Direct Sum Questions in Classical Communication Complexity

Denis Pankratov

February 20, 2012

Contents

1 Introduction 3

2 Acknowledgements 5

3 Direct-Sum Results in Other Contexts 6
  3.1 Matrix Multiplication .............................. 6
  3.2 Polynomial Evaluation ............................. 6
  3.3 XOR lemmas ...................................... 7
    3.3.1 Yao’s XOR lemma ............................. 7
    3.3.2 Viola-Wigderson XOR lemmas ................. 8
  3.4 Raz’s Parallel Repetition Theorem .................. 8
  3.5 Decision Trees .................................. 9
  3.6 Magnification Ratio ................................ 11

4 Communication Complexity Background 12

5 The Roots of Direct Sum in CC 16
6 The Birth of Direct Sum in CC

7 Direct Sum Violations 22

7.1 Deterministic Communication Complexity 22

7.2 Randomized Communication Complexity 24

7.3 Nondeterministic Communication Complexity 26

8 Direct Sum Approaches for General Functions and Relations 27

8.1 Nondeterministic Communication Complexity 27

8.2 Deterministic Communication Complexity 30

8.3 Randomized Communication Complexity 32

8.3.1 Information Theory Background 33

8.3.2 Information Cost Measure 34

8.3.3 Public vs Private Randomness in the Information Cost: a New Result 38

A Upper Bound on the Monotone Universal Composition Relation 44
Abstract

In 1988, Karchmer and Wigderson generalized Yao’s two-party communication model of functions to relations and showed a remarkable connection of this model to the Boolean circuit model. A few years later, continuing this line of work, Karchmer, Raz, and Wigderson proposed a program to separate $NC$ from $P$ through direct-sum-type inequalities in communication complexity. This spurred the study of this fundamental question in communication complexity: given problems $A$ and $B$, is it easier to solve $A$ and $B$ together than separately? It seems that we are still far from separating $NC$ from $P$; however, during the last 20 years of research our knowledge of the behavior of different communication complexity measures with respect to the direct sum has seen a lot of progress. We survey some of these results in this paper and make a new observation about the recent approach to the direct-sum question in the randomized setting.

1 Introduction

Consider a computational task $T$ and let $T^n$ denote the juxtaposition of $n$ independent copies of $T$. Let $\mathcal{M}$ be a complexity measure on the set of tasks. The following is one of the fundamental questions about $\mathcal{M}$.

**Question 1.1** (Direct Sum). Is it true that for all $T$ we have $\mathcal{M}(T^n) = \Theta(n\mathcal{M}(T))$?

Economists might immediately recognize this question as whether $\mathcal{M}$ admits the “economies of scale”. This question appears all over computational complexity theory and other areas of mathematics (see Section 3), and it is known under various names, such as “direct-sum theorems”, “direct-product theorems” (if $\mathcal{M}$ is in some sense “multiplicative”), and “XOR lemmas”. In
almost all cases $M(T^n) = O(nM(T))$ is trivially true, and the real question is if $M(T^n) = \Omega(nM(T))$ holds. Quite often the gut reaction is “yes, of course;” however, over the years of research these questions have proved to be rather elusive. The answers vary from “strong no/strong yes” to being wide open depending on the definitions of $M$ and $T$.

The main subject of this paper is the direct-sum question in the classical two-party communication complexity (CC). This question has been studied for over 20 years. The original motivation came from the problem of separating complexity classes $NC^1$ and $NC^2$, which turned out to have an interesting connection with communication complexity [KW88] and, in particular, the direct sum in CC [KRW91]. The direct-sum question is very sensitive to the model of communication (deterministic, nondeterministic, randomized) and to the task at hand (relation, function, partial function). Nondeterministic communication complexity is the most understood model in this regard. Two works [FKNN95] and [KKN95] showed that solving $k$ copies of a relation $R$ takes essentially $k$ times the amount of the nondeterministic communication, i. e., $C_N(R^k) = \Omega(k(C_N(R) - \log n))$, where $n$ is the number of bits required to describe an input for $R$ and $C_N(R)$ denotes the nondeterministic communication complexity of $R$. This immediately implies a weak direct-sum result for the deterministic communication complexity of functions, because the separation between the nondeterministic communication complexity and the deterministic communication complexity of functions can be at most quadratic ([AUY83] and [HR88]). The randomized communication complexity saw no progress until information-theoretic techniques were introduced in 2001 [CSWY01]. A new notion of complexity called information cost was defined in [Bra11]. As of now, the most promising approach to the direct sum in the randomized setting is to show that the information cost cannot be much lower than the communication
complexity. An interesting feature of the information cost is that it is defined in terms of protocols that use both public and private randomness. In this work we show that the question whether the information cost can be achieved with protocols that use only public randomness is essentially equivalent to the direct-sum question.

The rest of the paper is organized as follows. In Section 3 we provide a quick summary of the most famous instances of the direct-sum questions in different contexts other than communication complexity. In Section 4 we give an overview of communication complexity. In Section 5 we describe Karchmer-Wigderson games [KW88] that revealed a surprising relationship between communication complexity and the Boolean-circuit depth and motivated the study of the direct-sum question. Section 6 describes the connection between Karchmer-Wigderson games and the direct-sum problem, i.e., the Karchmer-Raz-Wigderson program to separate $NC^1$ from $NC^2$. In Section 7 we exhibit the known examples when the direct-sum statements, at least in their strongest forms, fail to hold. Section 8 surveys the past and modern approaches to the direct-sum questions. In Section 8.3.3 we present our new result.

2 Acknowledgements

I would like to thank Laci Babai for reviewing earlier versions of this paper and for his helpful comments that considerably improved the presentation of the material. I would like to thank Sasha Razborov for introducing me to the direct-sum problem in communication complexity and getting me in touch with a leading expert in the area. I thank Mark Braverman for the insightful conversations, continued encouragement and support in working on the information cost. I also thank Mark for pointing out Proposition 8.20.
3 Direct-Sum Results in Other Contexts

3.1 Matrix Multiplication

The computational task \(T\) is to multiply a fixed \(n \times n\) matrix \(M\) with entries from \(F_2\) (the field of two elements) by a vector \(v\). The complexity measure \(\mathcal{M}\) is the size of the smallest Boolean circuit computing \(Mv\) correctly for all \(v\). A simple counting argument shows that there is a matrix \(M\) requiring a circuit of size \(\Omega(n^2/\log n)\), hence \(\mathcal{M}(T) = \Omega(n^2/\log n)\). If we now consider multiplying \(M\) by \(n\) vectors \(v_1, \ldots, v_n\), then this task is equivalent to multiplying two \(n \times n\) matrices and can be done with a circuit of size \(O(n^\omega)\), where \(\omega\) is the matrix multiplication constant. The history of \(\omega\) is fascinating on its own, but for our purposes it suffices to name the best known bound for \(\omega < 2.373\) [VW11]. Therefore, the direct sum is strongly violated in this case, as we have

\[
\mathcal{M}(T^n) = O(n^{2.373}) \ll n\mathcal{M}(T) = \Omega(n^3/\log n).
\]

3.2 Polynomial Evaluation

Let \(p(x) = \sum_{i=0}^n a_ix^i\) be a polynomial. Straightline program \(\alpha\) for computing \(p\) is a sequence of instructions \(A \odot B\), where \(\odot \in \{\times, \div, +, -\}\) and \(A, B\) can be scalars, intermediate results, or input \(x\). The cost of \(\alpha\) is the number of multiplication and division instructions that do not solely depend on coefficients \(a_0, \ldots, a_n\). This allows preprocessing of coefficients of \(p\) for free. The following classical results demonstrate that the direct sum is violated for polynomial evaluation in a very strong sense.

**Theorem 3.1** (Motzkin [Mot55], Winograd [Win70]). Let \(p(x) = \sum_{i=0}^n a_ix^i\) be a polynomial over \(\mathbb{C}\). If \(a_0, \ldots, a_n\) are algebraically independent over \(\mathbb{Q}\), then any straightline program computing \(p\) has cost at least \(n/2\).
Theorem 3.2 (Fiduccia [Fid72]). Let \( p(x) = \sum_{i=0}^{n} a_i x^i \) be a polynomial over \( \mathbb{C} \) and \( \varepsilon_1, \ldots, \varepsilon_n \in \mathbb{C} \) then \( p(\varepsilon_1), \ldots, p(\varepsilon_n) \) can be computed by a straightline program of cost \( O(n \log n) \).

