices of χ, both $\mathcal{E}_0(\chi)$ and $\mathcal{E}_1(\chi)$ hold. Fix a $\hat{\chi}$ for which the two events hold. In the next three steps we will show that $\hat{\chi} \circ f$ is close to a coloring distribution with parameters claimed above.

Step 3. By Lemma 7.4,

$$\Pr_{\ell \in L(\Gamma)}[\mathsf{path}(\ell) \text{ has at least } \tfrac{n}{S} \text{ bad nodes}] = \frac{CS}{T} = \alpha.$$

Since $\mathcal{E}_0(\hat{\chi})$ holds, we have

$$\Pr_{\ell \in L(\Gamma)}[\mathsf{path}(\ell) \text{ has at least } \tfrac{n}{S} \text{ nodes that are not well-colored}] \leq \alpha + \sqrt{\delta}. \tag{6}$$

Let $\mathcal{G} = \{D : D \text{ is a distribution on } [n] \text{ that is well-colored by } \hat{\chi}\}$. In terms of $\mathcal{G}$, (6) says

$$\Pr_{\ell \in L(\Gamma)}[|\{v \in \mathsf{path}(\ell) : D_v \in \mathcal{G}\}| \leq n(1 - \frac{1}{S})] \leq \alpha + \sqrt{\delta}.$$

Note that since $\mathcal{E}_1(\hat{\chi})$ holds, the uniform distribution on $[n]$ belongs to $\mathcal{G}$. Then, by Lemma 7.5 Γ is within $\alpha + \sqrt{\delta}$ of a tree Γ' such that

$$\Pr_{\ell \in L(\Gamma')}[|\{v \in \mathsf{path}(\ell) : D_v \in \mathcal{G}\}| \leq n(1 - \frac{1}{S})] = 0.$$

Step 4. Let f' be the random function (from $[n]$ to $[n]$) corresponding to Γ'. By Lemma 7.6, $\hat{\chi} \circ f'$ is a (k', ϵ')-coloring, where

$$\begin{aligned} k' &= \log n - \log \frac{\epsilon\alpha(\epsilon A - 1)}{C} \leq \log n - \log \frac{\alpha\epsilon^2 A}{C}; \\ \epsilon' &= 3\epsilon. \end{aligned}$$

Step 5. Note that if the distributions of f and f' are within β of each other, then the distributions of $\hat{\chi} \circ f$ and $\hat{\chi} \circ f'$ are also within β of each other (see Proposition 4.3). ■

7.3 Bounds for well-colored trees

Lemma 7.6 *Let f be a random function from $[n]$ to $[n]$. We will think of f as a probability distribution on $[n]^n$. Let Γ be the probability tree corresponding to f. Let $\chi : [n] \to \{0,1\}$. We say that a node v of Γ is* well-colored *by χ if*

$$|\mathop{\mathrm{E}}_{i \in D_v}[\chi(i)] - \frac{1}{2}| \leq \delta.$$

For $\ell \in L(\tau)$ with $(v_1, v_2, \ldots, v_n)$ as the nodes of τ on the path from the root to ℓ, let

$$\mathsf{WC}(\ell) \stackrel{\mathrm{def}}{=} \{i \in [n] : \textit{the vertex } v_i \textit{ is well-colored}\}.$$

Suppose, Γ and D satisfy the following conditions.

1. $\Pr_{\ell \in L(\Gamma)}[|\mathsf{WC}(\ell)| \geq (1-\epsilon)n] = 1.$

2. *D has min-entropy k.*

Let X be the random variable taking values in $\{0,1\}^n$ generated by f and χ (that is $X_i \stackrel{\mathrm{def}}{=} \chi(f(i))$, for $i = 1, \ldots, n$). Then, for all $t \geq 0$,

$$\Pr_X[|\sum_{i=1}^n D(i)X_i - \frac{1}{2}| \geq \delta + \frac{\epsilon n}{2^k} + t] \leq 2\exp(-\frac{t^2 2^{2k}}{2n}). \tag{7}$$

Proof: Let v be a node of Γ at distance i from the root. Now, all leaves in the subtree rooted at v give rise to the same values for $f(1), \ldots, f(i-1)$, and hence for $\chi(f(1)), \chi(f(2)), \ldots, \chi(f(i-1))$. We refer to $\langle \chi(f(1)), \chi(f(2)), \ldots, \chi(f(i-1)) \rangle$ as the *coloring history* of v.

