Electronic Colloquium on Computational Complexity, Report No. 14 (2007)

Lower Bounds for Multi-Player Pointer Jumping

Amit Chakrabarti
Dartmouth College
ac@cs.dartmouth.edu

December 7, 2006

Abstract

We consider the k-layer pointer jumping problem in the one-way multi-party number-on-the-forehead communication model. In this problem, the input is a layered directed graph with each vertex having outdegree 1, shared amongst k players: Player i knows all layers *except* the ith. The players must communicate, in the order $1, 2, \ldots, k$, to determine the vertex reached by following edges from a special start vertex. This problem has been considered by a number of researchers in the past because sufficiently strong lower bounds for it would have major consequences in circuit complexity.

We take an information complexity approach to this problem and obtain three lower bounds that improve upon earlier work. For myopic protocols (where players may see only one layer ahead but arbitrarily far behind), we greatly improve a lower bound due to Gronemeier (2006). Our new lower bound is $\Omega(n/k)$, where n is the number of vertices per layer. For conservative protocols (where players may see arbitrarily far ahead but not behind, instead seeing only the vertex reached by following the pointers up to their layer), we extend an $\Omega(n/k^2)$ lower bound due to Damm, Jukna and Sgall (1998) so that it applies for all k.

The above two bounds apply even to the Boolean version of pointer jumping. Our third lower bound is for the non-Boolean case and for $k \leq \log^* n$. We obtain an $\Omega(n \log^{(k-1)} n)$ bound for myopic protocols. Damm et al. had obtained a similar bound for deterministic conservative protocols. All our lower bounds apply directly to randomised protocols.

1 Introduction

Communication complexity has been a central technique in proving a number of lower bounds, even in models of computation that do not involve communication. In particular, it has some well known connections to circuit complexity: proving sufficiently strong lower bounds for certain specific communication problems would place them outside certain restricted, but well-studied, classes of circuits. For example, the celebrated super-logarithmic lower bound on the depth of a monotone circuit for undirected connectivity, due to Karchmer and Wigderson [KW90], was proven via a lower bound on a related communication problem.

Our focus here is on the *pointer jumping* (also called *pointer chasing*) problem and its multi-party communication complexity in the so-called *number-on-the-forehead* (NOF) model, introduced by Chandra, Furst and Lipton [CFL83]. Due to known connections between this model and circuits [Yao90, HG91, BT94], a strong enough communication lower bound for pointer jumping would place the problem outside the complexity class ACC^0. We say more about this connection in Section 1.2. In this work we introduce an approach to proving such communication lower bounds via *information complexity*, a concept formally introduced by Chakrabarti et al. [CSWY01] and refined by Bar-Yossef et al. [BJKS02]. Our approach results

ISSN 1433-8092

in lower bounds for pointer jumping in certain restricted one-way NOF communication models. Our lower bounds are at least as high as (in fact, much higher than) would be required to prove non-membership in ACC^0; proving similar bounds in a less restricted communication model would imply that pointer jumping is not in ACC^0.

1.1 The Problem and Our Results

The term "pointer jumping" has been used to refer to any of a family of related problems, all of which involve following pointers (i.e., directed edges) out of a starting vertex in a given input graph. The variant called *multi-layer pointer jumping* with k layers, denoted $\widehat{\text{MPJ}}_k$, is defined on a fixed underlying graph G_k^n whose vertex set consists of $k+1$ layers of vertices: layer 0 has a single vertex v_0 and layers 1 through k have n vertices each, and every vertex in layer i has a directed edge to every vertex in layer $i+1$. The input is a subgraph of G_k^n in which every vertex (except those in layer k) has outdegree 1. The desired output is the name of the unique vertex in layer k reachable from v_0, i.e., the final vertex reached by "jumping along pointers" starting at v_0. The output is therefore $\lceil \log n \rceil$ bits long.[1] We can also consider a Boolean version, denoted MPJ_k, by shrinking layer k so that it consists of 2 vertices. We give a more formal definition later.

A couple of other variants of pointer jumping that have been studied before are *tree pointer jumping* (TPJ_k), where the underlying graph G_k^n is replaced by a complete n-ary tree of height $k+1$, and *bipartite pointer jumping* (BPJ_k), where G_k^n is replaced by a bipartite graph with directed edges in both directions and one is required to follow k edges (pointers) from a designated start vertex.

In the number-on-the-forehead (NOF) model of communication, there are k players who share an input $(x_1, \ldots, x_k) \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_k$ as follows: Player i sees every x_j where $j \neq i$. We think of x_i as being written on Player i's forehead. The goal is to exchange messages according to a *protocol* so as to jointly compute a function $f : \mathscr{A}_1 \times \cdots \times \mathscr{A}_k \to \mathscr{B}$. For the purposes of proving lower bounds against ACC^0 circuits, it suffices to consider *simultaneous message* protocols, where all players simultaneously send their messages to a referee (who is not one of the k players) who sees no input and computes the desired output as function of the messages he receives. In this paper, as in some earlier work [NW93, PRS97, DJS98], we consider the more general *one-way blackboard communication* model, where players communicate one after another, in the fixed order $1, 2, \ldots, k$, by writing their messages on a blackboard visible to all. Player k's message is the desired output.

It is natural to consider k-player NOF protocols for MPJ_k where the input on Player i's forehead describes the ith layer of edges in the input graph (i.e., edges from vertices in layer $i-1$ to vertices in layer i). Note that it is important that the players speak in the order $1, 2, \ldots, k$ in order for the problem to be nontrivial: any other order of speaking leads to an easy protocol with only $O(\log n)$ communication.

Unfortunately, we are unable to prove our results in the unrestricted one-way model. Instead, we work with two different restrictions of the model. Our first lower bound applies to *myopic protocols*: those in which Player i only sees $x_1, \ldots, x_{i-1}$ and x_{i+1}. This model was recently introduced by Gronemeier [Gro06] who proved a lower bound of $\Omega(n^{(1-\varepsilon)/k} \log n)$ for $\widehat{\text{MPJ}}_k$ in this model, for ε-error protocols.[2] Note that this bound becomes trivial for $k = O(\log n)$ players. We prove the following, much stronger, lower bound.

Theorem 1. *A randomised myopic protocol for* MPJ_k *must communicate* $\Omega(n/k)$ *bits.*

[1] Throughout this paper we use "log" to denote logarithm to the base 2.

[2] Gronemeier defines myopic protocols using information theoretic terminology. In fact, the notion he defines should be described as "protocol that is myopic for a particular input distribution." In his work, he only applies his definition with the uniform distribution on inputs, in which case his information theoretic definition reduces to our structural one. Indeed, protocols myopic for *arbitrary* input distributions can communicate essentially nothing, for one could always consider distributions that perfectly correlate the inputs on the players' foreheads.

Our second lower bound applies to *conservative protocols*: those in which Player i only sees $x_{i+1}, \ldots, x_k$ and the function $g_{x,i} : \mathscr{A}_i \times \mathscr{A}_{i+1} \times \cdots \times \mathscr{A}_k \to \mathscr{B}$ given by $g_{x,i}(z_i, \ldots, z_k) = f(x_1, \ldots, x_{i-1}, z_i, \ldots, z_k)$. For pointer jumping, this amounts to saying that Player i sees all layers $i+1, \ldots, k$ of edges (i.e., the layers following the one on her forehead), but not layers $1, \ldots, i-1$; however, she does see the result of following $i-1$ pointers starting from v_0. This model was introduced by Damm, Jukna and Sgall [DJS98] who proved a lower bound of $\Omega(n/k^2)$ for $\widehat{\text{MPJ}}_k$ for deterministic protocols involving up to $k = o\left((n/\log n)^{1/3}\right)$ players (their argument also applies to MPJ_k and can be extended to randomised protocols using some careful estimation). Here, we obtain the same lower bound without an extra restriction on k, and via different techniques.

Theorem 2. *A randomised conservative protocol for* MPJ_k *must communicate* $\Omega(n/k^2)$ *bits.*

Although these models are quite restrictive, we note that the only known nontrivial upper bound for pointer jumping, due to Damm et al. [DJS98], is via a protocol that is *both* myopic and conservative (but see Section 1.2, below). Their improvement over a trivial upper bound is for $\widehat{\text{MPJ}}_k$ only: they give a (conservative and myopic) protocol for it with communication $O(n \log^{(k-1)} n)$ for $k \leq \log^* n$ and $O(n)$ for $k > \log^* n$.[3] The trivial upper bound would have been $O(n \log n)$. This shows that both restricted models do allow nontrivial protocols. They also give a matching $\Omega(n \log^{(k-1)} n)$ lower bound for deterministic conservative protocols; their proof does *not* generalise to randomised protocols. Here, we give a matching lower bound for randomised *myopic* protocols.