Let \( T \) be the task of evaluating a given polynomial \( p(x) = \sum_{i=0}^{n} a_i x^i \) with \( a_0, \ldots, a_n \) algebraically independent over \( \mathbb{Q} \). Let \( \mathcal{M} \) denote the cost of a best straightline program for \( p \). Then we have

\[
\mathcal{M}(T^n) = O(n \log n) \ll n\mathcal{M}(T) = \Omega(n^2).
\]

3.3 XOR lemmas

Definition 3.3. For a pair of Boolean functions \( f, g : \mathcal{X} \to \{\pm 1\} \) the correlation between \( f \) and \( g \) with respect to distribution \( \mu \) on \( \mathcal{X} \) is defined as

\[
cor_{\mu}(f, g) = |\mathbb{E}_{x \sim \mu}(f(x)g(x))|.
\]

For a class of Boolean functions \( C \) the correlation between \( f \) and \( C \) is the maximum of \( \text{cor}_{\mu}(f, g) \) over all \( g \in C \). When \( \mu \) is the uniform distribution, we shall simply write \( \text{cor}(f, g) \).

3.3.1 Yao’s XOR lemma

We use \( f^{\oplus n} \) to denote the XOR of \( n \) copies of \( f \). Let \( C(k, s) \) denote the class of Boolean functions on \( k \)-bit inputs computable by Boolean circuits of size \( s \). Then Yao’s XOR lemma [Yao82] (see also [GNW95]) says the following.

Theorem 3.4 (Yao [Yao82]). For any \( k, s, n \in \mathbb{N}, \epsilon, \alpha > 0, f : \{\pm 1\}^k \to \{\pm 1\} \) we have

\[
\text{cor}(f, C(k, s)) \leq \epsilon \Rightarrow \text{cor}(f^{\oplus n}, C(nk, s \left( \frac{\alpha}{nk} \right)^2)) \leq \epsilon^n + \alpha.
\]
We can restate the above theorem in the direct-sum-like terms given in the introduction. Let $T$ be the task of computing $f$. Let $T^n$ denote the task of computing $f^\oplus n$. Define $\mathcal{M}(g) := -\log \text{cor}(g, C(k, s))$ and $\mathcal{M}'(h) := -\log \text{cor}(h, C(nk, s(\epsilon^{\Omega(n)/nk})^2))$. Then Yao’s lemma says

$$\mathcal{M}'(f^\oplus n) = \Omega(n \mathcal{M}(f)).$$

Note that two different complexity measures appear on the two sides of the inequality, making it hard to judge the strength of this direct-sum result. In fact, the class of circuits computing $f^\oplus n$ is smaller than the class of circuits computing $f$ in Yao’s XOR lemma.

### 3.3.2 Viola-Wigderson XOR lemmas

Viola and Wigderson [VW08] showed XOR lemmas for two models: polynomials over $\mathbb{F}_2$ and multiparty communication complexity. More specifically, they showed that if a Boolean function has correlation $\epsilon < 1/2$ with $k$-bit $k$-party protocols, then the correlation of XOR of $m$ copies of the same Boolean function with $c$-bit $k$-party protocols drops to $2^c \epsilon^m/2^k$. For polynomials over $\mathbb{F}_2$, Viola and Wigderson showed that if a Boolean function has correlation $\leq 1 - 1/2^d$ with degree-$d$ polynomials, then XOR of $m$ copies of this Boolean function has correlation at most $\exp(-\Omega(m/(4^d d)))$.

### 3.4 Raz’s Parallel Repetition Theorem

Consider the game where the referee chooses a pair $(x, y)$ according to a publicly known distribution, sends $x$ to Alice, $y$ to Bob, who respond with $a$ and $b$, respectively. Alice and Bob win the game if a publicly known predicate $Q(x, y, a, b)$ holds. This game originated from the study of multi-prover interactive proof systems [FRS88].
More specifically, the two-player referee game \( G \) is a pair \((\mu, Q)\), where \( \mu \) is a probability distribution on \( X \times Y \) and \( Q \) is a predicate on \( X \times Y \times A \times B \). The strategy for this game is a pair \((h_a, h_b)\), where \( h_a : X \to A \), \( h_b : Y \to B \). The winning probability of the strategy is \( P_{(x, y) \sim \mu} [Q(x, y, h_a(x), h_b(y)) = 1] \). The value of the game \( \nu(G) \) is the maximum over all strategies of the winning probability of a strategy.

The parallel repetition of \( G \) is another game \( G^n \) defined as \((\mu^n, Q^\wedge n)\), where \( \mu^n \) is the \( n \)-fold product distribution on \( X^n \times Y^n \) and \( Q^\wedge n \) is a predicate on \( X^n \times Y^n \times A^n \times B^n \) defined by \( Q^\wedge n(x_1, \ldots, x_n, y_1, \ldots, y_n, a_1, \ldots, a_n, b_1, \ldots, b_n) = \bigwedge_i Q(x_i, y_i, a_i, b_i) \). A natural direct sum question is how the value of the repeated game is related to the value of the single instance of the game. The following celebrated result due to Raz [Raz98] provides the answer.

**Theorem 3.5** (Raz [Raz98]). For any game \( G \) with value \( \nu(G) < 1 \) there exists \( \bar{\nu} < 1 \) (depending only on \( \nu \)) such that

\[
\nu(G^n) = \bar{\nu}^n / \log(|A||B|).
\]

### 3.5 Decision Trees

Let \( f \subseteq \{0,1\}^n \times Y \) be the following promise problem: given \( x \in \{0,1\}^n \) there is at least one \( y \in Y \) such that \((x, y) \in f\), and we are asked to find one such \( y \). A decision tree \( T \) is a full binary tree, in which internal nodes are labeled with the \( i \), where\(^1 \) \( i \in [n] \). The edge going to the left child is labeled 0 and the edge going to the right child is labeled 1. The leaves are labeled by elements of \( Y \). We define the output function \( \text{out}_T(v, x) \) on input \( x \in \{0,1\}^n \) and node \( v \) recursively. If \( v \) is a leaf \( \text{out}_T(v, x) \) is the label of the leaf. If \( v \) is the internal

\(^1\)[n] = \{1, 2, \ldots, n\}.\n
9
node with label \(i\) and left child \(u\) and right child \(w\), then

\[
\text{out}_T(v, x) = \begin{cases} 
\text{out}_T(u, x) & \text{if } x_i = 0, \\
\text{out}_T(w, x) & \text{if } x_i = 1.
\end{cases}
\]

The output of the decision tree on \(x\) is \(\text{out}_T(x) := \text{out}_T(r, x)\), where \(r\) is the root of \(T\). The decision tree \(T\) is said to compute \(f\) if for all \(x \in \{0, 1\}^n\) we have \((x, \text{out}_T(x)) \in f\). The deterministic decision tree complexity of \(f\), denoted by \(\text{DTC}(f)\), is the depth of the minimum-depth decision tree computing \(f\).

The \(k\)-fold product \(f^k\) is defined naturally as a subset of \((\{0, 1\}^n)^k \times Y^k\), where \((x^1, \ldots, x^k, y^1, \ldots, y^k) \in f^k\) if and only if \((x^i, y^i) \in f\) for all \(i \in [k]\).

Jain, Klauck, and Santha [JKS10] proved the optimal direct-sum result for the deterministic decision tree complexity of promise problems \(f\).

**Theorem 3.6** (Jain, Klauck, and Santha [JKS10]). For every promise problem \(f \subseteq \{0, 1\}^n \times Y\) and for every \(k\) we have \(\text{DTC}(f^k) = k \cdot \text{DTC}(f)\).