Now, fix $t \geq 0$. We wish to bound the left hand side of (7). We may assume that if two nodes v_1 and v_2 of Γ have the same coloring history, then the subtrees of Γ rooted at v_1 and v_2 are identical. This can be justified by the following observation: We consider all nodes v with the same coloring history (such that there is non-zero probability of reaching any of them). We consider the subtree rooted at each node and calculate the conditional probability

$$\Pr_X[|\sum_{i=1}^n D(i)X_i - \frac{1}{2}| \geq \delta + \frac{\epsilon n}{2^k} + t] \tag{8}$$

in that subtree. We then replace all the subtrees by the one in which this conditional probability is the maximum. Note that this only increases the probability we are trying to calculate.

Let τ be the probability tree corresponding to X. We say that a node v of τ is balanced if $\frac{1}{2} - \delta \leq p_1(v) \leq \frac{1}{2} + \delta$. For $\ell \in L(\tau)$ with $(v_1, v_2, \ldots, v_n)$ as the nodes of τ on the path from the root to ℓ. Let

$$\mathsf{Bal} \stackrel{\text{def}}{=} \{i \in [n] : \text{the vertex } v_i \text{ is balanced}\}.$$

Now, since $|\mathsf{WC}(\ell)| \geq (1-\epsilon)n$ with probability 1 for $\ell \in L(\Gamma)$, we have $|\mathsf{Bal}(\ell)| \geq (1-\epsilon)n$ with probability 1 for $\ell \in L(\tau)$. Since $D(i) \leq 1/2^k$ for all $i \in [n]$, this implies that

$$\Pr_{\ell \in L(\tau)}[\sum_{i \in \mathsf{Bal}(\ell)} D(i) \geq 1 - \frac{\epsilon n}{2^k}] = 1.$$

The inequality (7) now follows from Lemma 7.7 below. ■

Lemma 7.7 (Tail bounds) *Let X be a random variable taking values in $\{0,1\}^n$. Let τ be the corresponding probability tree. We say that a node v of τ is balanced if $\frac{1}{2} - \delta \leq p_1(v) \leq \frac{1}{2} + \delta$. For $\ell \in L(\tau)$ with $(v_1, v_2, \ldots, v_n)$ as the nodes of τ on the path from the root to ℓ. Let*

$$\mathsf{Bal}(\ell) \stackrel{\text{def}}{=} \{i \in [n] : \textit{the vertex } v_i \textit{ is balanced}\}.$$

Let D be a distribution on $[n]$, satisfying the following two conditions.

1. $\Pr_{\ell \in L(\tau)}[\sum_{i \in \mathsf{Bal}(l)} D(i) \geq 1 - \epsilon] = 1.$

2. *$D(i) \leq \alpha$ for all $i \in [n]$; that is, the min-entropy of D is $\log 1/\alpha$.*

Then, for all $t \geq 0$,

$$\Pr_{\ell \in L(\tau)}[|\sum_{i=1}^{n} D(i)X_i - \frac{1}{2}| \geq \epsilon + \delta + t] \leq 2\exp(-\frac{t^2}{2\alpha^2 n}). \tag{9}$$

Proof: The bound (9) will follow if we show the following:

$$\Pr_{\ell \in L(\tau)}[\sum_{i=1}^{n} D(i)X_i - \frac{1}{2} \geq \epsilon + \delta + t] \leq \exp(-\frac{t^2}{2\alpha^2 n}); \tag{10}$$

$$\Pr_{\ell \in L(\tau)}[\frac{1}{2} - \sum_{i=1}^{n} D(i)X_i \geq \epsilon + \delta + t] \leq \exp(-\frac{t^2}{2\alpha^2 n}). \tag{11}$$

Note that (11) follows from (10) by considering the random variable Y (taking values in $\{0,1\}^n$) defined by $Y_i = 1 - X_i$, for $i = 1, \ldots, n$. We may thus restrict our attention to (10). Fix $t \geq 0$; then, by Lemma 7.8, we may assume that τ and D satisfy the following conditions.

2'. $\Pr_{\ell \in L(\tau)}[1 - \epsilon \leq \sum_{i \in \mathsf{Bal}(\ell)} D(i) \leq 1 - \epsilon + \alpha] = 1.$

3. For all nodes v in τ, $p_1(v) = \frac{1}{2} + \delta$ or 1.

Now, for such 'regularized' distributions τ and D, which satisfy these stronger conditions, we can apply the method of bounded differences (McDiarmid [10]) and obtain inequality (10). Details appear in Lemma 7.9 below. ■

7.4 Regularizing the distributions

Lemma 7.8 *Let τ and D be distributions as in Lemma 7.7. For each $T \geq 0$, there is a probability tree τ' such that*

(a) For all nodes v of τ', $p_1(v) = \frac{1}{2} + \delta$ or 1.