Theorem 3. *A randomised myopic protocol for* $\widehat{\text{MPJ}}_k$*, involving* $k \leq \log^* n$ *players, must communicate* $\Omega(n \log^{(k-1)} n)$ *bits.*

Our techniques in fact allow us to combine and extend Theorems 1 and 2 by relaxing the restrictions on the communication model somewhat. Rather than constrain every player in the same way, we can consider protocols where some players are myopic and others conservative. We define specific players to be myopic or conservative in the natural way; e.g., Player i is myopic if she only sees inputs $x_1, \ldots, x_{i-1}$ and x_{i+1}. Let us define a (k_m, k_c)-*split protocol* to be a one-way NOF protocol with $(k_m + k_c)$ players such that players 1 through k_m are myopic and the rest are conservative.

Theorem 4. *Let* $k = k_m + k_c$ *where* $0 \leq k_m \leq k$*. A randomised* (k_m, k_c)*-split protocol for* MPJ_k *must communicate* $\Omega\left(\min\{n/k_m, n/k_c^2\}\right)$ *bits.*

1.2 Related Work: Motivation and Prior Results

The complexity class ACC^0 is defined to be the class of all Boolean functions computable using circuits with constant depth and polynomial size that consist of (unbounded fan-in) AND, OR, NOT, and MOD_m gates, for arbitrary values of m. This is about the smallest well-studied class for which we do not know an explicit non-member. Finding an explicit function not in ACC^0 is a major open problem in complexity theory. The function MPJ_k is often considered a good candidate, partly because it is complete for $\mathsf{LOGSPACE}$, which contains ACC^0, and partly because it seems amenable to a communication complexity approach that we now describe.

A series of papers by Yao [Yao90], Håstad and Goldmann [HG91], and Beigel and Tarui [BT94] showed that ACC^0 is included in SYM^+, the class of depth-2 circuits with polylogarithmic fan-in AND gates at the

[3] We use $\log^{(k)} n$ to denote the kth iterated logarithm of n. More precisely, $\log^{(1)} n = \log n$, and $\log^{(k)} n = \log\left(\log^{(k-1)} n\right)$ for $k > 1$. We use $\log^* n$ to denote the smallest integer r such that $\log^{(r)} n \leq 1$.

input level and a single quasi-polynomial fan-in symmetric gate at the output level. This in turn means that for every function $f : \{0,1\}^n \to \{0,1\}$ in ACC^0 and every possible way of splitting its input bits into $k = \text{poly}(\log n)$ parts, the corresponding multi-player communication problem $f(x_1, \ldots, x_k)$ has a simultaneous message (hence, one-way) NOF protocol that communicates $\text{poly}(\log n)$ bits. Therefore, removing the restrictions (myopia/conservativeness) on the communication model in either of our Theorems 1 or 2 would imply $\text{MPJ}_k \notin \mathsf{ACC}^0$. This is our primary motivation.

We have already mentioned the work of Damm et al. [DJS98] and Gronemeier [Gro06] on lower bounds for MPJ_k. One other significant lower bound in the area is due to Wigderson (unpublished, but see Babai, Hayes and Kimmel [BHK01] for an exposition), building on the work of Nisan and Wigderson [NW93]: it shows that an *unrestricted* deterministic one-way NOF protocol for MPJ_3 requires $\Omega(\sqrt{n})$ bits of communication. Improving this bound is a key open question, as is proving *any* unrestricted $\Omega(n^\varepsilon)$ bound for MPJ_4. We hope that this work provides new insights and spurs progress on these problems.

An important potential obstacle in proving more such unrestricted lower bounds was identified by Pudlák, Rödl and Sgall [PRS97]. They showed, via an ingenious non-constructive probabilistic argument, that a special case of MPJ_3, where the middle layer is a *permutation*, has a one-way NOF protocol with communication $O((n \log\log n)/\log n)$. The protocol is neither myopic nor conservative. This result should be viewed as cautioning against a hasty conjecture of an $\Omega(n)$ lower bound for MPJ_3. However, such a lower bound is not yet ruled out, because the protocol does not work for a general instance of MPJ_3.

There is also a long line of work on the two-party complexity of the aforementioned variants BPJ_k and $\widehat{\text{BPJ}}_k$, starting with Papadimitriou and Sipser [PS84] and continuing with Nisan and Wigderson [NW93], Ponzio, Radhakrishnan and Venkatesh [PRV01], Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01] and Jain, Radhakrishnan and Sen [JRS02]. We refer the reader to the latter paper for more details and history. There is some work on the variant TPJ_k by Klauck et al. [KNTZ01]. Some of these papers also consider quantum communication settings.

1.3 Organisation of the Paper

The rest of the paper is organised as follows. In Section 2, we outline the basic plan that all our proofs follow. We then introduce our terminology and notation formally. In Section 3 we introduce some information theoretic tools used in the proofs. We then use these tools to perform certain "protocol manipulations" in Section 4, culminating in a couple of *round elimination lemmas* that form the heart of the argument. Section 5 uses the round elimination lemmas to prove Theorems 1, 2 and 3. Finally, in Section 6 we comment on some open problems and give a brief sketch of how our techniques can be extended to prove Theorem 4.

2 Preliminaries

2.1 Plan of the Proofs

Our proof formalises the following intuitive argument. Suppose there is a k-player one-way NOF protocol P for MPJ_k in which each player communicates at most αn bits, for some "small" quantity α. Let us run P on a random input and consider the information revealed by Player 1's message about the second layer of pointers (i.e., the input on Player 2's forehead). This layer consists of n pointers. Since Player 1 sends at most αn bits, there exists an $i \in \{1, 2, \ldots, n\}$ such that she reveals at most α bits of information about the ith pointer.

Now, consider instances of $\textsc{mpj}_k$ in which the pointer from v_0 always points to the ith vertex in layer 1; note that such instances are effectively instances of $\textsc{mpj}_{k-1}$. We thus have a k-player protocol for $\textsc{mpj}_{k-1}$, with the inputs written on the foreheads of Players 2 through k. In and of itself, such a protocol is silly: the first player can simply compute the final answer and reveal it. However, our protocol has the additional property that Player 1 reveals only $\alpha \ll 1$ bits about the input on Player 2's forehead. Using an appropriate tool from information theory, we can argue that it does not make much difference if we alter Player 1's behaviour so she sends *zero* information about that input. More precisely, the protocol's error probability increases by $O(\sqrt{\alpha})$. At this point, Player 2 can emulate Player 1, so we may eliminate Player 1 from the game altogether. We now have $(k-1)$-player protocol Q for $\textsc{mpj}_{k-1}$ with slightly larger error probability than P.

Iterating this construction $k-2$ times, we eventually arrive at a 2-player protocol for $\textsc{mpj}_2$, which is simply a restatement of INDEX problem. At this point, we can apply standard two-party one-way communication lower bounds for INDEX. Note that in order for the error to have only increased by a constant, we need $\alpha = O(1/k^2)$, limiting us to an $\Omega(n/k^2)$ lower bound. A more careful analysis gives a higher $\Omega(n/k)$ bound for myopic protocols.

When seeking a super-linear lower bound for $\widehat{\textsc{mpj}}_k$, the above outline runs into trouble because $\alpha > 1$, which means that $O(\sqrt{\alpha})$ additional error is intolerable. Therefore, we need a different information theoretic tool. The details appear below, but for readers familiar with the work of Chakrabarti and Regev [CR04], we mention that the tool we need has the flavour of combining a "message compression lemma" and a "message switching lemma" from that work. The compression lemma is in turn inspired by the work of Jain, Radhakrishnan and Sen [JRS03].

Some earlier lower bounds on pointer jumping in traditional two-player settings (i.e., for $\textsc{bpj}_k$, $\widehat{\textsc{bpj}}_k$ and $\textsc{tpj}_k$) were proven using similar information theoretic ideas [KNTZ01, JRS02] in a quantum communication setting. However, extra complications are introduced when dealing with $\textsc{mpj}_k$ and the NOF model, which makes new technical ideas necessary in our work.

2.2 Terminology and Notation

For the rest of the paper, "protocols" shall be assumed to be public coin randomised protocols in the one-way NOF model, unless explicitly qualified otherwise. The more common Alice-and-Bob protocols with messages exchanged between two players shall be called "traditional protocols."

We shall assume that each message in a protocol has a predetermined length independent of the actual input; this makes no asymptotic difference in communication cost. Let P be a k-player protocol in which Player i's message has length ℓ_i. We say that the *signature* of P is $\langle \ell_1, \ell_2, \ldots, \ell_k \rangle$ or, equivalently, that P is an $\langle \ell_1, \ldots, \ell_k \rangle$-protocol. We define $\mathrm{cost}(P) := \ell_1 + \cdots + \ell_k$. We denote the error probability of P (over its internal coin tosses) on its worst case input by $\mathrm{err}(P)$. For deterministic as well as randomised protocols, we define the distributional error of P with respect to input distribution $\mathcal{D}$ by $\mathrm{err}(P, \mathcal{D})$.