The randomized decision tree \(T\) is a probability distribution on the deterministic decision trees. The depth of a randomized decision tree \(T\) is the maximum depth of a deterministic decision tree in the support of \(T\). The \(\epsilon\)-error randomized query complexity of \(f\), denoted by \(\text{RQC}_\epsilon(f)\), is the minimum depth of a randomized decision tree \(T\) computing \(f\) with probability of error at most \(\epsilon\) on every input. The following direct-sum result is due to Jain, Klauck, and Santha [JKS10].

**Theorem 3.7** (Jain, Klauck, Santha [JKS10]). Let \(f \subseteq \{0, 1\}^n \times Y\) be a promise problem, \(k \in \mathbb{N}, \delta > 0\). Then \(\text{RQC}_\epsilon(f^k) \geq \delta^2 k \cdot \text{RQC}_\epsilon(f),\) where \(\epsilon' = \epsilon/(1-\delta)+\delta\).
3.6 Magnification Ratio

In additive combinatorics, the original proof of the Plünnecke-Ruzsa inequality relied heavily on a direct sum result. The Plünnecke graph of level \(h\) is a directed graph \(G = (V,E)\), where\(^2\) \(V = \bigsqcup_{i=0}^{h} V_i\) and \(E \subseteq \bigcup_{i=1}^{h} V_{i-1} \times V_i\), satisfying the following two conditions:

- if \((u,v) \in E\) and \((v,w_1), \ldots, (v,w_k) \in E\) then there exists \(v_1, \ldots, v_k\) such that for all \(i\) we have \((u,v_i) \in E\) and \((v_i, w_i) \in E\),
- if \((u_1,v), \ldots, (u_k,v) \in E\) and \((v,w) \in E\) then there exists \(v_1, \ldots, v_k\) such that for all \(i\) we have \((u_i, v_i) \in E\) and \((v_i, w) \in E\).

For \(X \subseteq V_0\) we define the set of neighbors of \(X\) at level \(h\) by \(N_h(X) = \{v \in V_h \mid\) there exists a path from some \(x \in X\) to \(v\}\}. The magnification ratio of a Plünnecke graph \(G\) is defined as \(||G|| = \min_{\emptyset \neq X \subseteq V_0} \{|N_h(X)|/|X|\}\).

Given two Plünnecke graphs \(G_1 = (\bigsqcup_i V_{i,1}, E_1)\) and \(G_2 = (\bigsqcup_i V_{i,2}, E_2)\) of level \(h\), the product graph \(G = G_1 \times G_2 = (V,E)\) is defined as follows \(V = \bigsqcup_i V_{i,1} \times V_{i,2}\) and \(((v,w), (x,y)) \in E\) if and only if \((v,x) \in E_1\) and \((w,y) \in E_2\).

The direct-sum theorem for the magnification ratio due to Plünnecke and Ruzsa (see [Nat96]) is stated as follows.

**Theorem 3.8** (Plünnecke, Ruzsa). Let \(G_1\) and \(G_2\) be two Plünnecke graphs. Then we have

\[ ||G_1 \times G_2|| = ||G_1|| \cdot ||G_2||. \]

To obtain the formulation of this direct-sum result in the terms from the introduction, let \(T\) stand for a Plünnecke graph, \(T^n\) the \(n\)-fold product of such a graph with itself, and \(\mathfrak{M}(T)\) the logarithm of the magnification ratio of \(T\).

\(^2\)The symbol \(\bigsqcup\) denotes disjoint union.
4 Communication Complexity Background

In 1979, Yao [Yao79] introduced the two-party communication model for computing functions. In 1988, Karchmer and Wigderson [KW88] generalized this model to handle relations. In the generalized version, two parties, traditionally called Alice and Bob, are trying to collaboratively “compute” a known relation $R \subseteq X \times Y \times Z$ in the following sense. Each party is computationally unbounded; however, Alice is only given input $x \in X$ and Bob is only given $y \in Y$, and their job is to output $z \in Z$ such that $(x, y, z) \in R$. To that end, Alice and Bob communicate in accordance with an agreed-upon communication protocol $\pi$. Protocol $\pi$ specifies as a function of transmitted bits only whether the communication is over and, if not, who sends the next bit. Moreover, $\pi$ specifies as a function of the transmitted bits and $x$ the value of the next bit to be sent by Alice. Similarly for Bob. The communication is over as soon as both players know $z$ such that $(x, y, z) \in R$. The cost of the protocol $\pi$, is the number of bits exchanged on the worst input. The deterministic communication complexity of $R$, denoted by $C(R)$, is the least cost of a protocol computing $R$.

Several remarks about the above definitions are in order:

- An input pair $x, y$ such that there is no $z$ with $(x, y, z) \in R$ is called illegal. A relation $R$ with illegal inputs is called partial. For partial problems $R$ we assume that Alice and Bob are never presented with illegal inputs. Alternatively, we can think of illegal inputs as being in relation with any $z \in Z$.

- If for each pair $x, y$ there is at most one $z$ with $(x, y, z) \in R$ then $R$ is, in fact, a partial function. Thus Yao’s original model is the special case of this model.

- Another possibility for a termination condition of the protocol is to finish
communication as soon as one of the parties knows the answer\textsuperscript{3}. The difference in the complexity measures between the two conventions is\textsuperscript{4} at most $\log |Z|$.  

- In this paper we shall only consider $|\mathcal{X}|, |\mathcal{Y}|, |Z| < \infty$.  

With a relation $R \subseteq \mathcal{X} \times \mathcal{Y} \times Z$ we associate its communication matrix $M_R$ of size $|\mathcal{X}| \times |\mathcal{Y}|$, where the entry $M_R(x, y)$ is a set of all $z \in Z$ such that $(x, y, z) \in R$. A combinatorial rectangle is a subset of $\mathcal{X} \times \mathcal{Y}$ of the form $A \times B$ for some $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$. A combinatorial rectangle $A \times B$ is called monochromatic if there is a $z \in Z$ such that for all $x \in A$ and $y \in B$ we have $z \in M_R(x, y)$ (to emphasize particular $z$ we also call such rectangle $z$-monochromatic). Every protocol partitions the communication matrix into a set of nonoverlapping monochromatic combinatorial rectangles.

Let $\text{NCov}(M_R)$ denote the minimum number of monochromatic rectangles needed to cover $M_R$ allowing overlaps. The nondeterministic communication complexity of $R$ is defined as $C_N(R) = \log \text{NCov}(M_R)$. Similarly, let $\text{NCov}_z(M_R)$ denote the minimum number of $z$-monochromatic rectangles needed to cover the entries of $M_R$ containing $z$. The corresponding “NP-like” version of nondeterministic communication complexity is defined as $C_{N,z}(R) = \log \text{NCov}_z(M_R)$.

The following observation relating nondeterministic communication complexity and deterministic communication complexity of functions is due to Halstenberg and Reischuk [HR88]. A weaker version of this theorem was originally proven by Aho, Ullman, and Yannakakis [AUY83].

\textsuperscript{3}This was the original definition due to Yao [Yao79]. Lovász [Lov90] argued that this is, indeed, a natural termination condition. For example, with this condition the trivial protocol for a Boolean function $F : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ requires $n$ bits of communication rather than $n + 1$, when the last party has to broadcast the message. We decided to use the other termination rule to be consistent with the majority of literature on the direct sum problem.

\textsuperscript{4}All logarithms in this paper are to the base 2 unless otherwise stated.
Theorem 4.1 (Halstenberg, Reischuk [HR88]). For every function \( f : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\} \) we have

\[
C(f) = O(C_{N,0}(f)C_{N,1}(f)) = O(C_N(f)^2).
\]

The deterministic communication complexity model can be extended by granting Alice and Bob access to random strings. This leads to two natural types of protocols: private-coin, where Alice and Bob each have their own random string concealed from another player, and public-coin, where Alice and Bob have access to a shared random string. The \( \epsilon \)-error public-coin (private-coin) protocol is a randomized protocol that outputs the correct value with probability at least \( 1 - \epsilon \) on every input. The worst-case cost of the best \( \epsilon \)-error public-coin (private-coin) protocol is denoted by \( C_{R,\epsilon} \) (\( C_{R,\epsilon}^p \), respectively). Here and in what follows, when \( \epsilon \) is omitted it is assumed to be \( \epsilon = 1/3 \).

