Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity*

Amit Chakrabarti Yaoyun Shi Anthony Wirth Andrew Yao

Department of Computer Science, Princeton University
35 Olden Street, Princeton, NJ 08544
{amitc, shiyy, awirth, yao}@cs.princeton.edu

Abstract

Given m copies of the same problem, does it take m times the amount of resources to solve these m problems? This is the direct sum problem, a fundamental question that has been studied in many computational models. We study this question in the simultaneous message (SM) model of communication introduced by Yao [Y79].

The equality problem for n-bit strings is well known to have SM complexity $\Theta(\sqrt{n})$. We prove that solving m copies of the problem has complexity $\Omega(m\sqrt{n})$; the best lower bound provable using previously known techniques is $\Omega(\sqrt{mn})$. We also prove similar lower bounds on certain Boolean combinations of multiple copies of the equality function. These results can be generalized to a broader class of functions.

We introduce a new notion of informational complexity *which is related to SM complexity and has nice direct sum properties. This notion is used as a tool to prove the above results; it appears to be quite powerful and may be of independent interest.*

1 Introduction

We consider the problem of computing a function $f(x, y)$ whose inputs are split between two parties: Alice, who holds x, and Bob, who holds y. The two-party communication model was introduced by Yao [Y79] to study the number of bits Alice and Bob need to communicate to each other to compute f. In the same paper Yao also introduced the simultaneous message (SM) model in which Alice and Bob do not talk to each other, but send messages to a third party, a referee, who then computes f based on these messages. This SM model is the subject of this paper.

Kushilevitz and Nisan's book [KN97] provides a comprehensive treatment of communication complexity. The survey paper by Babai [B01] provides an overview of recent research.

Considerable effort has been spent in understanding the SM complexity of the (Boolean) equality function, EQ_n, defined as $\mathsf{EQ}_n(x, y) = 1$ iff $x = y$, for $x, y \in \{0, 1\}^n$. It is easy to show that deterministic protocols for EQ_n require n bits of communication, both in the two-party and the SM models. The problem becomes considerably more interesting if we allow randomized protocols (that err with low probability) with Alice and Bob flipping private coins to decide on their messages to the referee. Various SM protocols[1] for EQ_n that communicate only $O(\sqrt{n})$ bits were discovered by Ambainis [A96], Newman and Szegedy [NS96], and Babai and Kimmel [BK97], amongst others. The last two papers also prove that this is tight, and therefore that the SM complexity of EQ_n is $\Theta(\sqrt{n})$.

Babai and Kimmel [BK97] actually prove the following general lower bound on SM complexity:

Theorem 1.1 ([BK97]) *For any f, its SM complexity $C(f)$ and its deterministic two-party communication complexity $C_0(f)$ are related as follows: $C(f) = \Omega(\sqrt{C_0(f)})$.*

To our knowledge, the lower bound results on some explicit functions proven here are the first that are stronger than what [BK97] can prove.

The *direct sum problem* asks whether m copies of a problem require m times as much resources to solve as one copy. For a function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ and integer m, the direct sum $f^m : \{0, 1\}^{mn} \times \{0, 1\}^{mn} \to \{0, 1\}^m$ is defined as the function obtained by concatenating m copies of f with separate inputs for each copy. For communication

*This work was supported in part by NSF Grant CCR-96-23768, NSF Grant CCR-98-20855, ARO Grant DAAH04-96-1-0181, NEC Research Institute, and a Gordon Wu Fellowship.

[1] From here on, we shall assume that our SM model is randomized, with private coins for Alice and Bob.

complexity, it was proved by Feder, Kushilevitz, Naor and Nisan [FKNN91] that in the deterministic two-party model, there exists a partial function f with $C(f) = \Theta(\log n)$, whereas solving m copies takes only $C(f^m) = O(m + \log m \cdot \log n)$; thus the average cost per copy is $O(1)$ for large m. A lower bound $C(f^m) \geq m(\sqrt{C(f)/2} - \log n - O(1))$ was also proved. For the ε-error randomized two-party model, [FKNN91] showed that for the equality problem, $C_R(\mathsf{EQ}_n^m) = O(m + \log n)$, in contrast to the single copy case $C_R(\mathsf{EQ}_n) = \Theta(\log n)$. For the one-round deterministic model, [FKNN91] proved that, for any f, essentially $C(f^m) = mC(f)$. For the two-round deterministic model, Karchmer, Kushilevitz and Nisan [KKN92] showed that essentially $C(f^m) = mC(f)$ for any f. This latter relation was also shown true in [KKN92] for nondeterministic communication models.

The communication complexity model for *relations* was invented by Karchmer and Wigderson [KW90]: Alice and Bob want to agree on an output z such that a relation $R(x, y, z)$ is true. The direct sum problem for this model is wide open and is closely connected to the Boolean circuit depth problem [KW90, KRW91]; resolving it would lead to important results such as $\mathsf{NC}^1 \neq \mathsf{NC}^2$. Results for this model were obtained in [KRW91, KKN92] and by Edmonds, Impagliazzo, Rudich, and Sgall [EIRS91]. In this paper we will not study communication complexity for relations.

1.1 Results

In this paper we address the direct sum problem for the case of SM complexity. Henceforth $C(f)$ denotes the SM complexity of f. For the equality function we show

Theorem 1.2 $C(\mathsf{EQ}_n^m) = \Omega(m\sqrt{n})$.

Note that an application of Theorem 1.1 only yields a lower bound of $\Omega(\sqrt{mn})$.

We also obtain results for certain Boolean combinations of several equality functions. Suppose Alice and Bob receive n-bit strings $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$ respectively; define $b_i = 1$ if $x_i = y_i$ and $b_i = 0$ otherwise. Also define the functions

- $\mathsf{OREQ}_n^m(x_1, \ldots, x_m, y_1, \ldots, y_m) = \bigvee_{i=1}^m b_i$.
- $\mathsf{XOREQ}_n^m(x_1, \ldots, x_m, y_1, \ldots, y_m) = \bigoplus_{i=1}^m b_i$.
- $\mathsf{MAJEQ}_n^m(x_1, \ldots, x_m, y_1, \ldots, y_m) = 1$ iff $|\{i : b_i = 1\}| \geq \frac{1}{2}m$.
- $\mathsf{THREQ}_n^{m,k}(x_1, \ldots, x_m, y_1, \ldots, y_m) = 1$ iff $|\{i : b_i = 1\}| \geq k$.