(b) $\Pr_{\ell \in L(\tau')}[1 - \epsilon \leq \sum_{i \in \mathsf{Bal}(\ell)} D(i) \leq 1 - \epsilon + \alpha] = 1$. (Here balanced *and $Bal(\ell)$ are defined with respect to the tree τ').*

(c) Let $Y = Y_1, \ldots, Y_n$ be the random variable associated with the tree τ'. Then,

$$\Pr_X[\sum_{i=1}^{n} D(i) X_i \geq T] \;\leq\; \Pr_Y[\sum_{i=1}^{n} D(i) Y_i \geq T].$$

Proof:

Enforcing conditions (a) and (c). Let τ^* be a tree for which condition 1 of the previous lemma and condition (c) above hold, and for which the maximum number of nodes v satisfy the condition $p_1(v) = \frac{1}{2} + \delta$ or 1. (Clearly, such a tree exists, because τ itself meets these conditions.) We claim that, then τ^* satisfies (a). For, otherwise, let v be the node farthest from the root such that $p_1(v)$ is neither $\frac{1}{2} + \delta$ nor 1. Let τ_v be the subtree of τ^* rooted at v. We will replace τ_v by another tree τ'_v such that the condition (c) continues to hold for the resulting tree. In τ'_v, $p_1(v)$ will be $\frac{1}{2} + \delta$ if v was balanced in τ^* and 1 if it was not balanced. Furthermore, the number of nodes $w \neq v$ with $p_1(w) \neq \frac{1}{2} + \delta$ or 1 in τ_v and τ'_v will be the same. But then, this new tree contradicts the choice of τ'. It, thus, remains only to describe how the tree τ'_v is obtained.

Suppose v appears at level i (from the top) of τ^*. Let v_0 be the child of v corresponding to the value $X_i = 0$, and v_1 the child corresponding to the value $X_i = 1$; let τ_0 and τ_1 be the subtrees rooted at v_0 and v_1 respectively. Let

$$\alpha \stackrel{\text{def}}{=} \Pr_{\ell \in L(\tau_0)}[\sum_{j=1}^{n} D(j) X_j \geq T];$$

$$\beta \stackrel{\text{def}}{=} \Pr_{\ell \in L(\tau_1)}[\sum_{j=1}^{n} D(j) X_j \geq T].$$

Now, if $p_0(v) = 1$ or $\alpha \geq \beta$, then replace τ_1 with a copy of τ_0 (otherwise, τ_0 and τ_1 remain the same). If v was balanced then set $p_1(v) = \frac{1}{2} + \delta$, and if v was unbalanced, then set $p_1(v) = 1$. The resulting tree obtained from τ_v is τ'_v. It can be checked that this new tree satisfies the requirements stated above.

Enforcing condition (b). On each path $v_1, v_2, \ldots$ (with non-zero probability) in τ', consider the first i such that

$$\sum_{j=1, v_j \text{ balanced}}^{i} D(j) \geq 1 - \epsilon.$$

By condition 1 of the previous lemma, such an i exists, and by condition 2

$$\sum_{j=1, v_j \text{ balanced}}^{i+1} D(j) \leq 1 - \epsilon + \alpha.$$

Replace the subtree of τ rooted at v_i by the tree of height $n-i+1$, such that $p_1(v)=1$ for all its nodes. After this has been done for all paths with non-zero probabilities, condition (b) will be satisfied. Also, conditions (a) and (c) continue to hold. ■

7.5 Tail bounds for regularized distributions

Lemma 7.9 *Let X be a random variable taking values in $\{0,1\}^n$; let τ be the probability tree corresponding to X. Let D be a distribution on $[n]$. Suppose τ and D satisfy the following conditions.*

1. *For all nodes v of τ, $p_1(v) = \frac{1}{2}+\delta$ or 1; we say that v is a* balanced *if $p_1(v)=\frac{1}{2}+\delta$.*
2. *$D(i) \leq \alpha$ for all $i \in [n]$; that is, D has min-entropy $\log 1/\alpha$.*
3. *For $\ell \in L(\tau)$ with $(v_1, v_2, \ldots, v_n)$ as the nodes of τ on the path from the root to ℓ. Let*
$$\mathsf{Bal}(\ell) \stackrel{\text{def}}{=} \{i \in [n] : \textit{the vertex } v_i \textit{ is balanced}\}.$$
There exist $\epsilon, \beta \geq 0$, such that
$$\Pr_{\ell \in L(\tau)}[1-\epsilon \leq \sum_{i \in \mathsf{Bal}(l)} D(i) \leq 1-\epsilon+\beta] = 1.$$

Then,
$$\Pr_X[\sum_{i=1}^n D(i)X_i - \frac{1}{2} \geq \epsilon+\delta+t] \leq \exp(-\frac{2t^2}{n(\alpha+\beta)^2}). \tag{12}$$

Proof: Let $f(X) = \sum_{i=1}^n D(i)X_i$. First, we will show that $\mathrm{E}_X[f(X)] \leq \frac{1}{2}+\epsilon+\delta$. Next, we will show, using the method of bounded differences, that the probability that $f(X)$ differs from its expectation by t is at most the right hand side of (12).