Clearly, \( C_{R,\epsilon} \leq C_{R,\epsilon}^p \) always holds. Ilan Newman [New91] showed that for partial functions the two measures are identical up to constant multiplicative factors and logarithmic additive terms.

**Theorem 4.2** (Newman [New91]). For every partial function \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \) we have

\[
C_{R}^p(f) = O(C_R(f) + \log n).
\]

Instead of introducing randomness to the players, we can introduce randomness to the inputs. Let \( \mu \) be a probability distribution over \( \mathcal{X} \times \mathcal{Y} \). We consider deterministic protocols, which err on at most \( \epsilon \)-fraction of inputs weighed according to \( \mu \). The worst-case cost of the best such protocol is called \( \epsilon \)-error distributional communication complexity and is denoted by \( C_{D,\epsilon}^\mu \).

Distributional and public-coin complexities are related via Yao’s Min-Max
Principle [Yao77].

**Proposition 4.3** (Yao [Yao77]).

\[ C_{R,\epsilon}(f) = \max_{\mu} C_{D,\epsilon}^{\mu}(f). \]

The *direct product* of two relations \( R \subseteq X \times Y \times Z \) and \( R' \subseteq X' \times Y' \times Z' \) is a relation \( R \times R' \subseteq (X \times X') \times (Y \times Y') \times (Z \times Z') \) such that \((x_1, x_2), (y_1, y_2), (z_1, z_2)) \in R \times R'\) if and only if \((x_1, y_1, z_1) \in R\) and \((x_2, y_2, z_2) \in R'\). The \( k \)-fold product of a relation \( R \) is defined recursively as \( R^k = R \times R^{k-1} \).

The *amortized* versions of the complexity measures are denoted by the tilde above the notational symbol for the given complexity measure. For example, the *amortized nondeterministic communication complexity* of \( R \) is \( \tilde{C}_N(R) = \limsup_{k \to \infty} C_N(R^k)/k \). The definitions of amortized versions of other complexity measures are analogous. We use \( C_{R,\epsilon}^{\mu}(f^n) \) to denote public-coin randomized communication complexity of \( f \) where a protocol is allowed to err with probability at most \( \epsilon \) in each coordinate on every input.

For any protocol \( \pi \) we use \( C(\pi) \) to denote the maximum number of bits exchanged on an input.

The *disjointness* function, denoted by \( \text{DISJ} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\} \), was introduced by Babai, Frankl, and Simon [BFS86] as an example of a coNP-complete problem in the communication complexity world. Since then it has become one of the central objects of study in CC. Formally, it is defined as

\[
\text{DISJ}(x, y) = \begin{cases} 
0 & \text{if there is an index } i \text{ such that } x_i = y_i = 1, \\
1 & \text{if for all } i \text{ we have } x_i = 0 \text{ or } y_i = 0.
\end{cases}
\]

In the promise version of the disjointness function, denoted by \( \text{UDISJ} \), Alice and Bob are promised that if there is an index \( i \) with \( x_i = y_i = 1 \) then such
an index is unique. The partial function $UDISJ$ is an easier problem than $DISJ$, and yet it has the $\Omega(n)$ lower bound on its bounded-error randomized communication complexity. This result is due to Bala Kalyanasundaram and Georg Schnitger \cite{KS92} (see also \cite{Raz92} and \cite[Chapter 4.6]{KN97}).

**Theorem 4.4** (Kalyanasundaram and Schnitger \cite{KS92}).

$$C_R(UDISJ) = \Omega(n).$$

The first difference function $FDIFF : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1,\ldots,n\}$ is defined as follows.

$$FDIFF(x, y) = \begin{cases} \min\{i \mid x_i \neq y_i\} & \text{if } x \neq y, \\ 0 & \text{otherwise}. \end{cases}$$

The following proposition is implicit in the work of Feige, Raghavan, Peleg, and Upfal \cite{FRPU94}.

**Proposition 4.5** (Feige et al. \cite{FRPU94} (implicit)).

$$C_{R,\epsilon}(FDIFF) = O\left(\log\left(\frac{n}{\epsilon}\right)\right).$$

For a thorough treatment of communication complexity we refer an interested reader to an excellent monograph by Kushilevitz and Nisan \cite{KN97} and a more recent survey by Lee and Shraibman \cite{LS09}.

## 5 The Roots of Direct Sum in CC

The study of direct-sum questions in communication complexity originated from the work of Karchmer and Wigderson \cite{KW88}. This work was described in the doctoral thesis of Karchmer \cite{Kar89}. For these findings Karchmer became the
first foreign winner of the ACM Distinguished Doctoral Dissertation Award.

Karchmer and Wigderson were interested in proving lower bounds on circuit depth. Consider Boolean circuits over the basis \(\{\land, \lor, \neg\}\), where \(\land\)- and \(\lor\)-gates have fan-in two (unless otherwise stated) and \(\neg\)-gates have fan-in one. The depth of a circuit is the number of fan-in two gates on the longest path from the root to a leaf. Note that \(\neg\)-gates do not count towards the depth. The depth of the shallowest circuit computing a given function \(f\) is denoted by \(d(f)\).

For a Boolean function \(f : \{0, 1\}^n \to \{0, 1\}\), Karchmer and Wigderson [KW88] defined a relation \(R_{KW}(f) \subseteq f^{-1}(1) \times f^{-1}(0) \times [n]\) by \((x, y, i) \in R\) if and only if \(x_i \neq y_i\). Similarly, for two disjoint subsets \(A, B \subseteq \{0, 1\}^n\) relation \(R(A, B) \subseteq A \times B \times [n]\) is defined to consist of all triples \((x, y, i)\) with \(x_i \neq y_i\).

Karchmer and Wigderson [KW88] showed the following remarkable equivalence.

**Theorem 5.1** (Karchmer, Wigderson [KW88]). For every Boolean function \(f : \{0, 1\}^n \to \{0, 1\}\) we have

\[
d(f) = C(R_{KW}(f)).
\]

We remark that from the point of view of Karchmer-Wigderson games, the deterministic model is the most interesting due to the following.

**Proposition 5.2.** For every Boolean function \(f : \{0, 1\}^n \to \{0, 1\}\), we have

1. \(C_N(R_{KW}(f)) = O(\log n)\),
2. \(C_R(R_{KW}(f)) = O(\log n)\),
3. \(C^p_R(R_{KW}(f)) = O(\log n)\).

**Proof.** The first part is clear as Alice and Bob can simply guess an index \(i\) such that \(x_i \neq y_i\). The second part follows immediately from Proposition 4.5, and the third part is obtained from the second by applying Proposition 4.2. \(\square\)
In the monotone world, a result similar to Theorem 5.1 holds. Consider a monotone function \( f : \{0, 1\}^n \rightarrow \{0, 1\} \). Let \( d_m(f) \) denote the minimum depth of a monotone circuit (no \( \neg \)-gates) computing \( f \). A minterm \( S \subseteq [n] \) of \( f \) is a minimal subset of indices such that \((\forall x \in \{0, 1\}^n)(x|_S = 1 \Rightarrow f(x) = 1)\). A maxterm is similar except with 0 replacing 1 everywhere in the above definition. Let \( MIN(f) \) be the set of minterms of \( f \) and \( MAX(f) \) be the set of maxterms of \( f \). Note that a minterm always intersects a maxterm. Karchmer and Wigderson defined relation \( R_{KW}^m(f) \subseteq MIN(f) \times MAX(f) \times [n] \) by \((A, B, i) \in R_{KW}^m(f)\) if and only if \( i \in A \cap B \) and showed the following.