Theorem 1.3 *Let $0 < \lambda < \frac{1}{2}$ be a constant and suppose $m \leq \lambda 2^n$. If f is any of OREQ_n^m, XOREQ_n^m and MAJEQ_n^m, we have $C(f) = \Omega(m\sqrt{n})$. We also have $C(\mathsf{THREQ}_n^{m,k}) = \Omega((m-k+1)\sqrt{n})$.*

The above results are obtained using an information theoretic approach. For any communication problem f we define a quantity called its *informational complexity*, denoted $IC(f)$, which describes the minimum requirement on the capacity of the "communication channels" between Alice/Bob and the referee for a uniformly distributed input (for precise definitions please see Section 2). This quantity never exceeds the SM complexity $C(f)$.

For $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $X \subseteq \{0,1\}^n$, let $f|_X$ be the restriction of f to $X \times X$. We define $\tilde{C}(f) = \min_{X : |X| \geq (2/3)2^n} C(f|_X)$. We call f *robust* if $\tilde{C}(f) = \Omega(C(f))$. From the downward self-reducibility of equality, it follows that if $|X| \geq \frac{2}{3} \cdot 2^n$ then $C(\mathsf{EQ}_n) \leq C(\mathsf{EQ}_n|_X) + 1 \leq 2C(\mathsf{EQ}_n|_X)$; thus EQ_n is robust.

As mentioned above, $IC(f) \leq C(f)$. Our Main Lemma goes in the other direction.

Lemma 1.4 (Main Lemma) *Every function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ satisfies $IC(f) = \Omega(\tilde{C}(f)) - O(\log n)$. In particular, if f is robust, $IC(f) = \Omega(C(f)) - O(\log n)$.*

This lemma is a powerful tool for proving lower bounds on SM complexity, because informational complexity is a "well-behaved" measure:

Theorem 1.5 (Additivity Theorem) *For any f and any m, $IC(f^m) \geq m \cdot IC(f)$.*

We remark that our informational complexity approach also works for the one round two-party communication model, where Alice sends a single message to Bob who then computes the function. A straightforward analogue of the concept of informational complexity can be defined and used in analogues of Lemma 1.4 and Theorem 1.5. We will not elaborate further on this topic here.

The rest of this paper is organized as follows. Section 2 defines several basic concepts. Section 3 contains a proof of the Main Lemma, divided into several subsections. Finally, Section 4 proves Theorem 1.5, and then Theorems 1.2 and 1.3 by applying the Main Lemma.

2 Preliminaries

2.1 Conventions

The following conventions are used throughout the paper. Random variables are denoted by boldface characters. For random variables $\mathbf{x}$ and $\mathbf{y}$, $H(\mathbf{x})$ denotes the Shannon entropy of $\mathbf{x}$, $H(\mathbf{x} \,|\, \mathbf{y})$ denotes the entropy of $\mathbf{x}$ conditioned on $\mathbf{y}$ and $I(\mathbf{x}, \mathbf{y})$ denotes the mutual information between $\mathbf{x}$ and $\mathbf{y}$. Recall that

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x} \,|\, \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y} \,|\, \mathbf{x}).$$

Matrices are denoted by uppercase letters and their entries by the corresponding lowercase letters with indices in parentheses. Thus, if A is a matrix, its $(i,j)^{\text{th}}$ entry is written as $a(i,j)$. The set $\{1,2,\ldots,k\}$ is denoted by $[k]$. All logarithms are to base 2.

2.2 Definitions

Consider a Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. A simultaneous message (SM) protocol for f with error bound ε is a quintuple (k,ℓ,P,Q,R) with k,ℓ integers and P,Q,R matrices of shape $2^n \times k$, $2^n \times \ell$ and $k \times \ell$ respectively. We assume that the rows of P and Q are indexed by elements of $\{0,1\}^n$. P and Q are required to be stochastic matrices (i.e. their rows sum to 1) and R is required to have entries in $\{0,1\}$. Finally, the product $F^* = PRQ^{\mathsf{T}}$ is required to satisfy

$$|f^*(x,y) - f(x,y)| < \varepsilon \qquad \text{for all } x,y \in \{0,1\}^n. \quad (1)$$

The protocol works as follows: suppose Alice receives $x \in \{0,1\}^n$ and Bob receives $y \in \{0,1\}^n$. Alice interprets row x of matrix P as a probability distribution over the set $[k]$ and picks a "message" from $[k]$ according to this distribution. Bob similarly encodes y into a message from $[\ell]$. They both send their messages to a referee, thereby communicating $\lceil \log k \rceil$ and $\lceil \log \ell \rceil$ bits respectively. The referee uses the messages to index into matrix R and outputs the corresponding entry of R. [2] Equation (1) says that the probability that the referee outputs $f(x,y)$ exceeds $1-\varepsilon$.

We denote the set of all such ε-error protocols for f by $\Pi_\varepsilon(f)$. For a protocol $\mathcal{P} = (k,\ell,P,Q,R) \in \Pi_\varepsilon(f)$ we define $\text{cost}_A(\mathcal{P}) = \lceil \log k \rceil$, $\text{cost}_B(\mathcal{P}) = \lceil \log \ell \rceil$ and $\text{cost}(\mathcal{P}) = \text{cost}_A(\mathcal{P}) + \text{cost}_B(\mathcal{P})$. The *simultaneous message communication complexity* (or SM complexity) of f is defined to be

$$C(f) = \min_{\mathcal{P} \in \Pi_{1/3}(f)} \text{cost}(\mathcal{P}).$$

For protocol $\mathcal{P}$, let $\mathbf{x},\mathbf{y}$ be random variables uniformly distributed over $\{0,1\}^n$. Let random variables $\mathbf{u}$ and $\mathbf{v}$ denote the messages that Alice and Bob (respectively) generate upon receiving inputs $\mathbf{x}$ and $\mathbf{y}$. Clearly the distributions of $\mathbf{u}$ and $\mathbf{v}$ depend on the matrices P and Q. The *informational cost*, $\text{icost}(\mathcal{P})$, of the protocol is defined to be $\text{icost}_A(\mathcal{P}) + \text{icost}_B(\mathcal{P})$ where $\text{icost}_A(\mathcal{P}) = I(\mathbf{x},\mathbf{u})$ and $\text{icost}_B(\mathcal{P}) = I(\mathbf{y},\mathbf{v})$. The *informational complexity* of f is defined to be

$$IC(f) = \min_{\mathcal{P} \in \Pi_{1/3}(f)} \text{icost}(\mathcal{P}).$$

[2] One can make a more general definition where R has entries in $[0,1]$ and the referee makes a probabilistic decision whether to output 0 or 1. However it is well known [NS96] that any such protocol can be converted to a protocol with deterministic referee by at most doubling the error.