Bounding the expectation. It can be verified that

$$\begin{aligned}
\mathop{\mathrm{E}}_X[f(X)] &= \mathop{\mathrm{E}}_{\ell \in L(\tau)}[(\frac{1}{2}+\delta)\sum_{i \in \mathsf{Bal}(\ell)} D(i) + \sum_{i \notin \mathsf{Bal}(\ell)} D(i)] \\
&= (\frac{1}{2}+\delta)(1 - \mathop{\mathrm{E}}_{\ell \in L(\tau)}[\sum_{i \notin \mathsf{Bal}(\ell)} D(i)]) + \mathop{\mathrm{E}}_{\ell \in L(\tau)}[\sum_{i \notin \mathsf{Bal}(\ell)} D(i)] \\
&= \frac{1}{2}+\delta+(\frac{1}{2}-\delta)\mathop{\mathrm{E}}_{\ell \in L(\tau)}[\sum_{i \notin \mathsf{Bal}(\ell)} D(i)] \qquad (13)
\end{aligned}$$

Since $\sum_{i \in \mathsf{Bal}(\ell)} D(i) \geq 1-\epsilon$ with probability 1, we have $\sum_{i \notin \mathsf{Bal}(\ell)} D(i) \leq \epsilon$ with probability 1. It follows that
$$\mathop{\mathrm{E}}_X[f(X)] \leq \frac{1}{2}+\delta+(\frac{1}{2}-\delta)\epsilon \leq \frac{1}{2}+\delta+\epsilon.$$

Bounding the probability of deviation. We will use the following result based on the method of bounded differences (see Corollary 6.10 of McDiarmid [10]).

Lemma 7.10 *Let X be a random variable taking values in $\{0,1\}^n$. Let f be any function from $\{0,1\}^n$ to $\Re$. Suppose for $j = 1, .., n$, and $x_i \in \{0,1\}$ $(i = 1, \dots, j-1)$*

$$\begin{aligned} |\,\mathrm{E}[f(X) \mid \mathcal{A}_1] - |\,\mathrm{E}[f(X) \mid \mathcal{A}_0]| &\leq c, \qquad (14)\\ \textit{where } \mathcal{A}_0 &\equiv \langle X_1, \dots, X_{j-1}\rangle = \langle x_1, \dots, x_{j-1}\rangle \wedge x_j = 0;\\ \textit{and } \mathcal{A}_1 &\equiv \langle X_1, \dots, X_{j-1}\rangle = \langle x_1, \dots, x_{j-1}\rangle \wedge x_j = 1. \end{aligned}$$

[The inequality (14) is required to hold whenever both conditional probabilities are defined.] Then, $\Pr[f(X) - \mathrm{E}[f(X)] \geq t] \leq \exp(-\frac{2t^2}{nc^2})$.

To obtain an appropriate value for c, fix $j \in [n]$ and $x_1, \dots, x_{j-1} \in \{0,1\}$. Consider the unique node v at distance $j-1$ from the root of τ, which corresponds to choices $x_1, \dots, x_{j-1}$ for $X_1, \dots, X_{j-1}$. Let w_0 and w_1 be the two children of v corresponding to the choices 0 and 1 for X_j. Let τ_0 be the subtree rooted at w_0 and τ_1 be the subtree rooted at w_1. Thus, we need to bound

$$|\mathop{\mathrm{E}}_{\ell \in L(\tau_1)}[f(X)] - \mathop{\mathrm{E}}_{\ell \in L(\tau_0)}[f(X)]| = |\mathop{\mathrm{E}}_{\ell \in L(\tau_1)}[\sum_{i=1}^{n} D(i)X_i] - \mathop{\mathrm{E}}_{\ell \in L(\tau_0)}[\sum_{i=1}^{n} D(i)X_i]|.$$