**Theorem 5.3** (Karchmer, Wigderson [KW88]). For every monotone Boolean function \( f : \{0, 1\}^n \rightarrow \{0, 1\} \) we have

\[
d_m(f) = C(R_{KW}^m(f)).
\]

Thus proving lower bounds on the depth of monotone circuits amounts to proving lower bounds on communication complexity of relations. Using this characterization, Karchmer and Wigderson [KW88] (see also Grigni and Sipser [GS95]) showed the separation \( mNC^1 \subset mAC^1 \). Using a similar argument, but a different communication complexity problem, Raz and McKenzie [RM97] proved the separation of \( mNC^i \) from \( mNC^{i+1} \) for all \( i \geq 1 \). Unfortunately, the techniques used to prove these results do not carry over to the non-monotone world, and separating \( NC^1 \) from \( NC^2 \) remains an open problem. However, this question can still be approached within the communication complexity framework. This idea is now known as Karchmer-Raz-Wigderson program and a certain direct-sum question lies at its core.
6 The Birth of Direct Sum in CC

For two Boolean functions \( f : \{0,1\}^n \rightarrow \{0,1\} \) and \( g : \{0,1\}^m \rightarrow \{0,1\} \) a
composed function \( f \circ g : \{0,1\}^{nm} \rightarrow \{0,1\} \) is defined by
\[
f \circ g(\vec{X}_1, \ldots, \vec{X}_n) = f(g(\vec{X}_1), \ldots, g(\vec{X}_n)),
\]
where \( \vec{X}_i \in \{0,1\}^m \). When a function is composed with itself \( k \) times we use
special\(^5\) notation \( f^{\circ k} = f \circ f^{\circ (k-1)} \).

Over two decades ago Karchmer, Raz, and Wigderson [KRW91] proposed
the following two problems, which remain wide open to this day.

**Open Problem 6.1** (Deterministic Direct Sum of Relations). What is the
relationship between \( C(R^k) \) and \( C(R) \)?

**Open Problem 6.2** (Deterministic Direct Sum for Karchmer-Wigderson Games
of Composed Functions). What is the relationship between \( C(R_{KW}(f^{\circ k})) \) and
\( C(R_{KW}(f)) \)?

Karchmer, Raz, and Wigderson [KRW91] showed that a “good answer” to
the second question would give a separation between \( NC^1 \) and \( NC^2 \). This
approach is now called Karchmer-Raz-Wigderson program.

**Theorem 6.3** (Karchmer, Raz, Wigderson [KRW91]). If for some \( \epsilon \in (0,1) \)
every \( f \) satisfies \( C(R_{KW}(f^{\circ 2})) \geq (1 + \epsilon)C(R_{KW}(f)) \) then \( NC^1 \neq NC^2 \).

**Proof.** Let \( k = \log n / \log \log n \) and take a hard function \( f \) on \( \log n \) variables,
i. e., \( d(f) = C(R_{KW}(f)) = \Omega(\log n) \). Then \( f^{\circ k} \) is a function on \( n \) variables.
This function is clearly in \( NC^2 \). Now, we have \( C(R_{KW}(f^{\circ k})) \geq (1 + \epsilon)C(R_{KW}(f^{\circ (k/2)})) \geq \ldots \geq (1 + \epsilon)^{\log k}C(R_{KW}(f)) = \Omega(\log^{1+\epsilon'} n/\log \log n) \) for
some \( \epsilon' > 0 \).

\(^5\)Not to be confused with \( f^k \), the juxtaposition of \( k \) copies of \( f \).
Karchmer et al. [KRW91] mention that the assumption of the above theorem can be tweaked in many ways without weakening the conclusion. For instance, the assumption can be changed as follows.

**Theorem 6.4** (Karchmer, Raz, Wigderson [KRW91]). *If for a random function $f$ and for every $g$ we have $C(R_{KW}(f \circ g)) \geq \epsilon \cdot C(R_{KW}(f)) + C(R_{KW}(g))$ then $NC^1 \neq NC^2$."

One of the very few steps made towards Open Problem 6.2 was a lower bound on the *universal composition relation*, denoted by $U_{n,k}$. Alice and Bob are each given a red/blue coloring of a complete $n$-ary tree of depth $k$. The players are promised that the roots are colored differently. Moreover, if node $u$ is colored in Alice’s tree differently from $u$ in Bob’s tree then there exists a child of $u$ that is also colored differently. The goal of Alice and Bob is to agree on a leaf that is colored differently in the two colorings.

In light of the following proposition, proving a lower bound on $C(U_{n,k})$ is an easier task than Open Problem 6.2 and was proposed by Karchmer, Raz, and Wigderson [KRW91] as a test of the feasibility of their program.

**Proposition 6.5** (Karchmer, Raz, Wigderson [KRW91]). *For every set of $k$ Boolean functions $f_1, \ldots, f_k : \{0,1\}^n \rightarrow \{0,1\}$ we have $C(R_{KW}(f_1 \circ \cdots \circ f_k)) \leq C(U_{n,k})$."

The lower bound for the universal composition relation was first proved by Edmonds et al. [ERIS91], who showed that $C(U_{n,k}) \geq kn - O(k^2 \sqrt{n \log n})$. A completely different approach was used by Håstad and Wigderson [HW97] who showed $C(U_{n,k}) \geq kn - O(k^3 \log k)$. Observe that both theorems leave open whether $C(U_{n,k}) = \Omega(nk)$ for $k \geq \sqrt{n}$. A rather tight upper bound $C(U_{n,k}) \leq k(n+2)$, is due to Tardos and Zwick [TZ97].

\footnote{Getting $C(U_{n,k}) \leq k(n + \log n)$ is trivial, a simple trick due to Karchmer [Kar89] gives $k(n + \log^* n)$.
Karchmer, Raz, and Wigderson [KRW91] also described the monotone version of the universal composition relation, denoted by $U_{n,k}^m$. As before, Alice and Bob are each given a red/blue coloring of a complete $n$-ary tree of depth $k$. This time the colorings are such that both roots are red and if a node is colored red in both inputs then there exists a child colored red in both inputs. The job of Alice and Bob is to find a leaf that is colored red in both inputs. Similarly, to Proposition 6.5 we have the following.

**Proposition 6.6** (Karchmer, Raz, Wigderson [KRW91]). For every set of $k$ monotone Boolean functions $f_1, \ldots, f_k : \{0,1\}^n \to \{0,1\}$ we have $C(R_{KRW}^m(f_1 \circ \cdots \circ f_k)) \leq C(U_{n,k}^m)$.

Similar to the nonmonotone case, the trivial upper bound on $C(U_{n,k})$ is $k(n + \log n)$ and a trick due to Karchmer [Kar89] can reduce it to $k(n + \log^* n)$. We show the upper bound $C(U_{n,k}) \leq k(n + 4)$ in Appendix A. Karchmer, Raz, and Wigderson [KRW91] show a lower bound $C(U_{n,k}^m) \geq kn - 2$ by a simple reduction from the intersection function, which leads us to ask the following question.

**Open Problem 6.7.** Is there a “reduction-style” proof of a lower bound for $C(U_{n,k})$?

Theorem 6.3 might be viewed as an indication that the direct-sum theorem for deterministic communication complexity of relations is out of reach of current techniques; however, it led researchers to consider direct-sum questions in other models of communication. In the next section we survey to which degree direct sum can fail in various CC models.
7 Direct Sum Violations

Quite often the direct-sum theorem in its strongest form fails to hold, i.e., the amount of resources needed to solve $A$ and $B$ together is not exactly the sum of the amount of resources needed to solve $A$ and $B$ separately. Many models admit some small savings, and then the question becomes to quantify the possible savings.

7.1 Deterministic Communication Complexity

Consider the following promise problem $f$, known under the name “league problem”. Alice is given $S \subseteq \{0, \ldots, m-1\}$ with $|S| = 2$ and Bob is given $x \in S$. Their task is to compute the rank of $x$ in $S$. The size of the input is $n = 2 \log m + \log m = 3 \log m$. This problem was studied by Orlitsky [Orl90] and Feder et al. [FKNN95], who showed the following.

**Theorem 7.1** (Orlitsky [Orl90]). *For the league problem $f : \{0,1\}^n \rightarrow \{0,1\}$ we have $C(f) = \Theta(\log n)$."

**Theorem 7.2** (Feder et al. [FKNN95]). *For the league problem $f : \{0,1\}^n \rightarrow \{0,1\}$ we have $\tilde{C}(f) = O(1)$."