We remark that the choice of the constant $1/3$ in the above definitions is arbitrary; the error can be reduced from any constant to any smaller constant by trivially "repeating" the protocol and increasing the cost or informational cost only by a constant factor.

For studying the direct sum question, we also need to define protocols for functions with multiple bits of output. For a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, we require Alice and Bob to behave as before, but let the referee use m 0-1 matrices $R_1,\ldots,R_m$, one for each bit. We call this an ε-error protocol if, for each i, the protocol given by R_i is an ε-error protocol for the i^{th} bit of f. [3] The concepts cost, icost, C and IC are defined exactly as before.

Proposition 2.1 *For any f, $IC(f) \le C(f)$.*

Proof: Consider any protocol $\mathcal{P} = (k,\ell,P,Q,R) \in \Pi_{1/3}(f)$ and let $\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v}$ be as above. Then $\text{icost}(\mathcal{P}) = H(\mathbf{u}) - H(\mathbf{u}\,|\,\mathbf{x}) + H(\mathbf{v}) - H(\mathbf{v}\,|\,\mathbf{y}) \le H(\mathbf{u}) + H(\mathbf{v}) \le \log k + \log \ell \le \lceil \log k \rceil + \lceil \log \ell \rceil = \text{cost}(\mathcal{P})$. □

Thus, a lower bound on $IC(f)$ automatically implies one on $C(f)$.

3 Proof of the Main Lemma

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function and let $\mathcal{P} = (k,\ell,P,Q,R) \in \Pi_\varepsilon(f)$ be an ε-error SM protocol for f, with ε a sufficiently small constant. We prove the Main Lemma in two major stages. In the first stage we prove that a protocol satisfying a property we call *uniformity* can be converted into a new protocol that has cost comparable to the informational cost of the original protocol. In the second stage we show how to convert a general protocol into one that is *near-uniform*, increasing the informational cost essentially by a constant factor. It turns out that the proof for the uniform case can be adjusted to work even for near-uniform protocols.

3.1 Outline

For each $x \in \{0,1\}^n$ and $u \in [k]$ let us define the following sets, which depend on the protocol $\mathcal{P}$.

$$\begin{aligned} U_x &= \{u' \in [k] : p(x,u') > 0\} \\ X_u &= \{x' \in \{0,1\}^n : p(x',u) > 0\} \end{aligned}$$

Definition 3.1 *$\mathcal{P}$ is said to be s-*uniform for Alice, *for an integer s, if each of her messages $u \in [k]$ satisfies*

- $s \le |X_u| < 2s$.

[3] Another plausible definition would have required the referee to correctly output the *entire* m-bit result with probability more than $1-\varepsilon$. Our weaker definition only makes our lower bound results stronger.

- *For all $x_1, x_2 \in X_u$ we have $p(x_1, u) = p(x_2, u)$.*

Uniformity for Bob is defined similarly.

Lemma 3.2 *For a sufficiently small constant ε, suppose $\mathcal{P} \in \Pi_\varepsilon(f)$ is s-uniform for Alice and s'-uniform for Bob, for some s and s'. Then there is a protocol $\mathcal{P}' \in \Pi_{1/3}(f)$ with $\mathrm{cost}(\mathcal{P}') \le \mathrm{icost}(\mathcal{P}) + O(\log n)$.*

As mentioned above, this lemma is not strong enough for our purposes. We need a further definition and a similar, but stronger, lemma. Recall the definition of $\mathbf{u}$ from Section 2.

Definition 3.3 *$\mathcal{P}$ is said to be* near-s-uniform *for Alice, with irregularity δ, for an integer s and a constant $0 < \delta < 1$, if there is a set $Z \subseteq [k]$ such that*

- *$s \le |X_u| < 2s$, for all $u \in [k] \setminus Z$.*
- *$p(x_1, u) = p(x_2, u)$, for all $u \in [k] \setminus Z$ and all $x_1, x_2 \in X_u$.*
- *$\Pr[\mathbf{u} \in Z] < \delta$.*

The messages in Z will be called irregular *messages and the rest* regular. *Near-uniformity for Bob is defined similarly.*

For brevity, we shall call a protocol which is near-s-uniform for Alice and near-s'-uniform for Bob, with irregularity δ on each side, a (δ, s, s')-near-uniform protocol.

Lemma 3.4 *Given a protocol $\mathcal{P} \in \Pi_\varepsilon(f)$ and a constant $0 < \delta < 1$, there is a protocol $\mathcal{P}' \in \Pi_\varepsilon(f)$ and integers s, s', such that $\mathcal{P}'$ is (δ, s, s')-near-uniform and $\mathrm{icost}(\mathcal{P}') = O(\mathrm{icost}(\mathcal{P})) + O(\log n)$.*

Lemma 3.5 *For sufficiently small constants ε and δ, suppose $\mathcal{P}' \in \Pi_\varepsilon(f)$ is (δ, s, s')-near-uniform for some s and s'. Then there is an $X \subseteq \{0,1\}^n$ and a protocol $\mathcal{P}'' \in \Pi_{1/3}(f|_X)$ such that $|X| \ge \frac{2}{3} \cdot 2^n$ and $\mathrm{cost}(\mathcal{P}'') \le \mathrm{icost}(\mathcal{P}') + O(\log n)$.*

Proof of Lemma 1.4 (Main Lemma): Suppose, for the moment, that the above lemmas have been proven. Applying Lemma 3.4 to a protocol $\mathcal{P} \in \Pi_\varepsilon(f)$ and then applying Lemma 3.5 to the resulting protocol $\mathcal{P}'$ gives a protocol $\mathcal{P}''$ with

$$\mathrm{cost}(\mathcal{P}'') = O(\mathrm{icost}(\mathcal{P})) + O(\log n),$$

and $\mathcal{P}'' \in \Pi_{1/3}(f|_X)$ for an X of size at least $\frac{2}{3} \cdot 2^n$. Clearly

$$\mathrm{cost}(\mathcal{P}'') \ge C(f|_X) \ge \tilde{C}(f).$$

Therefore $\mathrm{icost}(\mathcal{P}) = \Omega(\tilde{C}(f)) - O(\log n)$. □

In the next three subsections we prove Lemmas 3.2, 3.5 and 3.4 respectively.