Note that $X_1, \dots, X_{j-1}$ are the same for all leaves in $L(\tau_0) \cup L(\tau_1)$; also $X_j = 1$ for all $\ell \in L(\tau_1)$, and $X_j = 0$ for all $\ell \in L(\tau_0)$. Thus, we are left with

$$|D(j) + \mathop{\mathrm{E}}_{\ell \in L(\tau_1)}[\sum_{i=j+1}^{n} D(i)X_i] - \mathop{\mathrm{E}}_{\ell \in L(\tau_0)}[\sum_{i=j+1}^{n} D(i)X_i]|. \qquad (15)$$

$$\begin{aligned} \mathop{\mathrm{E}}_{\ell \in L(\tau_1)}[\sum_{i=j+1}^{n} D(i)X_i] &= (\frac{1}{2} + \delta)P + (1 - (\frac{1}{2} + \delta)P) \mathop{\mathrm{E}}_{\ell \in L(\tau_1)}[\sum_{i=j+1,\, i \notin \mathsf{Bal}(\ell)}^{n} D(i)];\\ \mathop{\mathrm{E}}_{\ell \in L(\tau_0)}[\sum_{i=j+1}^{n} D(i)X_i] &= (\frac{1}{2} + \delta)P + (1 - (\frac{1}{2} + \delta)P) \mathop{\mathrm{E}}_{\ell \in L(\tau_0)}[\sum_{i=j+1,\, i \notin \mathsf{Bal}(\ell)}^{n} D(i)],\\ \text{where } P &= p_{j+1} + \cdots + p_n. \end{aligned}$$

Since $\epsilon - \beta \leq \sum_{i \notin \mathsf{Bal}(\ell)} D(i) \leq \epsilon$ for all $\ell \in L(\tau)$, we conclude that as ℓ varies in $L(\tau_0) \cup L(\tau_1)$, the value of $\sum_{i=j+1,\, i \notin \mathsf{Bal}(\ell)}^{n} D(i)$ is restricted to some interval of the form $[\gamma, \gamma + \beta]$. Thus, (15) is bounded by

$$|D(j) + (1 - (\frac{1}{2} + \delta)(p_{j+1} + \cdots + p_n))\beta| \leq \alpha + \beta.$$

Thus,

$$|\,\mathrm{E}[f(X) \mid \langle X_1, \dots, X_{j-1}\rangle = \langle x_1, \dots, x_{j-1}\rangle \wedge x_j = 1]| \leq \alpha + \beta.$$

Our claim now follows by taking c to be $\alpha + \beta$ in Lemma 7.10. ∎

Acknowledgements. We thank the referees for several suggestions which helped improve the paper. We thank Carsten Damm for his comments in the early stages of this work.

References

[1] N. Alon, J. Spencer, *The probabilistic method*, Wiley-interscience, 1992.

[2] C. Damm, S. Jukna, J. Sgall, Some bounds on multiparty communication complexity of pointer jumping, *Computational Complexity*, 7, 1998, pages 109–127.

[3] P. Duris, Z. Galil, G. Schnitger, Lower bounds on communication complexity, *Information and computation*, 73, 1987, pages 1–22.

[4] B. Halstenberg, R. Reischuk, On different modes of communication, *In Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, 1988, pages 162–172.

[5] J. Hromkovič, G. Schnitger, Nondeterministic communication with a limited number of advice bits, *In Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, 1996, pages 451–460.

[6] M. Karchmer, A. Wigderson, Monotone circuits for connectivity require superlogarithmic depth, *In Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, 1988, pages 539–550.

[7] H. Klauck, Lower bounds for computation with limited nondeterminism, *In Proceedings of the Thirteenth Annual IEEE Conference on Computational Complexity*, 1998, pages 141–152.

[8] M. Klawe, W. Paul, N. Pippenger, M. Yannakakis, On monotone formulae with restricted depth, *In Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, 1984, pages 480–487.

[9] E. Kushilevitz, N. Nisan, *Communication complexity*, Cambridge university press, 1996.

[10] C. McDiarmid, On the method of bounded differences, *London mathematical society lecture notes series*, 141, 1989, pages 148–188.

[11] N. Nisan, A. Wigderson, Rounds in communication complexity revisited, *SIAM journal of computing*, 22, 1993, pages 211–219.

[12] N. Nisan, D. Zuckerman, Randomness is linear in space, *Journal of Computing systems and sciences*, 52, 1996, pages 43–52.

[13] C. Papadimitriou, M. Sipser, Communication complexity, *Journal of Computing systems and sciences*, 28, 1984, pages 260–269.

[14] S. J. Ponzio, J. Radhakrishnan, S. Venkatesh, The communication complexity of pointer chasing, Applications of entropy and sampling, *In Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, 1999, pages 602–611.

[15] A. Yao, Lower bounds by probabilistic arguments, *In Proceedings of the Twenty-Fourth IEEE Conference on Foundations of Computer Science*, 1983, pages 420–428.