Proof of Theorem 7.1 relies on the following observation.

**Proposition 7.3** (Orlitsky [Orl90]). *For every relation $R \subseteq X \times Y \times Z$ we have $C^1(R) \leq 2^{C(R)}$."

Proof. Let $\pi$ be an optimal two-way protocol for $R$. Enumerate all transcripts $\tau_1, \ldots, \tau_\ell$ in some order, where $\ell \leq 2^{C(F)}$. Now we describe a one-way communication protocol for $R$ that uses at most $2^{C(R)}$ bits. The first player sends to the
second player a binary string $s$ of length $\ell$, where bit $s_i$ indicates whether transcript $\tau_i$ is consistent with that player’s input or not. The second player, upon the receipt of the string $s$, can find an index $i$ such that $\tau_i$ is consistent with both players’ inputs. Since we require the answer to $R$ to be evident from the transcript, the second player can infer the answer. Correctness of this protocol is evident from the correctness of $\pi$.

Proof of Theorem 7.1. First we show that $C(f) = O(\log n)$. Given $S = \{s_1, s_2\}$, Alice can find an index $i$ such that $s^i_1 \neq s^i_2$ and send $i$ to Bob. Bob then replies with $x^i$, so Alice find out whether $x = s_1$ or $x = s_2$. The whole communication requires $\log \log m + 1 = O(\log n)$ bits.

For the lower bound, consider one-way protocols, in which Bob communicates first. Bob has to communicate at least $\log m$ bits, for otherwise two distinct inputs $x_1$ and $x_2$ will correspond to the same message. Such a protocol is guaranteed to make a mistake if Alice is given $S = \{x_1, x_2\}$. Thus, we conclude that $C^1(f) = \Omega(n)$. Applying Proposition 7.3, we obtain that $C(f) = \Omega(\log n)$.

In the proof of Theorem 7.2 we shall consider the following family of hash functions. Let $p$ be a prime such that $4 \log m \leq p \leq 8 \log m$. Define

$$H = \{h : \{0, \ldots, m-1\} \rightarrow \{0, \ldots, 7\} \mid h(x) = (ax \mod p) \mod 8, 1 \leq a \leq p-1\}.$$ 

We say that a function $h \in H$ is good for the set $S \subseteq \{0, \ldots, m-1\}$ if $h$ is one-to-one on $S$, and bad otherwise. To show Theorem 7.2 we shall need the following proposition due to Feder et al. [FKN95].

Proposition 7.4 (Feder et al. [FKN95]). Let $S_1, \ldots, S_\ell \subseteq \{0, \ldots, m-1\}$ be such that $|S_i| = 2$ for all $i$. Then there exists a set of $\log \ell + 1$ functions $h_1, \ldots, h_{\log \ell + 1} \in H$ with the property that function $h_i$ is good for at least a half of the $S_j$ for which all $h_1, \ldots, h_{i-1}$ are bad.
We are now ready to show Theorem 7.2.

Proof of Theorem 7.2. We present a protocol due to Feder et al. [FKNN95] to solve \( f^\ell \), where \( f \) is the league problem. Alice is given \( S_1, \ldots, S_\ell \subseteq \{0, \ldots, m-1\} \) where \( |S_i| = 2 \), and Bob is given \( x_1 \in S_1, \ldots, x_\ell \in S_\ell \). Alice finds a set of hash functions \( h_1, \ldots, h_{\log \ell + 1} \) as in Proposition 7.4. From the property described in Proposition 7.4 it follows that for each \( i \in [\ell] \) there is \( j(i) \) such that \( h_{j(i)} \) is good for \( S_i \). Alice sends the names of the functions along with the \( j(i) \) to Bob. Bob then replies with \( h_{j(i)}(x_i) \) to Alice, and Alice has now all the information necessary to compute the ranks of the \( x_i \) in the \( S_i \).

As for the communication cost, first observe that each hash function can be described by \( \log p = O(\log n) \) bits. Thus, Alice sending the names of the functions costs \( O(\log \ell \log n) \) in total. Next, total cost of sending all the \( j(i) \) can be made \( O(\ell) \), since Alice can encode the index of the function good for at least a half of the sets with 2 bits, for at least a half of the remaining sets with 3 bits, and so on. Bob’s reply costs \( O(\ell) \) bits in total. Then, the whole communication requires \( O(\ell + \log \ell \log n) \) bits.

7.2 Randomized Communication Complexity

In this section we shall consider the equality function \( EQ : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\} \) defined by \( EQ(x,y) = 1 \) if and only if \( x = y \).

Proposition 7.5.

\[
C_{R,\epsilon}(EQ) = \Theta(\log 1/\epsilon).
\]

Proof. For the upper bound, we can use the “inner-product protocol”. In round \( i \), Alice sends an inner product over \( \mathbb{F}_2 \) of \( x \) with a random string \( r_i \) to Bob. Since randomness is public, Bob knows \( r_i \) and can compare the bit sent by Alice to the inner product of \( y \) with \( r_i \). If the bits don’t match players agree that
\( x \neq y \), otherwise they continue to the next round. After \( k \) rounds requiring \( 2^k \) bits of communication, the probability that \( x \neq y \) is \( 2^{-k} \). We require \( \epsilon \leq 2^{-k} \), thus \( k = \log 1/\epsilon \) suffices.

For the lower bound, we exhibit a distribution \( \mu \) on \( \{0, 1\}^n \times \{0, 1\}^n \) with the property that \( C_{D, \epsilon}^\mu(EQ) = \Omega(\log 1/\epsilon) \). The result then follows by Proposition 4.3. Define \( \mu \) as follows

\[
\mu = \begin{cases} 
\frac{1}{2^n+t} & \text{if } x = y \\
\frac{1}{2^n+t}(2^n-1) & \text{otherwise.}
\end{cases}
\]

Consider a deterministic protocol \( \pi \) that solves \( EQ \) with probability of error at most \( \epsilon \) when inputs are sampled according to \( \mu \). Let \( R_1, \ldots, R_t \) be 1-rectangles of \( \pi \). Let \( R_i = X_i \times Y_i \). We may assume that for all \( i \) we have \( X_i = Y_i \) (if that’s not the case, with two extra bits of communication Alice and Bob can split the 1 rectangle into four, only one of which is a 1-rectangle and is of the desired form). Let \( s_i = |X_i| = |Y_i| \). The weight of the mistakes in such a rectangle is \( (s_i^2 - s_i)/(2^{n+1}(2^n - 1)) \). Since the weight of all mistakes is at most \( \epsilon \) we obtain \( \sum_i (s_i^2 - s_i)/(2^{n+1}(2^n - 1)) \leq \epsilon \). We also have \( \sum_i s_i = 2^n \) and \( \sum_i s_i^2 \geq 2^{2n}/t \) by Cauchy-Schwarz. Combining these inequalities together it follows \( t \geq 1/(4\epsilon) \), thus \( C(\pi) \geq \log 1/\epsilon - 2 \). \( \square \)

Feder et al. [FKNN95] showed how the “inner-product protocol” can be adapted to handle multiple instances of equality efficiently and showed the following result.

**Theorem 7.6** (Feder et al. [FKNN95]).

\[
C_{R, 2^{-\alpha(\epsilon \pi)}}(EQ^k) = O(k).
\]

Thus, if the error probability \( \epsilon \) is treated as a parameter we obtain a sep-
aration between $C_{R,\epsilon}(EQ) = \Theta(\log 1/\epsilon)$ and amortized $C_{R,\epsilon}(EQ^k)/k = O(1)$ already for $k = \Theta(\log^2 1/\epsilon)$.

However, no gap is known for public-coin randomized communication complexity of functions when $0 < \epsilon < 1/2$ is a fixed constant.

**Open Problem 7.7.** What is the largest gap between $C_R(f)$ and $\tilde{C}_R(f)$?

The log $n$ gap between the private-coin randomized communication complexity and its amortized version is known. This gap is achieved by the equality function. It is well known that $C_{pR}(EQ) = \Theta(\log n)$ (see [KN97], for example). By Theorem 7.6 we have $\tilde{C}_R(EQ) = O(1)$, and it follows by Theorem 8.25 that $\tilde{C}_{pR}(EQ) = O(1)$.