For random variables X, Y and Z, we use $\mathrm{H}(X)$ to denote the entropy of X (in bits), $\mathrm{I}(X : Y)$ to denote the mutual information between X and Y, and $\mathrm{H}(X \mid Z)$ and $\mathrm{I}(X : Y \mid Z)$ to denote conditional entropy and conditional mutual information, respectively. We use a number of basic results from information theory. For more on the subject we refer the reader to the textbook by Cover and Thomas [CT91].

In addition to the restrictions of myopia and conservativeness, defined above, we will need to consider the following unusual restriction.

Definition 1 (Quasi-private coin protocols). A protocol involving $k \geq 2$ players is said to be quasi-private coin if the random coin of Player 1 is private. Players 2 through k may continue to share a public coin.

Definition 2 (Information cost). Let P be a protocol for a problem $\phi : \mathscr{A}_1 \times \cdots \times \mathscr{A}_k \to \mathscr{B}$ and $\mathcal{D}$ a distribution on $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$. The information cost of P with respect to $\mathcal{D}$, denoted $\mathrm{icost}(P, \mathcal{D})$ is defined to be the following conditional mutual information:

$$\mathrm{icost}(P, \mathcal{D}) \ := \ \mathrm{I}(X_2 : M \mid X_3, \ldots, X_k)$$

where $(X_1, \ldots, X_k) \sim \mathcal{D}$ and M is the random message produced by Player 1 when she sees $(X_2, \ldots, X_k)$.

Notice that the information cost deals only with the *first* message of the protocol and only captures the information revealed by this message about the input unavailable to Player 2. We have the following simple lemma relating the information cost of a protocol to a part of its actual communication cost.

Lemma 5. *Let P be an $\langle \ell_1, \ell_2, \ldots, \ell_k \rangle$-protocol and $\mathcal{D}$ be any distribution on the input to P. Then* $\mathrm{icost}(P, \mathcal{D}) \leq \ell_1$.

Proof. Using the notation in Definition 2 we have

$$\mathrm{icost}(P, \mathcal{D}) \ = \ \mathrm{I}(M : X_2 \mid X_3, \ldots, X_k) \ \leq \ \mathrm{H}(M \mid X_3, \ldots, X_k) \ \leq \ \mathrm{H}(M) \ \leq \ |M| \ = \ \ell_1 \,. \qquad \square$$

Definition 3 (Pointer jumping). For a positive integer n, let $[n] := \{1, 2, \ldots, n\}$. For $k \geq 2$, we define $\widehat{\mathrm{MPJ}}_k : [n] \times \left([n]^{[n]}\right)^{k-1} \to [n]$ recursively, as follows. Here, $i \in [n]$ and $f, \ f_2, \ldots, \ f_k \in [n]^{[n]}$.

$$\begin{aligned} \widehat{\mathrm{MPJ}}_2(i, \ f) \quad &= \quad f(i) \,, \\ \widehat{\mathrm{MPJ}}_k(i, \ f_2, \ f_3, \ldots, \ f_k) \quad &= \quad \widehat{\mathrm{MPJ}}_{k-1}(\ f_2(i), \ f_3, \ldots, \ f_k) \,, \quad \forall\, k > 2 \,. \end{aligned}$$

We define $\mathrm{MPJ}_k : [n] \times \left([n]^{[n]}\right)^{k-2} \times \{0,1\}^n \to \{0,1\}$ similarly, except that we start with $\mathrm{MPJ}_2(i, x) = x_i$ for $i \in [n]$ and $x \in \{0,1\}^n$.

The crucial fact about pointer jumping that we exploit is that an instance of MPJ_{k-1} can be “embedded” in an instance of MPJ_k. This is made precise in the following lemma, whose trivial proof we omit.

Lemma 6. *For $f \in [n]^{[n]}$ and $i, a \in [n]$, define the function $f^{i:a} \in [n]^{[n]}$ as follows:*

$$f^{i:a}(j) \ = \ \begin{cases} a \,, & \text{if } j = i \,, \\ f(j) \,, & \text{otherwise.} \end{cases}$$

Then, for any $k \geq 3$, $i \in [n]$ and $g \in [n]^{[n]}$, we have $\mathrm{MPJ}_{k-1}(a, \ f_3, \ldots, \ f_k) = \mathrm{MPJ}_k(i, g^{i:a}, \ f_3, \ldots, \ f_k)$. A similar statement holds for $\widehat{\mathrm{MPJ}}_{k-1}$ and $\widehat{\mathrm{MPJ}}_k$. □

3 Information Theoretic Tools

We now present two key information theoretic tools that we shall use in our proofs. It may be helpful to keep in mind the following context while reading this section. We have two random variables — to be thought of as “input” and “response” — and a function that assigns a real-valued score to each input-response pair. We would like to alter the response in some way so as to simplify it without changing the expected score much. In Lemma 8 below, the input splits into two independent portions (A and B) and the response (C) carries a negligible amount of information about one of the portions (A); we show that the response can be made functionally independent of that portion. In Lemma 9 below, the response (B) carries a small amount of information about the input (A); we show that the response can be restricted to lie in a correspondingly small set.

The latter lemma is similar to (and stronger than) a lemma of Chakrabarti and Regev [CR04] that was used to compress the first message of a traditional protocol. We use it here for a very similar purpose. Lemma 8 is in the spirit of the Average Encoding Theorem of Klauck et al. [KNTZ01] and we use it here to eliminate "uninformative" messages. It explicates and generalises similar ideas in Sen [Sen03] and Chakrabarti and Regev [CR04].

We recall the following well known theorem from information theory (see, e.g, Lemma 12.6.1 of Cover and Thomas [CT91]).

Fact 7 (Pinsker's inequality). *Let $\mathcal{P}$ and $\mathcal{Q}$ be two probability distributions on the same domain. Then the Kullback-Leibler divergence $\mathrm{D_{KL}}(\mathcal{P}\|\mathcal{Q})$ and the L_1 distance $\|\mathcal{P}-\mathcal{Q}\|_1$ are related by*

$$\mathrm{D_{KL}}(\mathcal{P}\|\mathcal{Q}) \;\geq\; \frac{1}{2\ln 2}\|\mathcal{P}-\mathcal{Q}\|_1^2\,.$$

Lemma 8. *Let A, B and C be random variables with ranges $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$ respectively. Suppose A and B are independent. Then, for every function $f : \mathscr{A}\times\mathscr{B}\times\mathscr{C}\to[0,1]$, there exists a function $g:\mathscr{B}\to\mathscr{C}$ such that*

$$\mathrm{E}_{A,B}[f(A,B,g(B))] \;\leq\; \mathrm{E}_{A,B,C}[f(A,B,C)] + \sqrt{\frac{\ln 2}{2}\cdot \mathrm{I}(A:C\mid B)}\,.$$

Proof. Let Π be the joint distribution of (A,B,C) and let $\Pi_{\mathscr{A}}$, $\Pi_{\mathscr{B}\mathscr{C}}$, etc. be its marginals. Define the distribution Π' on $\mathscr{A}\times\mathscr{B}\times\mathscr{C}$ by $\Pi'(a,b,c) = \Pi_{\mathscr{A}}(a)\Pi_{\mathscr{B}\mathscr{C}}(b,c)$. By independence of A and B, we have

$$\mathrm{D_{KL}}(\Pi\|\Pi') \;=\; \mathrm{I}(A:BC) \;=\; \mathrm{I}(A:B)+\mathrm{I}(A:C\mid B) \;=\; \mathrm{I}(A:C\mid B)\,. \tag{1}$$

Observe that

$$\begin{aligned}
\sum_{b\in\mathscr{B}}\sum_{c\in\mathscr{C}}\Pi_{\mathscr{B}\mathscr{C}}(b,c)\sum_{a\in\mathscr{A}}\Pi_{\mathscr{A}}(a)\,f(a,b,c) \;&=\; \sum_{a\in\mathscr{A}}\sum_{b\in\mathscr{B}}\sum_{c\in\mathscr{C}}\Pi'(a,b,c)\,f(a,b,c) &(2)\\
&\leq\; \mathrm{E}_{A,B,C}[f(A,B,C)] + \frac{1}{2}\|\Pi-\Pi'\|_1 &(3)\\
&\leq\; \mathrm{E}_{A,B,C}[f(A,B,C)] + \frac{1}{2}\sqrt{(2\ln 2)\cdot\mathrm{D_{KL}}(\Pi\|\Pi')} &(4)\\
&=\; \mathrm{E}_{A,B,C}[f(A,B,C)] + \sqrt{\frac{\ln 2}{2}\cdot\mathrm{I}(A:C\mid B)}\,, &(5)
\end{aligned}$$

where (3) holds because f takes values in $[0,1]$, (4) follows from Pinsker's inequality and (5) follows from (1). Now, define $g:\mathscr{B}\to\mathscr{C}$ by

$$g(b) \;:=\; \operatorname*{argmin}_{c\in\mathscr{C}}\sum_{a\in\mathscr{A}}\Pi_{\mathscr{A}}(a)\,f(a,b,c)\,.$$