3.2 Uniform Protocols

We set ε to a sufficiently small constant; for definiteness' sake

$$\varepsilon = \frac{1}{3^4 \cdot 5^6}. \tag{2}$$

Suppose $\mathcal{P}$ is s-uniform for Alice and s'-uniform for Bob. Let $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}$ be as defined in Section 2. By the definition of uniformity, we clearly have

$$n - \log s - 1 < I(\mathbf{x}, \mathbf{u}) \le n - \log s. \tag{3}$$

The idea of the proof is to replace Alice's strategy by a new one in which her message set reduces in size from k to about $O(2^n/s)$, increasing the error by a constant factor. Alice's new message set will be an appropriate random sample of $[k]$, possibly with some elements of $[k]$ appearing multiple times (these multiple occurrences are, of course, considered distinct messages in the new protocol). With the right choice of parameters the new protocol will have cost on Alice's side comparable to $\mathrm{icost}_A(\mathcal{P})$.

Fix inputs x, y for Alice and Bob and let P_x, Q_y be the corresponding rows of matrices P, Q. Since $\mathcal{P}$ is a valid ε-error protocol, we have $|P_x R Q_y^{\mathsf{T}} - f(x,y)| < \varepsilon$. Consider the vector $R_y^* = RQ_y^{\mathsf{T}}$; its entries are to be thought of as being indexed by Alice's messages. Call a message $u \in U_x$ *defective* for y if the u^{th} entry of R_y^* differs from $f(x,y)$ by more than $\sqrt{\varepsilon}$. Note that this u^{th} entry is the probability that the referee outputs 1 given that Alice sent message u. Messages not in U_x are defined to be not defective. Let $T = (u_1, u_2, \ldots)$ be a finite sequence of messages and let $|T|$ denote its length.

Definition 3.6 *We say that T is* good *for (x, y) if*

$$|\{i : u_i \text{ is defective for } y\}| \le 4\sqrt{\varepsilon} \cdot |T \cap U_x|,$$

where $T \cap U_x$ is the subsequence of T consisting of only those elements that are in U_x.

Note that only those elements of T which belong to U_x affect whether or not T is good for (x, y).

We would like to prove that there exists a sequence T of appropriate length that is good for all (x, y). We use the probabilistic method: we argue that a T chosen at random in a suitable manner is good for any fixed (x, y) with probability greater than $1 - 2^{-2n}$. Since there are only 2^{2n} pairs (x, y) the result follows.

Define a random sequence $\mathbf{T}(\tau) = (\mathbf{u}_1, \ldots, \mathbf{u}_\tau)$ where $\mathbf{u}_i$ are independent random variables distributed identically to $\mathbf{u}$. The length τ is a parameter. Also let $\tilde{\mathbf{u}}_x$ be a random variable distributed according to P_x.

Lemma 3.7 *Fix $u \in U_x$. Then*

$$\frac{1}{2} \Pr[\tilde{\mathbf{u}}_x = u] < \Pr[\mathbf{u} = u \mid \mathbf{u} \in U_x] < 2 \Pr[\tilde{\mathbf{u}}_x = u].$$

Proof: Recall that the protocol is s-uniform for Alice. We have

$$\begin{aligned}\Pr[\mathbf{u} = u \mid \mathbf{u} \in U_x] &= \frac{\frac{1}{2^n}|X_u|p(x,u)}{\sum_{u' \in U_x} \frac{1}{2^n}|X_{u'}|p(x,u')} \\ &< \frac{2sp(x,u)}{\sum_{u' \in U_x} sp(x,u')} \\ &= 2p(x,u) \\ &= 2\Pr[\tilde{\mathbf{u}}_x = u].\end{aligned}$$

The other inequality follows similarly. □

Lemma 3.8 *Fix an index set $I \subseteq [\tau]$ and let $\mathbf{T} = \mathbf{T}(\tau)$. Let $\mathbf{T}_I$ denote the subsequence of $\mathbf{T}$ consisting of the elements indexed by I. Then*

$$\Pr[\mathbf{T} \text{ is good} \mid \mathbf{T} \cap U_x = \mathbf{T}_I] \geq 1 - e^{-\frac{1}{2}\sqrt{\varepsilon}|I|}$$

Proof: Conditioned on $\mathbf{T} \cap U_x = \mathbf{T}_I$, it is the case that $\mathbf{T}$ is good iff $\mathbf{T}_I$ is. Now under this condition each element of $\mathbf{T}_I$ is distributed identically to $(\mathbf{u} \mid \mathbf{u} \in U_x)$. By Lemma 3.7 we see that the expected number of defective elements in $\mathbf{T}_I$ is at most $2|I|$ times the probability that $\tilde{\mathbf{u}}_x$ is defective.

Note that $|P_x R_y^* - f(x,y)| < \varepsilon$, by (1); therefore $\tilde{\mathbf{u}}_x$ is defective with probability at most $\sqrt{\varepsilon}$. By standard Chernoff bounds, the probability that $\mathbf{T}_I$ has at least $4\sqrt{\varepsilon}|I|$ defective elements is at most $e^{-\frac{1}{2}\sqrt{\varepsilon}|I|}$ and the result follows. □

Lemma 3.9 *Set $t = 2^{n+3}n/(s\sqrt{\varepsilon})$. Then there exists a sequence T with $|T| = t$ that is good for all (x,y).*

Proof: Consider the random sequence $\mathbf{T}(t)$ and fix a particular (x,y). Each $\mathbf{u}_i$ lands in the set U_x with probability $\sum_{u \in U_x} \frac{1}{2^n}|X_u|p(x,u) \geq s/2^n$, by the s-uniformity condition. Therefore $\mathrm{E}[|\mathbf{T}(t) \cap U_x|] \geq ts/2^n = 8n/\sqrt{\varepsilon}$ and using a standard Chernoff bound, we have

$$\Pr\left[|\mathbf{T}(t) \cap U_x| \leq 4n/\sqrt{\varepsilon}\right] \leq e^{-n/\sqrt{\varepsilon}}. \qquad (4)$$

Let $I \subseteq [t]$ be an index set with $|I| \geq 4n/\sqrt{\varepsilon}$. By Lemma 3.8

$$\Pr[\mathbf{T}(t) \text{ is good} \mid \mathbf{T}(t) \cap U_x = \mathbf{T}(t)_I] \geq 1 - e^{-2n}.$$

Therefore,

$$\begin{aligned}&\Pr[\mathbf{T}(t) \text{ is good}] \\ &\geq \left(1 - e^{-2n}\right) \sum_{I \subseteq [t]:\, |I| \geq 4n/\sqrt{\varepsilon}} \Pr[\mathbf{T}(t) \cap U_x = \mathbf{T}(t)_I] \\ &= (1 - e^{-2n}) \cdot \Pr\left[|\mathbf{T}(t) \cap U_x| \geq 4n/\sqrt{\varepsilon}\right] \\ &> 1 - 2^{-2n}\end{aligned}$$

where the last inequality follows from (2) and (4).