### 7.3 Nondeterministic Communication Complexity

Consider the non-equality function $NE : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ defined by $NE(x, y) = 1$ if there exists $i$ such that $x_i \neq y_i$ and $NE(x, y) = 0$ otherwise. In this section we describe the following result due to Karchmer, Kushilevitz, and Nisan [KKN95]. For non-equality function $NE$ we have $C_{N_1}(NE) = \Theta(\log n)$ and $\tilde{C}_{N_1}(NE) = O(1)$.

A trivial protocol gives $C_{N_1}(NE) \leq \log n + 2$. This is essentially optimal due to the following folklore observation.

**Proposition 7.8.** For every Boolean function $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ we have

$$C(f) \leq 2^{C_{N_1}(f)} + 1.$$ 

**Proof.** Similar to the proof of Proposition 7.3. \qed

Via a simple logrank argument (first observed by Mehlhorn and Schmidt [MS82]) we have $C(NE) = \Theta(n)$. It follows that $C_{N_1}(NE) = \Theta(\log n)$. 

26
To establish the amortized nondeterministic complexity of $NE$, observe that the “inner-product protocol” for equality function (see part (2) of Proposition 5.2) gives a public-coin randomized protocol with one-sided error for $NE$ with $O(1)$ bits of communication. Hence by Theorem 8.6 we have that $\tilde{C}_{N,1}(NE) = O(1)$.

**Open Problem 7.9.** Is there a function $f$ such that $C_N(f) = \Theta(\log n)$ but $\tilde{C}_N(f) = O(1)$?

### 8 Direct Sum Approaches for General Functions and Relations

A common approach to proving a direct-sum theorem for measure $\mathcal{M}$ follows the steps of Program 1. Quite often the first step is used to replace a discrete measure $\mathcal{M}$ by an analytic $\mathcal{K}$, which is usually easier to handle.

In this section we describe several successful implementations of the above paradigm for different communication complexity models.

#### 8.1 Nondeterministic Communication Complexity

The direct-sum question in the nondeterministic setting has one of the most satisfying resolutions. Shortly after the question was raised, two different solutions appeared, first by Feder, Kushilevitz, Naor, and Nisan [FKNN95] and then by Karchmer, Kushilevitz, and Nisan [KKN95]. In this section we present the re-
sult due to Karchmer et al. [KKN95]. Along the way, their approach establishes a precise characterization of the nondeterministic communication complexity in terms of an analytic measure \( \mathfrak{R} \) in the spirit of the Program 1.

With a relation \( R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \) we can associate a hypergraph \( H_R = (V_R, E_R) \), where the set of vertices \( V_R = \mathcal{X} \times \mathcal{Y} \), and the set of hyperedges \( E_R \) consists of all monochromatic rectangles of \( R \). Consider the following integer program:

\[
N(H_R) = \min_{\phi} \sum_{e \in E_R} \phi(e)
\]
\[
\sum_{e \in E \mid v \in e} \phi(e) \geq 1 \quad (\forall v \in V_R)
\]
\[
\phi(e) \in \{0, 1\} \quad (\forall e \in E_R)
\]

Function \( \phi \) satisfying the constraints is called integral cover of \( H_R \). Note that the non deterministic communication complexity of \( R \) is simply equal to \( \log N(H_R) \).

It is quite natural to consider a relaxation of this problem:

\[
N^*(H_R) = \min_{\phi} \sum_{e \in E_R} \phi(e)
\]
\[
\sum_{e \in E \mid v \in e} \phi(e) \geq 1 \quad (\forall v \in V_R)
\]
\[
0 \leq \phi(e) \leq 1 \quad (\forall e \in E_R)
\]

We refer to any \( \phi \) satisfying the constraints of this linear program as a fractional cover of \( H_R \).

In terms of Program 1, we have \( \mathfrak{M}(R) = \log N(H_R) \) and \( \mathfrak{K}(R) = \log N^*(H_R) \).

We proceed to show steps two and three of the approach.

Given two hypergraphs \( H_1 = (V_1, E_1) \) and \( H_2 = (V_2, E_2) \) we can define a product hypergraph \( H_1 \times H_2 = (V_1 \times V_2, E_p) \) where \( E_p = \{e_1 \times e_2 \mid e_1 \in E_1, e_2 \in E_2\} \).
$E_1$ and $e_2 \in E_2$. The measures $N(H)$ and $N^*(H)$ were studied by Lovász [Lov75].

The following proposition demonstrates that the measure $N^*$ obeys the direct product theorem with respect to the product of hypergraphs.

**Proposition 8.1** (Lovász [Lov75]). $N^*(H_1 \times H_2) = N^*(H_1)N^*(H_2)$

Lovász [Lov75] also showed that the integrality gap is not too large.

**Proposition 8.2** (Lovász [Lov75]).

$$N^*(H) \geq \frac{N(H)}{1 + \ln |V(H)|}.$$  

For two relations $R$ and $S$, it is not necessary the case that $H_{R \times S} = H_R \times H_S$, thus we cannot directly invoke Proposition 8.1. The bridge is provided by the following proposition due to Karchmer, Kushilevitz and Nisan [KKN95].

**Proposition 8.3** (Karchmer, Kushilevitz, Nisan [KKN95]). Let $R$ and $S$ be two relations. Then

$$N^*(H_{R \times S}) = N^*(H_R \times H_S).$$

**Theorem 8.4** (Karchmer, Kushilevitz, Nisan [KKN95]). For every relation $R \subseteq X \times Y \times Z$ we have

(a) $k \log N^*(H_R) \leq C_N(R^k) \leq k \log N^*(H_R) + \log \log |X||Y| + \log k + 5$

(b) $\bar{C}_N(R) = \log N^*(H_R) \geq C_N(R) - \log \log |X||Y| - 5$

**Proof.**

**Part (a)** We have $C_N(R^k) = \log N(H_{R^k}) \geq \log N^*(H_{R^k}) = \log N^*(H_{R^k}) = k \log N^*(H_R)$, where the last two equalities follow from Proposition 8.3 and Proposition 8.1, respectively.

---

7 An interested reader who decides to follow up on this reference should be alerted that Lovász’s statements are in terms of duals of our graphs.
We have $C_N(R^k) = \log N(H_{R^k}) \leq \log((1+\ln |\mathcal{X}|^k|\mathcal{Y}|^k)N^*(H_{R^k})) \leq k \log N^*(H_{R^k}) + \log \log |\mathcal{X}| |\mathcal{Y}| + \log k + 5$, where the first inequality is due to Proposition 8.2.

**Part (b)** Immediately follows from Part (a). Also note that in the definition of $\tilde{C}_N(R)$ we can replace lim sup with just lim, as Part (a) implies that the limit exists.

If we repeat the above process for the one-sided “NP-like” version of non-deterministic communication complexity, i.e., when we require to cover only 1-inputs, we obtain a similar theorem.

**Theorem 8.5** (Karchmer, Kushilevitz, Nisan [KKN95]). For every relation $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ we have

$$\tilde{C}_{N,1}(R) = \log N^*_1(R),$$

where $N^*_1(R)$ denotes the solution to the relaxed linear program for covering 1-rectangles of $R$.

We end this section with an interesting corollary: an upper bound on $\tilde{C}_{N,1}(f)$ of a function $f$ in terms of one-sided randomized communication complexity $C_{R^1}(f)$.

**Corollary 8.6** (Karchmer, Kushilevitz, Nisan [KKN95]). For every function $f : \mathcal{X} \times \mathcal{Z} \rightarrow \{0, 1\}$ we have

$$\tilde{C}_{N,1}(f) \leq C_{R^1}(f) + 1.$$

### 8.2 Deterministic Communication Complexity

To apply Program 1 to the deterministic communication complexity, we need to define measure $\mathfrak{R}$. The idea due to Feder et al. [FKN95] is to apply Program 1
to $R = C_N$, since we already know the direct-sum theorem for $C_N$. Thus, Feder et al. [FKNN95] showed that for functions the separation between $C$ and $\tilde{C}$ can be at most quadratic.