Then, the sum on the left side of (2) is at least

$$\begin{aligned}
\sum_{b\in\mathscr{B}}\sum_{a\in\mathscr{A}}\Pi_{\mathscr{A}}(a)\,f(a,b,g(b))\sum_{c\in\mathscr{C}}\Pi_{\mathscr{B}\mathscr{C}}(b,c) \;&=\; \sum_{a\in\mathscr{A}}\sum_{b\in\mathscr{B}}\Pi_{\mathscr{A}}(a)\Pi_{\mathscr{B}}(b)\,f(a,b,g(b))\\
&=\; \mathrm{E}_{A,B}[f(A,B,g(B))]
\end{aligned}$$

which completes the proof. □

Lemma 9. *Let A and B be random variables with ranges $\mathscr{A}$ and $\mathscr{B}$ respectively. Then, for every function $f : \mathscr{A} \times \mathscr{B} \to [0, 1]$ and every $\lambda \geq 4\,\mathrm{I}(A : B)$, there exists $\mathscr{B}_0 \subseteq \mathscr{B}$ and a function $g : \mathscr{A} \to \mathscr{B}_0$ such that $|\mathscr{B}_0| \leq 2^\lambda$ and $\mathrm{E}_A[f(A, g(A))] \leq \mathrm{E}_{A,B}[f(A, B)] + \frac{5}{2}\sqrt{\mathrm{I}(A : B)/\lambda} + (1 + \log e)/\lambda$.*

Proof. This lemma is an analogue of Lemma 3.5 of Chakrabarti and Regev [CR04], but with tighter parameters. The proof is fairly technical. We give a complete self-contained proof in Appendix A. □

4 Protocol Manipulations

4.1 Removing Player 1's Message

We now prove a result (Lemma 11) that lets us remove Player 1's message in a protocol with a "slight" additive increase in error probability. The increase is in fact slight only when the information cost is low, to begin with. We use the result in our round elimination lemmas, below. The result requires the protocol to be quasi-private coin, so we begin with a preliminary lemma that addresses this requirement.

Lemma 10 (Quasi-privatisation lemma). *Let P be a myopic NOF protocol in which Player 2 is deterministic. Then there exists a quasi-private coin myopic protocol Q, with the same signature and information cost as P, that behaves identically to P on all inputs.*

Proof. If P involves just two players, there is nothing to prove. If it involves $k \geq 3$ players, we construct Q as follows. Let x_2 be the input on Player 2's forehead, R be the public random string used by all players in P to construct their messages, and $\mu^P(x_2, R)$ be the function computed by Player 1 to generate her first message in P. In Q, Player 1 still sends $\mu^P(x_2, R)$ but generates the random value R privately. Player 2 behaves the same as in P. Let $\mathcal{D}[x, m]$ denote the conditional distribution of $(R \mid \mu^P(x, R) = m)$. Players 3 through k, upon seeing the Player 1's message m_1, use a new public coin to generate a value R' distributed according to $\mathcal{D}[x_2, m_1]$ and then behave just as in P, using R' to provide the randomness in their messages. It is easy to see that Q has all the desired properties. □

Lemma 11. *Suppose $k \geq 3$. Let P be a quasi-private coin $\langle \ell_1, \ldots, \ell_k \rangle$-protocol for a function $\phi : \mathscr{A}_1 \times \cdots \times \mathscr{A}_k \to \mathscr{B}$, and let $\mathcal{D}$ be a distribution on $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$.*

(1) If $\mathcal{D}$ is a product distribution, there exists a deterministic $\langle 0, \ell_1 + \ell_2, \ell_3, \ldots, \ell_k \rangle$-protocol Q for ϕ such that $\mathrm{err}(Q, \mathcal{D}) \leq \mathrm{err}(P, \mathcal{D}) + \sqrt{\mathrm{icost}(P, \mathcal{D})}$.

(2) If P is myopic, there exists a deterministic myopic $\langle 0, \ell_2, \ell_3, \ldots, \ell_k \rangle$-protocol Q for ϕ such that $\mathrm{err}(Q, \mathcal{D}) \leq \mathrm{err}(P, \mathcal{D}) + \sqrt{\mathrm{icost}(P, \mathcal{D})}$.

(3) If P is myopic, then for every $\lambda \geq 4 \cdot \mathrm{icost}(P, \mathcal{D})$ there exists a deterministic myopic protocol Q for ϕ with signature $\langle 0, 2^\lambda \ell_2, \ell_3, \ldots, \ell_k \rangle$ such that $\mathrm{err}(Q, \mathcal{D}) \leq \mathrm{err}(P, \mathcal{D}) + 3\sqrt{\mathrm{icost}(P, \mathcal{D})/\lambda} + 3/\lambda$.

Proof. We give the full details of the argument for Part (1). The other two parts use much the same argument, so we merely point out the key differences.

Part (1). Let R_1 denote the random string used by Player 1 to generate her first message and let R_2 denote the random string shared by Players 2 through k. Let ε^P be the error indicator function for P, defined as follows: $\varepsilon^P(x_1, \ldots, x_k, m, r_2) = 0$ or 1 according as P produces a correct or an incorrect answer on input

$(x_1, \ldots, x_k)$, when $R_2 = r_2$ and Player 1 sends the message m. Let $\mu^P(x_2, \ldots, x_k, r_1)$ be the function that Player 1 computes to produce her message. Then

$$\mathrm{err}(P, \mathcal{D}) \;=\; \mathrm{E}_{X_1,\ldots,X_k,R_1,R_2}\left[\varepsilon^P(X_1, \ldots, X_k, \mu^P(X_2, \ldots, X_k, R_1), R_2)\right], \tag{6}$$

where $(X_1, \ldots, X_k) \sim \mathcal{D}$ and (R_1, R_2) is distributed uniformly. Let $\mathscr{M}$ be the domain of Player 1's message. Define $f : \mathscr{A}_2 \times \cdots \times \mathscr{A}_k \times \mathscr{M} \to [0, 1]$ by $f(x_2, \ldots, x_k, m) = \mathrm{E}_{X_1,R_2}[\varepsilon^P(X_1, x_2, \ldots, x_k, m, R_2)]$. Set $A := X_2$, $B := (X_3, \ldots, X_k)$, and $C := \mu^P(X_2, \ldots, X_k, R_1)$. Note that A and B are independent because $\mathcal{D}$ is a product distribution. Now, invoking Lemma 8 (and discarding the constant $(\ln 2)/2$ for simplicity) shows that there exists a function $g : \mathscr{A}_3 \times \cdots \times \mathscr{A}_k \to \mathscr{M}$ such that

$$\begin{aligned}\mathrm{E}_{A,B}[f(A, B, g(B))] \;&\le\; E_{X_2,\ldots,X_k,C}[f(X_2, \ldots, X_k, C)] + \sqrt{\mathrm{I}(X_2 : C \mid X_3, \ldots, X_k)}\\ &=\; \mathrm{err}(P, \mathcal{D}) + \sqrt{\mathrm{icost}(P, \mathcal{D})}\,,\end{aligned}$$

where the final equality follows from (6), the definition of f and the definition of icost.

Consider a protocol P' that is identical to P except that Player 1 sends the message $g(x_3, \ldots, x_k)$. Since the function ε^P has been parametrized by Player 1's message, we can use it to express the error probability of P' as well:

$$\mathrm{err}(P', \mathcal{D}) \;=\; \mathrm{E}_{X_1,\ldots,X_k,R_2}\left[\varepsilon^P(X_1, \ldots, X_k, g(X_3, \ldots, X_k), R_2)\right] \;=\; \mathrm{E}_{A,B}[f(A, B, g(B))]\,.$$

But note that Player 1's message in P' is a (deterministic) function of the inputs on the foreheads of Players 3 through k alone. Therefore, Player 2 has all the information necessary to generate this message. Therefore, there is a protocol P'' that behaves the same as P' on all inputs, but where Player 1 sends 0 bits and Player 2 sends $\ell_1 + \ell_2$ bits: the concatenation of Player 1's and Player 2's messages in P'. Finally, since we only care about distributional error under $\mathcal{D}$, we can fix the random coins of P'' to get a deterministic protocol Q that has the desired properties.

Part (2). We proceed almost exactly as in Part (1). The key difference is that Player 1 produces her message by computing a function $\mu^P(x_2, r_1)$, so when we construct P' as above, we end up with Player 1's message in P' being a constant. Therefore, there is no need for this message in P' at all and we can get the desired protocol Q by simply eliminating it and then fixing the resulting protocol's random coins.