Since there are only 2^{2n} pairs (x,y) it follows that $\mathbf{T}(t)$ is good for all (x,y) with positive probability. This proves the lemma. □

We are ready to prove the main result of this subsection.

Proof of Lemma 3.2: Let T be a sequence of length $t = 2^{n+3}n/(s\sqrt{\varepsilon})$ that is good for all (x,y). We change Alice's strategy as follows: on input x, Alice chooses one of the elements in the subsequence $T \cap U_x$ uniformly at random and sends it to the referee. Let $\mathcal{P}'$ be this new protocol.

To analyse the correctness of $\mathcal{P}'$, consider an arbitrary input (x_0, y_0). Since T is good for (x_0, y_0), the message chosen by Alice is defective for y_0 with probability at most $4\sqrt{\varepsilon}$. By definition of defectiveness, if Alice's message is not defective, the referee's output is wrong with probability at most $\sqrt{\varepsilon}$. Therefore $\mathcal{P}' \in \Pi_{5\sqrt{\varepsilon}}(f)$. Applying (3) we have $\mathrm{cost}_A(\mathcal{P}') = \lceil \log t \rceil \leq I(\mathbf{x},\mathbf{u}) + 4 + \log n - \log\sqrt{\varepsilon} = \mathrm{icost}_A(\mathcal{P}) + O(\log n)$.

Note that this transformation has not affected Bob's part of the protocol. Therefore, we can do another such transformation for Bob, this time not affecting Alice, and obtain a protocol $\mathcal{P}''$ with $\mathrm{cost}_A(\mathcal{P}'') = \mathrm{cost}_A(\mathcal{P}')$ and $\mathrm{cost}_B(\mathcal{P}'') = \mathrm{icost}_B(\mathcal{P}') + O(\log n) = \mathrm{icost}_B(\mathcal{P}) + O(\log n)$. The error now increases to at most $5\sqrt{5\sqrt{\varepsilon}}$ which, by (2), is $1/3$ and we are done. □

3.3 Near-uniform protocols

Now suppose $\mathcal{P}'$ is a (δ, s, s')-near-uniform protocol in $\Pi_\varepsilon(f)$. Let Z denote the set of irregular messages of Alice. Call an input x for Alice *active* if $\Pr[\tilde{\mathbf{u}}_x \in Z] < \sqrt{\delta}$. From the definition of near-uniformity it follows that at most a $\sqrt{\delta}$ fraction of Alice's inputs are inactive. The same is true of Bob's inputs as well. If we choose δ small enough, there exists an $X \subseteq \{0,1\}^n$ such that $|X| \geq \frac{2}{3} \cdot 2^n$ and inputs in $X \times X$ are active for both Alice and Bob.

We now transform $\mathcal{P}'$ into another protocol $\mathcal{P}''$ using a transformation like that in Section 3.2.

Proof of Lemma 3.5 (sketch): In essence, the proof technique of Lemma 3.2 works; we indicate the key differences. Define $V_x = U_x \setminus Z$ for every input x of Alice. Defectiveness is defined as before, except that messages not in V_x are defined to be not defective. We say that a sequence T of messages is *good* for (x,y) if

$$|\{i : u_i \text{ is defective for } y\}| \leq 6\sqrt{\varepsilon} \cdot |T \cap V_x|.$$

For an active x, if we fix a $u \in V_x$, we can show that

$$\frac{1}{3}\Pr[\tilde{\mathbf{u}}_x = u] < \Pr[\mathbf{u} = u \mid \mathbf{u} \in U_x] < 3\Pr[\tilde{\mathbf{u}}_x = u].$$

The proof is similar to that of Lemma 3.7 and uses the fact that δ is small. Next, we show that for such x

$$\Pr\left[\mathbf{T} \text{ is good} \mid \mathbf{T} \cap V_x = \mathbf{T}_I\right] \;\geq\; 1 - e^{-\frac{1}{2}\sqrt{\varepsilon}|I|},$$

where $\mathbf{T}$ and I are as in Lemma 3.8.

Set $t = 2^{n+4}n/(s\sqrt{\varepsilon})$. Note that this is twice as large as the t used in Lemma 3.9. We can show that, for any active x, $\mathrm{E}[|\mathbf{T}(t) \cap V_x|] \geq (1-\sqrt{\delta})ts/2^n \geq 8n/\sqrt{\varepsilon}$, for sufficiently small δ. Now, arguing exactly as in Lemma 3.9, we infer that there exists a sequence T with $|T| = t$ that is good for all (x, y) with x active. Finally, we change Alice's strategy as follows: on input x, she chooses one of the elements in $T \cap V_x$ uniformly at random and sends it to the referee.

Consider this new protocol's behaviour on an arbitrary input (x_0, y_0), with x_0 active for Alice. Alice's message is irregular with probability at most $\sqrt{\delta}$ and defective for y_0 with probability at most $6\sqrt{\varepsilon}$. When not irregular and not defective, her message causes the referee to err with probability at most $\sqrt{\varepsilon}$. Therefore the error of the new protocol is at most $7\sqrt{\varepsilon} + \sqrt{\delta}$.

As in the proof of Lemma 3.2, the cost of this protocol on Alice's side is $\lceil \log t \rceil \leq \mathrm{icost}_A(\mathcal{P}') + O(\log n)$.

Repeating the entire process for Bob gives a protocol $\mathcal{P}''$ whose error, on inputs in $X \times X$ (which are active for both Alice and Bob), can be made at most $1/3$ by choosing ε and δ appropriately. Thus $\mathcal{P}'' \in \Pi_{1/3}(f|_X)$ and we are done. □

3.4 General protocols

We now prove Lemma 3.4.

Call a message u *smooth* if $p(x_1, u) = p(x_2, u)$ for all $x_1, x_2 \in X_u$. Also, define the *degree* of a message u to be $|X_u|$. We convert the general protocol $\mathcal{P}$ into a near-uniform protocol $\mathcal{P}'$ in two steps, each time creating several new messages to replace one message. The first step smoothens the messages; the second equalizes the degrees up to a factor of 2.

Smoothening

We "split" Alice's messages one at a time. Fix a message $u \in [k]$ and let $\pi_u = \max_{x \in X_u} p(x, u)$. Let j be the integer that satisfies

$$\frac{1}{2^j} \;\leq\; \pi_u \;<\; \frac{1}{2^{j-1}}. \tag{5}$$

We shall replace message u by $n + c + 2$ new messages, $\{u^{(0)}, u^{(1)}, \ldots, u^{(n+c)}, u_{\mathrm{irr}}\}$, for some constant c, common to all the u's. The message u_{irr} will eventually end up an irregular message.

Let P' denote the matrix for Alice's new strategy. We "distribute" the probability $p(x, u)$ amongst the new probabilities given by P' as follows. Consider the binary representation of $p(x, u)$:

$$p(x, u) \;=\; \sum_{i=0}^{\infty} \frac{a_{j+i}}{2^{j+i}},$$

where $a_i \in \{0, 1\}$ (It does not matter that the a_i's might not be uniquely determined.) We set

$$p'\big(x, u^{(i)}\big) \;=\; \frac{a_{j+i}}{2^{j+i}} \tag{6}$$

for $0 \leq i \leq n + c$, and

$$p'(x, u_{\mathrm{irr}}) \;=\; \sum_{i=n+c+1}^{\infty} \frac{a_{j+i}}{2^{j+i}}. \tag{7}$$

The referee's matrix is modified so that, upon receiving any of these newly created messages, the referee behaves exactly as if the message u were received in the earlier protocol. This ensures that the modified protocol behaves identically to the original one.

We perform the above modification for each $u \in [k]$. The end result is a protocol with $k(n + c + 2)$ messages for Alice that behaves identically to the original one.

Let $Z_1 = \{u_{\mathrm{irr}} : u \in [k]\}$. From (6) we see that all the newly created messages not in Z_1 are smooth. Let $\mathbf{u}$ and $\mathbf{u}'$ denote the random messages that Alice sends in the original and modified protocols, respectively, upon receiving a uniformly distributed random input $\mathbf{x}$. We argue that $\mathbf{u}'$ is unlikely to land in Z_1. Consider an arbitrary $u \in [k]$. By (7),

$$p'(x, u_{\mathrm{irr}}) \;\leq\; \sum_{i=n+c+1}^{\infty} \frac{1}{2^{j+i}} \;=\; \frac{1}{2^{j+n+c}} \;\leq\; \frac{\pi_u}{2^{n+c}},$$

where the last inequality follows from (5). From the definition of π_u, we have

$$\pi_u \;\leq\; \sum_{x \in X_u} p(x, u) \;=\; 2^n \Pr[\mathbf{u} = u].$$

Therefore $p'(x, u_{\mathrm{irr}}) \leq \Pr[\mathbf{u} = u]/2^c$. Now

$$\begin{aligned} \Pr[\mathbf{u}' \in Z_1] \;&=\; \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \sum_{u \in [k]} p'(x, u_{\mathrm{irr}}) \\ &\leq\; \sum_{u \in [k]} \frac{\Pr[\mathbf{u} = u]}{2^c} \\ &=\; \frac{1}{2^c}. \end{aligned} \tag{8}$$

We must show that our modification of the protocol does not increase its informational cost too much. The lemma below allows us to bound the change.

Lemma 3.10 *If we replace each of Alice's messages by at most K new messages then the gain in the mutual information $I(\mathbf{x},\mathbf{u})$ is at most $\log K$.*

Proof: Consider a particular message u that is split into K_u messages, $\{u^{(1)},u^{(2)},\ldots,u^{(K_u)}\}$, $K_u \le K$, with associated new probabilities $p'\big(x,u^{(i)}\big)$ and new distribution $\mathbf{u}'$. Let $p(u)=\Pr[\mathbf{u}=u]$ and $p'\big(u^{(i)}\big)=\Pr[\mathbf{u}'=u^{(i)}]$. Then $H(\mathbf{u})=\sum_{u\in[k]}-p(u)\log p(u)$.

Because of the way the probabilities are redistributed, $p(u)=\sum_{i\in[K_u]}p'\big(x,u^{(i)}\big)$. Therefore

$$\begin{aligned}
H(\mathbf{u}') &= \sum_{u\in[k]}\sum_{i\in[K_u]} -p'\big(u^{(i)}\big)\log p'\big(u^{(i)}\big)\\
&= \sum_u K_u \sum_i -\frac{1}{K_u}p'\big(u^{(i)}\big)\log p'\big(u^{(i)}\big)\\
&\le \sum_u K_u\cdot\left(-\frac{p(u)}{K_u}\log\frac{p(u)}{K_u}\right) \qquad (9)\\
&= \sum_u -p(u)\,(\log p(u)-\log K_u)\\
&\le H(\mathbf{u})+\log K,
\end{aligned}$$

where (9) is obtained by applying Jensen's inequality to the inner summations.

It is also clear that, for any fixed x, $H(\mathbf{u}'\,|\,x)\ge H(\mathbf{u}\,|\,x)$. Therefore $H(\mathbf{u}'\,|\,\mathbf{x})\ge H(\mathbf{u}\,|\,\mathbf{x})$. Thus,

$$\begin{aligned}
I(\mathbf{x},\mathbf{u}') &= H(\mathbf{u}')-H(\mathbf{u}'\,|\,\mathbf{x})\\
&\le H(\mathbf{u})+\log K-H(\mathbf{u}\,|\,\mathbf{x})\\
&= I(\mathbf{x},\mathbf{u})+\log K.
\end{aligned}$$

□

Since each message is split into $n+c+2$ new messages, we can bound the increase in the mutual information by $\log(n+c+2)=O(\log n)$.

Equalizing degrees of messages

We now have a protocol in which all messages, except those in Z_1, are smooth. Their degrees, however, could vary considerably. We replace each message not in Z_1 with a set of new messages, but this time we also partition the inputs for each message so that the degrees become more balanced.