Note that we did *not* require $\mathcal{D}$ to be a product distribution. This is because the condition that A and B are independent was satisfied vacuously.

Part (3). We proceed as in Part (2). Since P is myopic, Player 1's message is given by a function $\mu^P(x_2, r_1)$ and we have

$$\mathrm{err}(P, \mathcal{D}) \;=\; \mathrm{E}_{X_1,\ldots,X_k,R_1,R_2}\left[\varepsilon^P(X_1, \ldots, X_k, \mu^P(X_2, R_1), R_2)\right] \;=\; \mathrm{E}_{X_2,R_1}\left[f(X_2, \mu^P(X_2, R_1)\right],$$

where $f(x, m) := \mathrm{E}_{X_1,X_3,\ldots,X_k,R_2}[\varepsilon^P(X_1, x_2, X_3, \ldots, X_k, m, R_2)]$. Let $\mathscr{M}$ be the domain of Player 1's message. Setting $A := X_2$ and $B := \mu^P(X_2, R_1)$ and invoking Lemma 9 (and weakening the constants slightly), we see that there exists $\mathscr{M}_0 \subseteq \mathscr{M}$ and a function $g : \mathscr{A}_2 \to \mathscr{M}_0$ such that $|\mathscr{M}_0| \le 2^\lambda$ and

$$\mathrm{E}_A[f(A, g(A)] \;\le\; \mathrm{err}(P, \mathcal{D}) + 3\sqrt{\frac{\mathrm{icost}(P, \mathcal{D})}{\lambda}} + \frac{3}{\lambda}\,.$$

Consider a protocol P' that is identical to P except that Player 1 sends the message $g(x_2)$. As in Part (1), we have $\mathrm{err}(P', \mathcal{D}) = \mathrm{E}_A[f(A, g(A)]$. Also, P' is myopic. In particular, every player except Player 2 can compute Player 1's message in P'. Therefore, P' behaves identically to a protocol P'' constructed as follows. In P'', Player 1 sends 0 bits. Player 2 sends her response to each of the $|\mathscr{M}_0|$ messages that Player 1 could have sent in P'. Note that this requires $|\mathscr{M}_0| \cdot \ell_2 \leq 2^\lambda \ell_2$ bits. Players 3 through k determine Player 1's would-be message in P' and pick out the appropriate response to it from Player 2's long message and continue the rest of the protocol exactly as in P'.

Clearly, the signature of P'' is $\langle 0, 2^\lambda \ell_2, \ell_3, \ldots, \ell_k \rangle$. Fixing the random coins of P'' gives us a deterministic protocol Q with all the desired properties. □

4.2 Round Elimination for Pointer Jumping

Here we prove our two central lemmas, showing how to eliminate the first message — and hence the first player — of certain NOF protocols for MPJ_k and $\widehat{\mathrm{MPJ}}_k$, and thereby obtain NOF protocols for MPJ_{k-1} and $\widehat{\mathrm{MPJ}}_{k-1}$, respectively.

Definition 4. We use $\mathcal{U}^k$ to denote the uniform distribution on inputs to MPJ_k.

Lemma 12 (Round elimination, Boolean case). *Suppose* MPJ_k *has a deterministic* $\langle \ell_1, \ell_2, \ldots, \ell_k \rangle$*-protocol* P *with* $\mathrm{err}(P, \mathcal{U}^k) \leq \varepsilon$*, for some* $k \geq 3$*.*

(1) If P *is conservative, then* MPJ_{k-1} *has a deterministic conservative* $\langle \ell_1 + \ell_2, \ell_3, \ldots, \ell_k \rangle$*-protocol* Q *with* $\mathrm{err}(Q, \mathcal{U}^{k-1}) \leq \varepsilon + \sqrt{\ell_1/n}$*.*

(2) If P *is myopic, then* MPJ_{k-1} *has a deterministic myopic* $\langle \ell_2, \ell_3, \ldots, \ell_k \rangle$*-protocol* Q *with* $\mathrm{err}(Q, \mathcal{U}^{k-1}) \leq \varepsilon + \sqrt{\ell_1/n}$*.*

Proof. For each $j \in [n]$, we construct a randomised protocol P_j for MPJ_{k-1}, using k players: the input $(a,\ f_3, \ldots,\ f_k)$ to MPJ_{k-1} is written on the foreheads of Players 2 through k and Player 1's forehead is left blank. The players use a public coin to generate a uniform random layer of pointers $G \in [n]^{[n]}$. They then behave as they would have in protocol P on input $(j, G^{j:a},\ f_3, \ldots,\ f_k)$. In other words, if Player 1 would have sent the message $\mu^P(f_2, \ldots,\ f_k)$ in P, then she sends $\mu^P(G^{j:a},\ f_3, \ldots,\ f_k)$ in P_j. From Lemma 6, it follows that that P_j is correct whenever P is, on the constructed input $(j, G^{j:a},\ f_3, \ldots,\ f_k)$. Thus,

$$\frac{1}{n} \sum_{j=1}^{n} \mathrm{err}(P_j, \mathcal{U}^{k-1}) \ = \ \mathrm{err}(P, \mathcal{U}^k) \ \leq \ \varepsilon\,. \tag{7}$$

The information cost of P can be decomposed into the sum of the information costs of the P_js as follows.

$$\begin{aligned}
\mathrm{icost}(P, \mathcal{U}^k) &= \mathrm{I}(F_2 : \mu^P(F_2, \ldots, F_k) \mid F_3, \ldots, F_k) \\
&\geq \sum_{j=1}^{n} \mathrm{I}(F_2(j) : \mu^P(F_2, \ldots, F_k) \mid F_3, \ldots, F_k) \qquad (8) \\
&= \sum_{j=1}^{n} \mathrm{I}(A : \mu^P(G^{j:A}, F_3, \ldots, F_k) \mid F_3, \ldots, F_k) \\
&= \sum_{j=1}^{n} \mathrm{icost}(P_j, \mathcal{U}^{k-1})\,, \qquad (9)
\end{aligned}$$

where (8) holds because the n random variables $F_2(1), \ldots, F_2(n)$ are independent given $F_3, \ldots, F_k$. Combining (7) and (9), and using the concavity of the square root function, we get

$$\frac{1}{n}\sum_{j=1}^{n}\left(\mathrm{err}(P_j, \mathcal{U}^{k-1}) + \sqrt{\mathrm{icost}(P_j, \mathcal{U}^{k-1})}\right) \;\le\; \varepsilon + \sqrt{\frac{\mathrm{icost}(P, \mathcal{U}^k)}{n}} \;\le\; \varepsilon + \sqrt{\frac{\ell_1}{n}}\,,$$

where the final inequality follows from Lemma 5. Therefore, there exists a j such that $\mathrm{err}(P_j, \mathcal{U}^{k-1}) + \sqrt{\mathrm{icost}(P_j, \mathcal{U}^{k-1})} \le \varepsilon + \sqrt{\ell_1/n}$. We now prove the two parts of the lemma separately.

Part (1). Consider the protocol P_j. If P is conservative, then for any $i \ge 3$, the message of Player i in P_j can only depend on $f_{i+1}, \ldots, f_k$ and on the value $f_{i-1} \circ \cdots \circ f_3 \circ F_2(j)$ where $F_2 = G^{j:a}$. Although F_2 is randomly chosen, $F_2(j) = G^{j:a}(j) = a$, which means that Player i is in fact deterministic. Player 2 is trivially deterministic, irrespective of whether or not P is conservative. Thus, Player 1 is the only player to use randomness in P_j. In particular, P_j is a quasi-private coin protocol. By Part (1) of Lemma 11, there exists a deterministic $\langle 0, \ell_1 + \ell_2, \ell_3, \ldots, \ell_k\rangle$-protocol P' for MPJ_{k-1} such that $\mathrm{err}(P', \mathcal{U}^{k-1}) \le \varepsilon + \sqrt{\ell_1/n}$. In this protocol, Player 1 neither has an input on her forehead nor does she communicate any bits, so we effectively have a $(k-1)$-player $\langle \ell_1 + \ell_2, \ell_3, \ldots, \ell_k\rangle$-protocol Q with the desired properties.

Part (2). If P is myopic, then so is P_j. Moreover, Player 2 is deterministic in P_j. Invoking the quasi-privatisation lemma (Lemma 10), we can replace P_j with an equivalent quasi-private coin protocol P_j'. Applying Part (2) of Lemma 11 to P_j' and removing Player 1 as before gives us the desired deterministic $\langle \ell_2, \ldots, \ell_k\rangle$-protocol Q. □

Notice that the above lemma does not provide an interesting result when $\ell_1 \ge n$. But we must deal with $\ell_1 \ge n$ we are working with the non-Boolean problem, $\widehat{\mathrm{MPJ}}_k$, and wish to prove a communication lower bound higher than n. To this end, we introduce another round elimination lemma, below. The fact that $\widehat{\mathrm{MPJ}}_k$ is a non-Boolean problem does not play a significant role in its proof. However, for our application later, we need to work with randomised protocols in this lemma, rather than with deterministic protocols and distributional error.

Lemma 13 (Round elimination, non-Boolean case). *Suppose $\widehat{\mathrm{MPJ}}_k$ has a myopic $\langle \ell_1, \ell_2, \ldots, \ell_k\rangle$-protocol P, for some $k \ge 3$. Then, for $\lambda \ge 4\ell_1/n$, $\widehat{\mathrm{MPJ}}_{k-1}$ has a myopic protocol Q with signature $\langle 2^\lambda \ell_2, \ell_3, \ldots, \ell_k\rangle$ and with* $\mathrm{err}(Q) \le \mathrm{err}(P) + 3\sqrt{\ell_1/(n\lambda)} + 3/\lambda$.

Proof. We use much the same argument as in Part (2) of Lemma 12 but without fixing a specific input distribution like $\mathcal{U}^k$. Let $\mathcal{D}^{k-1}$ be an arbitrary input distribution for $\widehat{\mathrm{MPJ}}_{k-1}$. By Yao's minimax principle [Yao77], it suffices to demonstrate a deterministic protocol Q' with signature $\langle 2^\lambda \ell_2, \ell_3, \ldots, \ell_k\rangle$ and with $\mathrm{err}(Q', \mathcal{D}^{k-1}) \le \mathrm{err}(P) + 3\sqrt{\ell_1/(n\lambda)} + 3/\lambda$. Let $\mathcal{D}^k$ denote the distribution of the random input $(J, G^{J:A}, F_3, \ldots, F_k)$, where J is drawn uniformly from $[n]$, each of $G(1), \ldots, G(n)$ is drawn independently from the first marginal of $\mathcal{D}^{k-1}$ and $(A, F_3, \ldots, F_k) \sim \mathcal{D}^{k-1}$. By Yao's minimax principle again (the easy half, this time) there is a deterministic protocol P' for $\widehat{\mathrm{MPJ}}_k$ with the same signature as P and with $\mathrm{err}(P', \mathcal{D}^k) \le \mathrm{err}(P)$.

For each $j \in [n]$, we now design a protocol P_j for $\widehat{\mathrm{MPJ}}_{k-1}$ just as before, the only difference being that the random layer of pointers G is drawn from the first marginal of $\mathcal{D}^{k-1}$. Arguing as in the derivation of (7) and (9), we now have

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{err}(P_j, \mathcal{D}^{k-1}) \;\le\; \mathrm{err}(P)\,, \qquad \text{and} \qquad \sum_{j=1}^{n}\mathrm{icost}(P_j, \mathcal{D}^{k-1}) \;\le\; \mathrm{icost}(P, \mathcal{D}^k)\,.$$

We now combine these two inequalities appropriately to conclude that there exists a j such that

$$\mathrm{err}(P_j,\mathcal{D}^{k-1}) + 3\sqrt{\frac{\mathrm{icost}(P_j,\mathcal{D}^{k-1})}{\lambda}} + \frac{3}{\lambda} \;\le\; \mathrm{err}(P) + 3\sqrt{\frac{\ell_1}{n\lambda}} + \frac{3}{\lambda}\,.$$

Applying the quasi-privatisation lemma (Lemma 10) followed by Part (3) of Lemma 11 to P_j, and removing Player 1 as before, we obtain the desired protocol Q'. □

5 The Lower Bounds

Let Σ be a finite alphabet. We shall let Σ-INDEX denote the following traditional (i.e., not NOF) communication problem. There are two players: Alice, who holds a string $x = x_1x_2\ldots x_n \in \Sigma^n$ and Bob, who holds an index $i \in [n]$. Alice must send Bob a (possibly randomised) message, after which Bob must determine x_i. More precisely, the error of the protocol is defined to be the probability that Bob's output differs from x_i. The following lower bound is an easily proven generalisation of the well known lower bound for $\{0,1\}$-INDEX [Abl96]. The function H is the binary entropy function: $H(\alpha) = -\alpha\log\alpha-(1-\alpha)\log(1-\alpha)$.

Fact 14. *Let $\mathcal{U}$ denote the uniform distribution on inputs to Σ-INDEX. Any traditional protocol for Σ-INDEX with error at most ε on $\mathcal{U}$ must communicate at least $(1 - H(\varepsilon))\,n\log|\Sigma|$ bits.*

Theorem 15 (Precise restatement of Theorem 2). *Let P be a conservative protocol for MPJ$_k$ such that* $\mathrm{err}(P) \le \frac{1}{6}$*. Then* $\mathrm{cost}(P) = \Omega(n/k^2)$.

Proof. We first note that a 2-player NOF protocol for MPJ$_2$ is simply a traditional protocol for $\{0,1\}$-INDEX. Now, suppose MPJ$_k$ has an $\frac{1}{6}$-error randomised conservative $\langle\ell_1,\ldots,\ell_k\rangle$-protocol P for some $k \ge 3$. By the easy half of Yao's minimax principle, MPJ$_k$ has a deterministic conservative $\langle\ell_1,\ldots,\ell_k\rangle$-protocol P' with $\mathrm{err}(P',\mathcal{U}^k) \le \frac{1}{6}$. Applying Part (1) of Lemma 12 to P' repeatedly (i.e., $k-2$ times), we see that MPJ$_2$ has a deterministic protocol Q with $\mathrm{cost}(Q) \le \ell_1 + \cdots + \ell_k$ and

$$\mathrm{err}(Q,\mathcal{U}^2) \;\le\; \frac{1}{6} + \sqrt{\frac{\ell_1}{n}} + \sqrt{\frac{\ell_1+\ell_2}{n}} + \cdots + \sqrt{\frac{\ell_1+\cdots+\ell_{k-2}}{n}} \;\le\; \frac{1}{6} + k\sqrt{\frac{\ell_1+\cdots+\ell_k}{n}}\,.$$

Suppose $\mathrm{cost}(P) \le n/(36k^2)$. Then $\ell_1+\cdots+\ell_k \le n/(36k^2)$, so $\mathrm{err}(Q,\mathcal{U}^2) \le \frac{1}{6}+\frac{1}{6} = \frac{1}{3}$. By Fact 14, we have $\mathrm{cost}(Q) \ge \left(1 - H\left(\frac{1}{3}\right)\right) n \ge n/13$, a contradiction. □

Theorem 16 (Precise restatement of Theorem 1). *Let P be a myopic protocol for MPJ$_k$ with* $\mathrm{err}(P) \le \frac{1}{3}$*. Then* $\mathrm{cost}(P) = \Omega(n/k)$.

Proof. Proceeding as above, suppose MPJ$_k$ has an $\frac{1}{6}$-error randomised myopic $\langle\ell_1,\ldots,\ell_k\rangle$-protocol P for some $k \ge 3$. Applying Yao's minimax principle, followed by $k-2$ applications of Part (2) of Lemma 12, we get a deterministic protocol Q for MPJ$_2$ with $\mathrm{cost}(Q) \le \ell_{k-1}+\ell_k$ and

$$\mathrm{err}(Q,\mathcal{U}^2) \;\le\; \frac{1}{6} + \sqrt{\frac{\ell_1}{n}} + \cdots + \sqrt{\frac{\ell_{k-2}}{n}} \;\le\; \frac{1}{6} + \sqrt{\frac{k(\ell_1+\cdots+\ell_k)}{n}}\,,$$

where the final inequality is obtained by applying Cauchy-Schwarz. As before, we can obtain a contradiction if we assume that $\mathrm{cost}(P) \le n/(36k)$. □

Theorem 17 (Precise restatement of Theorem 3). *Every $\frac{1}{6}$-error myopic protocol for $\widehat{\text{MPJ}}_k$ with $k \le \log^* n$ must communicate $\Omega(n\log^{(k-1)} n)$ bits.*

Proof. Let $\mathcal{A}_k$ denote the statement "$\widehat{\mathrm{MPJ}}_k$ has a myopic protocol with error at most $\frac{1}{6}$ in which each player communicates at most $(n \log^{(k-1)} n)/400$ bits". Fact 14, applied to $[n]$-INDEX, implies that $\mathcal{A}_2$ is false. To complete the proof, we show that $\mathcal{A}_k \Rightarrow \mathcal{A}_{k-1}$ for each $k \geq 3$.