For every message $u\notin Z_1$, let $I(u)=H(\mathbf{x})-H(\mathbf{x}\,|\,u)$; since u is smooth

$$I(u) \;=\; n-\log|X_u|. \qquad (10)$$

Let $Z_2=\{u:\ u\notin Z_1 \text{ and } I(u)>bI(\mathbf{x},\mathbf{u})\}$, for some constant b. The messages in Z_2 will not be split and will end up as irregular messages. Since we have $I(\mathbf{x},\mathbf{u})=\sum_{u\in U}\Pr[\mathbf{u}=u]\,I(u)$, it is clear that

$$\Pr[\mathbf{u}\in Z_2] \;\le\; 1/b. \qquad (11)$$

Now let Z, the final set of irregular messages, be defined as $Z=Z_1\cup Z_2$. Combining (8) and (11), and setting b and c to large enough constants, we can ensure that

$$\Pr[\mathbf{u}\in Z] \;\le\; \frac{1}{2^c}+\frac{1}{b} \;<\; \delta$$

which bounds the irregularity of the final protocol.

Let $s=2^{n-bI(\mathbf{x},\mathbf{u})}$. By (10), we have $|X_u|\ge s$ for all $u\notin Z_2$. We split every $u\notin Z_2$ into $K_u=\lfloor|X_u|/s\rfloor$ new messages, $\{u^{(1)},\ldots,u^{(K_u)}\}$. The probabilities are distributed as follows: partition the inputs in X_u into K_u parts; if x is in the i^{th} part then assign $p'\big(x,u^{(i)}\big)=p(x,u)$, otherwise assign $p'\big(x,u^{(i)}\big)=0$. This partitioning can be arranged so that the degrees of the new messages (i.e., the sizes of the parts) are all in the range $[s,2s)$. Clearly the new messages are all smooth and so we have a protocol that is near-s-uniform for Alice with irregularity δ.

When a message u is split, it is split into

$$K_u \;=\; \lfloor|X_u|/s\rfloor \;\le\; 2^n/s \;=\; 2^{bI(\mathbf{x},\mathbf{u})}$$

new messages. By Lemma 3.10 the gain in mutual information that results is at most $\log K_u \le bI(\mathbf{x},\mathbf{u}) = O(\mathrm{icost}_A(\mathcal{P}))+O(\log n)$.

Lastly, we repeat the entire transformation for Bob to obtain a (δ,s,s')-near-uniform protocol $\mathcal{P}'$, with $\mathrm{icost}(\mathcal{P}')=O(\mathrm{icost}(\mathcal{P}))+O(\log n)$, that behaves identically to $\mathcal{P}$. This completes the proof of Lemma 3.4.

4 SM complexity lower bounds

In this section we prove the Additivity Theorem for informational complexity. We also show that the m-fold increase in informational complexity holds even when EQ_n^m is composed with certain specific Boolean functions. Together with the Main Lemma, these results imply communication complexity lower bounds on EQ_n^m and on the Boolean compositions.

4.1 Direct sum problems

The most obvious way to deal with a direct sum problem is to repeat a protocol for the base problem. Therefore, for any f, $IC(f^m)=O(m)\cdot IC(f)$. What is interesting is that this is essentially the best we can hope for; this is what the Additivity Theorem tells us.

Proof of Theorem 1.5 (Additivity Theorem): We use the symbols x_i,y_i to denote fixed strings in $\{0,1\}^n$ and the symbols $\mathbf{x}_i,\mathbf{y}_i$ to denote random strings distributed uniformly in $\{0,1\}^n$.

Let $\mathcal{P}\in\Pi_{1/3}(f^m)$. For $1\le i\le m$ and arbitrary strings $a=x_1x_2\ldots x_{i-1}\in\{0,1\}^{(i-1)\cdot n}$ and $b=$

$y_1 y_2 \ldots y_{i-1} \in \{0,1\}^{(i-1)\cdot n}$, define the protocol $\mathcal{P}^{(i)}_{a,b}$ for f as follows: on receiving inputs x and y, Alice and Bob simulate $\mathcal{P}$ on the random inputs $x_1 \ldots x_{i-1} x\, \mathbf{x}_{i+1} \ldots \mathbf{x}_m$ and $y_1 \ldots y_{i-1} y\, \mathbf{y}_{i+1} \ldots \mathbf{y}_m$ respectively, and the referee uses the matrix for the i^{th} bit of output to determine f. Clearly, the referee will correctly output $f(x,y)$ with probability at least $2/3$, whence $\mathcal{P}^{(i)}_{a,b} \in \Pi_{1/3}(f)$.

Consider the probability distribution of Alice's message when she follows protocol $\mathcal{P}$ with the random input $z_1 \ldots z_j\, \mathbf{z}_{j+1} \ldots \mathbf{z}_m$. Let $H_j(z_1, \ldots, z_j)$ denote the entropy of this distribution and let h_j denote the expectation $\mathrm{E}[H_j(\mathbf{z}_1, \ldots, \mathbf{z}_j)]$. It is easy to see that

$$\mathrm{icost}_A(\mathcal{P}) \;=\; h_0 - h_m. \tag{12}$$

Further,

$$\begin{aligned}\mathrm{icost}_A\big(\mathcal{P}^{(i)}_{a,b}\big) \;&=\; H_{i-1}(x_1, \ldots, x_{i-1}) \\ &\quad -\mathrm{E}_{\mathbf{x}}[H_i(x_1, \ldots, x_{i-1}, \mathbf{x})],\end{aligned}$$

whence

$$\mathrm{E}_{\mathbf{a}}\left[\mathrm{icost}_A\big(\mathcal{P}^{(i)}_{\mathbf{a},b}\big)\right] \;=\; h_{i-1} - h_i, \tag{13}$$

where the expectation is over all $\mathbf{a}$ consisting of $i-1$ blocks of n-bit strings. Since this statement is true for all b, we can average over all $2^{(i-1)\cdot n}$ possible values of b in (13) and combine the resulting equation with (12) to obtain

$$\mathrm{icost}_A(\mathcal{P}) \;=\; \sum_{i=1}^{m} \mathrm{E}_{\mathbf{a},\mathbf{b}}\left[\mathrm{icost}_A\big(\mathcal{P}^{(i)}_{\mathbf{a},\mathbf{b}}\big)\right].$$

A similar equation can be derived for Bob's part of the protocols, whence

$$\begin{aligned}\mathrm{icost}(\mathcal{P}) \;&=\; \sum_{i=1}^{m} \mathrm{E}_{\mathbf{a},\mathbf{b}}\left[\mathrm{icost}\big(\mathcal{P}^{(i)}_{\mathbf{a},\mathbf{b}}\big)\right] \\ &\geq\; \sum_{i=1}^{m} IC(f)\end{aligned}$$

Thus, $IC(f^m) \geq m \cdot IC(f)$. □

Proof of Theorem 1.2: Recall that EQ_n is robust and that $C(\mathsf{EQ}_n) = \Omega(\sqrt{n})$. By the Main Lemma, $IC(\mathsf{EQ}_n) = \Omega(\sqrt{n}) - O(\log n) = \Omega(\sqrt{n})$. By the Additivity Theorem, we now have $IC(\mathsf{EQ}^m_n) = \Omega(m\sqrt{n})$. Applying Proposition 2.1 gives $C(\mathsf{EQ}^m_n) = \Omega(m\sqrt{n})$ as desired. □

4.2 Boolean compositions

To prove Theorem 1.3 we simply combine the Lemma below with Proposition 2.1. We give the proof for OREQ^m_n only: proofs for XOREQ^m_n, MAJEQ^m_n and $\mathsf{THREQ}^{m,k}_n$ are similar.

Lemma 4.1 *For any constant $0 < \lambda < \frac{1}{2}$ and integers m, n with $m \leq \lambda 2^n$ we have*

$$IC(\mathsf{OREQ}^m_n) \;\geq\; \Omega(m\sqrt{n}).$$

Proof: We start with a protocol in $\mathcal{P}_0 \in \Pi_{1/3}(\mathsf{OREQ}^m_n)$ and "repeat" it $O(1)$ times to obtain a protocol $\mathcal{P} \in \Pi_\varepsilon(\mathsf{OREQ}^m_n)$ for some constant $\varepsilon < \frac{1}{2} - \lambda$. For $1 \leq i \leq m$ and an arbitrary string $a = w_1 w_2 \ldots w_{i-1} \in \{0,1\}^{(i-1)\cdot n}$, define the protocol $\mathcal{Q}^{(i)}_a$ for EQ_n as follows: on receiving inputs x and y, Alice and Bob simulate $\mathcal{P}$ on the random inputs $w_1 \ldots w_{i-1} x\, \mathbf{x}_{i+1} \ldots \mathbf{x}_m$ and $\overline{w}_1 \ldots \overline{w}_{i-1} y\, \mathbf{y}_{i+1} \ldots \mathbf{y}_m$ respectively, and the referee uses the same matrix as before. Here $\overline{w}_i$ denotes the bitwise complement of w_i. Since

$$\Pr_{\mathbf{u},\mathbf{v} \in \{0,1\}^n}[\mathsf{EQ}_n(\mathbf{u},\mathbf{v}) = 1] \;=\; 2^{-n},$$

the referee fails to output $\mathsf{EQ}_n(x,y)$ correctly with probability at most $\varepsilon + (m-i)\cdot 2^{-n} \leq \varepsilon + m/2^n \leq \varepsilon + \lambda$, which is a constant less than $\frac{1}{2}$. Therefore we can get a protocol $\mathcal{P}^{(i)}_a \in \Pi_{1/3}(\mathsf{EQ}_n)$ by "repeating" $\mathcal{Q}^{(i)}_a$ an appropriate constant number of times, say c times.

Arguing exactly as in the proof of Theorem 1.5, we have

$$\begin{aligned}\mathrm{icost}(\mathcal{P}) \;&=\; \sum_{i=1}^{m} \mathrm{E}_{\mathbf{a}}\left[\mathrm{icost}\big(\mathcal{Q}^{(i)}_{\mathbf{a}}\big)\right] \\ &=\; \sum_{i=1}^{m} \mathrm{E}_{\mathbf{a}}\left[\frac{1}{c}\cdot \mathrm{icost}\big(\mathcal{P}^{(i)}_{\mathbf{a}}\big)\right] \\ &\geq\; \frac{1}{c}\sum_{i=1}^{m} IC(\mathsf{EQ}_n) \\ &=\; \Omega(m\sqrt{n})\end{aligned}$$

and so $\mathrm{icost}(\mathcal{P}_0) = \Omega(m\sqrt{n})$. The result follows. □

5 Acknowledgment

We wish to thank Sanjeev Arora for his insightful comments on an earlier draft of the paper.

References

[A96] Ambainis, A. Communication complexity in a 3-computer model, *Algorithmica*, **16** (1996), 298–301.

[BK97] Babai, L., Kimmel, P.G. Randomized simultaneous messages, *Proc. 12th IEEE Symp. on Computational Complexity* (1997), 239–246.

[B01] Babai, L. *Communication complexity*, Tech. Report TR-2001-09, University of Chicago (2001).

[EIRS91] Edmonds, J., Impagliazzo, R., Rudich, S., Sgall, J. Communication complexity towards lower bounds on circuit depth, *Proc. 32nd IEEE FOCS* (1991), 249–257.

[FKNN91] Feder, T., Kushilevitz, E., Naor, M., Nisan, N. Amortized communication complexity, *SIAM J. Comput.*, **24** (1995), 736–750. (Preliminary version in *Proc. 32nd IEEE FOCS* (1991), 239–248.)

[KRW91] Karchmer, M., Raz, R., Wigderson, A. On proving super-logarithmic depth lower bounds via the direct sum in communication complexity, *Structures in Complexity Theory '91* (1991), 299–304.

[KKN92] Karchmer, M., Kushilevitz, E., Nisan, N. Fractional covers and communication complexity, *Structures in Complexity Theory '92* (1992), 262–274.

[KW90] Karchmer, M., Wigderson, A. Monotone circuits for connectivity require super-logarithmic depth, *SIAM J. Discrete Math.*, **3** (1990), 255–265.

[KN97] Kushilevitz, E. and Nisan, N. *Communication Complexity*, Cambridge University Press, 1997.

[NS96] Newman, I., Szegedy, M. Public vs. private coin flips in one round communication games, *Proc. 28th ACM STOC* (1996), 561–570.

[Y79] Yao, A.C. Some complexity questions related to distributive computing, *Proc. 11th ACM STOC* (1979), 209–213.