Assume $\mathcal{A}_k$, for some $k \geq 3$, and let P be the protocol whose existence is guaranteed by $\mathcal{A}_k$. By padding the messages of the players if necessary, we can assume that the signature of P is $\langle \ell, \ell, \ldots, \ell \rangle$ with $\ell = (n \log^{(k-1)} n)/400$. Set $\lambda = 399\ell/n$. By Lemma 13, there exists a $\langle 2^{\lambda}\ell, \ell, \ldots, \ell \rangle$-protocol Q for $\widehat{\mathrm{MPJ}}_{k-1}$ with

$$\mathrm{err}(Q) \;\leq\; \frac{1}{6} + 3\sqrt{\frac{\ell}{n(399\ell/n)}} + \frac{3}{399\ell/n} \;\leq\; \frac{1}{3}\,.$$

Consider a random variable $X_m \sim \mathcal{B}(m, \frac{1}{3})$, where $\mathcal{B}(m, p)$ denotes the binomial distribution with parameters m and p. Let c be the smallest integer satisfying $\Pr[X_c \geq c/2] \leq \frac{1}{6}$. Then, if we repeat a $\frac{1}{3}$-error protocol for some communication problem c times in parallel and report the majority output, we obtain a $\frac{1}{6}$-error protocol for the same problem. This continues to be true even if the problem is non-Boolean: there may not exist a majority output, but we can simply output something arbitrary in such cases. The upshot is that Q can be repeated c times in parallel to obtain a $\frac{1}{6}$-error $\langle 2^{\lambda}c\ell, c\ell, \ldots, c\ell \rangle$-protocol Q'. Now,

$$2^{\lambda}c\ell \;=\; \frac{2^{(399 \log^{(k-1)} n)/400} \cdot cn \log^{(k-1)} n}{400} \;=\; \frac{cn \left(\log^{(k-2)} n\right)^{399/400} \log^{(k-1)} n}{400} \;\leq\; \frac{n \log^{(k-2)} n}{400}\,,$$

for sufficiently large n. Therefore, the existence of Q' implies $\mathcal{A}_{k-1}$. □

6 Concluding Remarks and an Extension

We have obtained improved lower bounds on the one-way NOF communication complexity of pointer jumping in certain previously studied restricted models. Our approach is based on the information complexity paradigm and leads to proofs that have the nice feature of being formalisations of intuitive arguments. We believe that these results show the promise of this paradigm in attacking questions about NOF communication complexity.

At the same time, our proofs help bring out the limitations of the present way of applying information complexity. A key step in the paradigm is to solve a "simple" problem (in this case, MPJ_{k-1}) by simulating the actions of a protocol for a "compound" or "direct sum" problem (in this case, MPJ_k). In a NOF model, in order to create suitably distributed inputs for this larger problem, the players require public coins. This presents a challenge because round elimination seems to require the message under consideration to be generated using private coins. A meaningful measure of information complexity in a public coin setting requires conditioning on the public random string (for more on this, see Appendix B of Bar-Yossef et al. [BJKS02]) and this seems to stymie our argument. Here, we are able to work around this issue when handling either myopic or conservative protocols. There might, however, be a more sophisticated way of applying information complexity that can deal with less restricted models.

We can, in fact, relax our restrictions somewhat and consider *split protocols*, as in Theorem 4. Here is a brief sketch of its proof; the details are straightforward. In a split protocol, if Player 1 is conservative, so is every other player. Therefore, we may apply Theorem 2. If Player 1 is myopic, our round elimination argument still goes through, after a suitable modification to the quasi-privatisation lemma. The modified lemma works with protocols in which those players that do not see Player 2's input are all deterministic. Now, carrying out calculations very similar to those in the proofs of Theorems 1 and 2 completes the proof.

The most obvious open problem is to remove the restrictions from our lower bounds, thereby proving $\textsc{mpj}_k \notin \mathsf{ACC}^0$. Less ambitious goals include improving the known $\Omega(\sqrt{n})$ lower bound for $\textsc{mpj}_3$ and proving nontrivial lower bounds for $\textsc{mpj}_4$, both in the unrestricted one-way NOF model. It is tempting to conjecture an $\Omega(n)$ lower bound for $\textsc{mpj}_3$, but the protocol of Pudlák et al. [PRS97] sounds a note of caution.

References

[Abl96] Farid Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 175(2):139–159, 1996.

[BHK01] László Babai, Thomas P. Hayes, and Peter G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

[BJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 209–218, 2002.

[BT94] Richard Beigel and Jun Tarui. On ACC. *Comput. Complexity*, 4:350–366, 1994.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proc. 15th Annual ACM Symposium on the Theory of Computing*, pages 94–99, 1983.

[CR04] Amit Chakrabarti and Oded Regev. An optimal randomised cell probe lower bound for approximate nearest neighbour searching. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 473–482, 2004.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, 1991.

[DJS98] Carsten Damm, Stasys Jukna, and Jiří Sgall. Some bounds on multiparty communication complexity of pointer jumping. *Comput. Complexity*, 7(2):109–127, 1998. Preliminary version in *Proc. 13th International Symposium on Theoretical Aspects of Computer Science*, pages 643–654, 1996.

[Gro06] Andre Gronemeier. NOF-multiparty information complexity bounds for pointer jumping. In *Proc. 31st International Symposium on Mathematical Foundations of Computer Science*, 2006.

[HG91] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1:113–129, 1991.

[JRS02] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 429–438, 2002.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proc. 30th International Colloquium on Automata, Languages and Programming*, pages 300–315, 2003.

[KNTZ01] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proc. 33rd Annual ACM Symposium on the Theory of Computing*, pages 124–133, 2001.

[KW90] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Disc. Math.*, 3(2):255–265, 1990. Preliminary version in *Proc. 20th Annu. ACM Symp. Theory Comput.*, pages 539–550, 1988.

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SICOMP*, 22(1):211–219, 1993. Preliminary version in *Proc. 23rd Annu. ACM Symp. Theory Comput.*, pages 419–429, 1991.

[PRS97] Pavel Pudlák, Vojtěch Rödl, and Jiří Sgall. Boolean circuits, tensor ranks and communication complexity. *SIAM J. Comput.*, 26(3):605–633, 1997.

[PRV01] Stephen Ponzio, Jaikumar Radhakrishnan, and Srinivasan Venkatesh. The communication complexity of pointer chasing. *J. Comput. Syst. Sci.*, 62(2):323–355, 2001. Preliminary version in *Proc. 31st Annual ACM Symposium on the Theory of Computing*, pages 602–611, 1999.

[PS84] Christos Papadimitriou and Michael Sipser. Communication complexity. *J. Comput. Syst. Sci.*, 28(2):260–269, 1984. Preliminary version in *Proc. 14th Annual ACM Symposium on the Theory of Computing*, pages 196–200, 1982.

[Sen03] Pranab Sen. Lower bounds for predecessor searching in the cell probe model. In *Proc. 18th Annual IEEE Conference on Computational Complexity*, pages 73–83, 2003.

[Yao77] Andrew C. Yao. Probabilistic computations: Towards a unified measure of complexity. In *Proc. 18th Annual IEEE Symposium on Foundations of Computer Science*, pages 222–227, 1977.

[Yao90] Andrew C. Yao. On ACC and threshold circuits. In *Proc. 31st Annual IEEE Symposium on Foundations of Computer Science*, pages 619–627, 1990.

A Proof of Lemma 9

Theorem 18 (Restatement of Lemma 9). *Let A and B be random variables with ranges $\mathscr{A}$ and $\mathscr{B}$ respectively. Then, for every function $f : \mathscr{A} \times \mathscr{B} \to [0,1]$ and every $\lambda \geq 4\,\mathrm{I}(A : B)$, there exists $\mathscr{B}_0 \subseteq \mathscr{B}$ and a function $g : \mathscr{A} \to \mathscr{B}_0$ such that $|\mathscr{B}_0| \leq 2^{\lambda}$ and $\mathrm{E}_A[\,f(A, g(A))] \leq \mathrm{E}_{A,B}[\,f(A,B)] + \frac{5}{2}\sqrt{\mathrm{I}(A:B)/\lambda} + (1 + \log e)/\lambda$.*

Proof. Let Π denote the (marginal) distribution of B and Π_a the distribution of B conditioned on $A = a$. For each $a \in \mathscr{A}$, we introduce a fraction $\rho_a \in (0,1)$, whose precise value we set later. Define the sets S_a and T_a as follows:

$$S_a \;:=\; \{b \in \mathscr{B} :\; \rho_a \Pi_a(b) \leq \Pi(b)\}\,; \qquad T_a \;:=\; \{b \in \mathscr{B} :\; \rho_a \Pi_a(b) > \Pi(b)\}\,.$$

Define $\delta_a := \Pi_a(T_a)$. It will help to think of δ_a as being very small. Consider the function $h : [0,1] \times \mathscr{A} \to \mathscr{B}$ defined by the following algorithm.

```
Algorithm $h(r, a)$:
Inputs: $r \in [0,1]$, $a \in \mathscr{A}$.
Note: Designed to be invoked with an $r$ chosen at random, uniformly.
    Repeat forever:
        Using $r$ as a source of random bits, generate $b \in \mathscr{B}$ according to $\Pi$.
        Using $r$ again, return $b$ with probability $\min\{\rho_a \Pi_a(b)/\Pi(b),\ 1\}$.
```

Let Π'_a denote the distribution of $h(R, a)$, where R denotes a uniform random real in [0, 1] independent of A and B. Define σ_a to be the probability that the algorithm stops (i.e., returns some value) in a particular iteration. Then

$$\sigma_a \;=\; \sum_{b \in \mathscr{B}} \Pi(b) \cdot \min\{\rho_a \Pi_a(b)/\Pi(b),\, 1\} \;=\; \rho_a \Pi_a(S_a) + \Pi(T_a) \;=\; \rho_a(1-\delta_a) + \Pi(T_a)\,; \tag{10}$$

$$\text{and} \quad \Pi'_a(b) \;=\; \sum_{k=0}^{\infty} (1-\sigma_a)^k \cdot \Pi(b) \cdot \min\{\rho_a \Pi_a(b)/\Pi(b),\, 1\} \;=\; \frac{\min\{\rho_a \Pi_a(b), \Pi(b)\}}{\sigma_a}\,.$$

Therefore,

$$\begin{aligned}
\|\Pi_a - \Pi'_a\|_1 &= \sum_{b \in S_a} \left| \frac{\rho_a \Pi_a(b)}{\sigma_a} - \Pi_a(b) \right| + \sum_{b \in T_a} \left| \frac{\Pi(b)}{\sigma_a} - \Pi_a(b) \right| \\
&\le \left(\frac{\rho_a}{\sigma_a} - 1 \right) + \left(\frac{\Pi(T_a)}{\sigma_a} + \Pi_a(T_a) \right) \\
&= \frac{\rho_a - \sigma_a + \Pi(T_a)}{\sigma_a} + \delta_a \\
&= \frac{\rho_a \delta_a}{\rho_a(1-\delta_a) + \Pi(T_a)} + \delta_a & (11) \\
&\le \frac{\delta_a}{1-\delta_a} + \delta_a\,, & (12)
\end{aligned}$$

where (11) follows from (10).

Let $n(r, a)$ denote the number of iterations of the infinite loop performed by the above algorithm before it returns a value. Notice that $n(R, a)$ is a geometric random variable with expectation $1/\sigma_a$. Let $h'(r, a)$ be a function that uses a slightly modified version of the algorithm, where the infinite loop is replaced by a loop that makes at most 2^λ iterations. If no value is returned within those many iterations, the modified algorithm returns some arbitrary fixed element of $\mathscr{B}$. Let Π''_a denote the distribution of $h'(R, a)$. Then we have

$$\begin{aligned}
\frac{1}{2}\|\Pi''_a - \Pi'_a\|_1 &\le \Pr[h'(R,a) \ne h(R,a)] \\
&\le \Pr[n(R,a) > 2^\lambda] \\
&\le \mathrm{E}_R[\log n(R,a)]/\lambda & (13) \\
&\le \log \mathrm{E}_R[n(R,a)]/\lambda \\
&= (-\log \sigma_a)/\lambda \\
&\le \frac{-\log \rho_a - \log(1-\delta_a)}{\lambda}\,. & (14)
\end{aligned}$$

where (13) follows from Markov's inequality and (14) follows from (10). Combining (12) and (14) using the triangle inequality, we get

$$\|\Pi_a - \Pi''_a\|_1 \;\le\; \frac{2(-\log \rho_a - \log(1-\delta_a))}{\lambda} + \frac{\delta_a}{1-\delta_a} + \delta_a \tag{15}$$

Consider the two-point distributions $P = (\Pi_a(S_a), \Pi_a(T_a))$ and $Q = (\Pi(S_a), \Pi(T_a))$. By monotonicity of the Kullback-Leibler divergence, we have

$$\begin{aligned}\mathrm{D}_{\mathrm{KL}}(\Pi_a \| \Pi) &\geq \mathrm{D}_{\mathrm{KL}}(P \| Q) \\ &= \Pi_a(S_a) \log \frac{\Pi_a(S_a)}{\Pi(S_a)} + \Pi_a(T_a) \log \frac{\Pi_a(T_a)}{\Pi(T_a)} \\ &\geq (1-\delta_a)\log(1-\delta_a) + \delta_a \log \frac{1}{\rho_a} \\ &\geq -\delta_a \log e - \delta_a \log \rho_a \,,\end{aligned}$$

where the penultimate inequality follows from the definitions of S_a, T_a, and δ_a. For $\rho_a < 1/e$ this implies

$$\delta_a \leq \frac{\mathrm{D}_{\mathrm{KL}}(\Pi_a \| \Pi)}{-\log \rho_a - \log e} \,. \tag{16}$$

We would like to have Π''_a close to Π. Considering inequality (15), we notice that the first term on the right hand side is a decreasing function of ρ_a, whereas the second and third terms are increasing functions of δ_a, which is in turn upper bounded by an increasing function of ρ_a, according to (16). Therefore, to minimise $\|\Pi_a - \Pi''_a\|_1$, we should choose ρ_a neither too large nor too small. The asymptotically optimal choice turns out to be given by

$$-\log \rho_a = \sqrt{\frac{\lambda}{\mathrm{I}(A:B)}} \cdot \mathrm{D}_{\mathrm{KL}}(\Pi_a \| \Pi) + \log e \,.$$

Plugging this into (16), we get $\delta_a \leq \sqrt{\mathrm{I}(A:B)/\lambda}$. The condition on λ implies $\delta_a \leq 1/2$, which in turn gives $\delta_a/(1-\delta_a) + \delta_a \leq 3\delta_a \leq 3\sqrt{\mathrm{I}(A:B)/\lambda}$. We also have $-\log(1-\delta_a) \leq -\log(1-\frac{1}{2}) = 1$. Using these bounds in (15), we get

$$\|\Pi_a - \Pi''_a\|_1 \leq \frac{2 \cdot \mathrm{D}_{\mathrm{KL}}(\Pi_a \| \Pi)}{\sqrt{\lambda \cdot \mathrm{I}(A:B)}} + \frac{2(1+\log e)}{\lambda} + 3\sqrt{\frac{\mathrm{I}(A:B)}{\lambda}} \,.$$

Let $p_a := \Pr[A=a]$. Then $\sum_{a \in \mathscr{A}} p_a \mathrm{D}_{\mathrm{KL}}(\Pi_a \| \Pi) = \mathrm{I}(A:B)$. Therefore

$$\sum_{a \in \mathscr{A}} p_a \|\Pi_a - \Pi''_a\|_1 \leq 5\sqrt{\frac{\mathrm{I}(A:B)}{\lambda}} + \frac{2(1+\log e)}{\lambda} \,. \tag{17}$$

Recalling that $h'(R,a) \sim \Pi''_a$, we have

$$\begin{aligned}\mathrm{E}_R[\mathrm{E}_A[f(A,h'(R,A))]] &= \mathrm{E}_A[\mathrm{E}_R[f(A,h'(R,A))]] \\ &= \sum_{a \in \mathscr{A}} p_a \, \mathrm{E}_R[f(a,h'(R,a))] \\ &= \sum_{a \in \mathscr{A}} p_a \sum_{b \in \mathscr{B}} \Pi''_a(b) f(a,b) \\ &\leq \sum_{a \in \mathscr{A}} \frac{p_a}{2} \|\Pi_a - \Pi''_a\|_1 + \sum_{a \in \mathscr{A}} p_a \sum_{b \in \mathscr{B}} \Pi_a(b) f(a,b) \\ &\leq \mathrm{E}_{A,B}[f(A,B)] + \frac{5}{2}\sqrt{\frac{\mathrm{I}(A:B)}{\lambda}} + \frac{1+\log e}{\lambda} \,,\end{aligned}$$

where the penultimate inequality holds because f takes values in $[0,1]$ and the final inequality follows from (17). Therefore, there exists some fixed $r_0 \in [0,1]$ such that

$$\mathrm{E}_A[f(A, h'(r_0, A))] \;\leq\; \mathrm{E}_{A,B}[f(A,B)] + \frac{5}{2}\sqrt{\frac{\mathrm{I}(A:B)}{\lambda}} + \frac{1+\log e}{\lambda}\,.$$

Let $g : \mathscr{A} \to \mathscr{B}$ be defined by $g(a) = h'(r_0, a)$ for $a \in \mathscr{A}$, and let $\mathscr{B}_0 \subseteq \mathscr{B}$ be the range of g. Since the algorithm for h' stops within 2^λ iterations by design, we have $|\mathscr{B}_0| \leq 2^\lambda$. Thus, the function g has all the desired properties. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/