**Theorem 8.7** (Feder et al. [FKNN95]). Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a function. Then

$$\tilde{C}(f) = \Omega(\sqrt{C(f)} - \log \log |\mathcal{X}| |\mathcal{Y}| - 1).$$

**Proof.** We have

$$C(f^k) \geq C_N(f^k) \geq k \log N^*(H_f) \quad \text{by Theorem 8.4 part (a)}$$

$$\geq k (C_N(f) - \log \log |\mathcal{X}| |\mathcal{Y}| - 5) \quad \text{by Theorem 8.4 part (b)}$$

$$= \Omega \left( k \left( \sqrt{C(f)} - \log \log |\mathcal{X}| |\mathcal{Y}| - 1 \right) \right) \quad \text{by Theorem 4.1.}$$

Unfortunately, this approach does not work for relations or even partial functions. The gap between $C(f)$ and $C_N(f)$ can be exponential if $f$ is partial. The following example is implicit in [Raz90] and (is hinted at) in [FKNN95]. Let $x, y \in \{0, 1\}^n$ and let $x^{(0)}$ denote the first half of $x$ and let $x^{(1)}$ denote the second half of $x$. Define $y^{(0)}$ and $y^{(1)}$ similarly. Consider partial function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ defined by $f(x, y) = z$ if $x^{(z)} = y^{(z)}$ and $x^{(1-z)} \neq y^{(1-z)}$ where $z = 0, 1$. Function $f$ is undefined on all other inputs. We have $C_N(f) = O(\log n)$, since the players can guess an index $i$ such that $x_i \neq y_i$ and depending on whether $i \in [n/2]$ or $i \in [n] \setminus [n/2]$ answer 1 and 0, respectively. The lower bound $C(f) = \Omega(n)$ follows from a simple fooling set argument. Consider set $S = \{(x, x) \mid x \in \{0, 1\}^n\}$. Suppose that a combinatorial rectangle $R$ contains more than $2^{n/2}$ entries from $S$. Then by pigeonhole principle there exist inputs $(x, x)$ and $(x', x')$ from $S$ such that $x^{(0)} = x'^{(0)}$ and $x^{(1)} \neq x'^{(1)}$. Thus, if rectangle
were monochromatic it would have to be labeled 0. However, by the same argument $R$ also has to contain $(x, x)$ and $(x', x')$ such that $x(0) \neq x'(0)$ and $x(1) = x'(1)$, thus $R$ has to be labeled 1. Therefore a monochromatic rectangle cannot contain more than $2^{n/2}$ entries from $S$. Since $|S| = 2^n$, a deterministic protocol is required to produce at least $2^{n/2}$ rectangles at its leaves. The claimed lower bound follows immediately.

8.3 Randomized Communication Complexity

The direct sum for randomized communication complexity did not see much progress until Chakrabarti et al. [CSWY01] introduced an information theoretic approach to the area\(^8\). They showed that functions satisfying a certain robustness criterion (the equality function is one such example) obey the direct sum in the simultaneous communication complexity. In this model, there is only one round of communication and both players submit their messages simultaneously. In their work, Chakrabarti et al. introduced the notion of the information cost of a protocol, which behaves well with respect to the direct sum. Using similar techniques Jain et al. [JRS03] show a direct-sum for distributional communication complexity of bounded-round protocols under product distributions.

Bar-Yossef et al. [BYJKS04] observed that the direct sum for the information cost of a protocol can be used to show a lower bound on the number-in-hand (NIH) $t$-party communication complexity of the disjointness function. The disjointness function can be viewed as $\bigvee_{i=1}^{n} \bigwedge_{j=1}^{t} x_{ij}$. Thus, the lower bound on CC of the disjointness problem reduces to proving a lower bound on the information cost of the AND function $\bigwedge_{j=1}^{t} x_{ij}$. Using this paradigm, Bar-Yossef et al. [BYJKS04] showed $\Omega(n/t^2)$ lower bound on the CC of the

\(^8\)Information theory was used in the communication complexity before, see for example the work of Ablayev [Abl96], but not in the context of direct sum.
t-party disjointness. Gronemeier [Gro09] closed the gap between upper bound and lower bound of the NIH CC of disjointness by providing a tighter analysis of the information cost of the AND function and ultimately showing $\Omega(n/t)$ lower bound for the disjointness function.

The first direct-sum result for general randomized protocols is due to Barack, Braverman, Chan, and Rao [BBCR10]. They showed that the communication required for computing $n$ instances of a problem is at least $\sqrt{n}$ times the communication required to compute a single instance of a problem. A promising direction to the direct-sum question in the randomized communication complexity is to better understand the information cost measure. Since the paper by Barack et al. [BBCR10], the information cost measure was studied in [BR11], [Bra11], and [BW11].

8.3.1 Information Theory Background

**Definition 8.8.** The entropy of the probability distribution $p$ on sample space $\Omega$ is defined as

$$H(p) = \sum_{\omega \in \Omega} p(\omega) \log \frac{1}{p(\omega)}.$$ 

The entropy of a random variable $X$ is defined as the entropy of the induced probability distribution on the range of $X$.

Let $X$ and $Y$ be two random variables. For $y \in \text{range}(Y)$ we can consider the conditional probability distribution $p(X \mid Y = y)$. We denote the entropy of this probability distribution by $H(X \mid Y = y)$.

**Definition 8.9.** The conditional entropy of random variable $X$ given $Y$ is defined as

$$H(X \mid Y) = \mathbb{E}_y (H(X \mid Y = y)).$$

For two random variables $X,Y$ we have $H(X,Y) = H(X) + H(Y \mid X)$. 

33
Definition 8.10. The mutual information of two random variables $X$ and $Y$ is defined as

$$I(X;Y) = H(X) - H(X|Y).$$

Note that we also have $I(X;Y) = H(Y) - H(Y|X)$.

Definition 8.11. The total variation distance between two probability distributions $p$ and $q$ on $\Omega$ is defined as

$$||p - q|| = \frac{1}{2} \sum_{\omega \in \Omega} |p(\omega) - q(\omega)|.$$

Definition 8.12. Kullback-Leibler divergence between two probability distributions $p$ and $q$ on $\Omega$ is defined as

$$d_{KL}(p \parallel q) = \sum_{\omega \in \Omega} p(\omega) \log \frac{p(\omega)}{q(\omega)}.$$

8.3.2 Information Cost Measure

Suppose that $\mu$ is a probability distribution on $\mathcal{X} \times \mathcal{Y}$ and $\pi$ is a randomized protocol with both public and private randomness computing a function on $\mathcal{X} \times \mathcal{Y}$. Informally, the information cost of such a protocol measures how much information about the inputs the parties reveal to each other (internal information cost) or to an independent observer (external information cost) during the execution of the protocol on input $(x,y) \sim \mu$. We use random variable $R$ to denote public randomness of $\pi$ and $\pi_r$ to denote the private-coin protocol obtained from $\pi$ by fixing $R = r$. Formally, the two measures are defined as follows.

Definition 8.13. Given a distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$, the external information cost of protocol $\pi$ with respect to $\mu$ is defined as

$$\text{IC}^{\text{ext}}(\pi, \mu) = I(X,Y; R_{\pi_R}(X,Y)).$$
and the internal information cost of protocol $\pi$ with respect to $\mu$ is defined as

$$IC^i(\pi, \mu) = I(X; R_{\pi_R}(X, Y)|Y) + I(Y; R_{\pi_R}(X, Y)|X).$$

Previously, when we dealt with the communication complexity of a protocol we never considered protocols with both public and private randomness. Indeed, we did not have to. Any protocol with both public and private randomness can be replaced by a protocol with just public randomness without any increase in the communication complexity. Alice and Bob simply partition the shared random string into three disjoint parts, use one of the parts for the shared randomness, and the other two for “private” randomness. The distinction between private and public randomness becomes crucial to the definition of the information cost measures, as making private bits public may increase the information cost dramatically. For example, suppose that in the course of a protocol Alice wants to send the first bit of her input XORed with a random bit to Bob. If the random bit was generated privately, Alice’s message reveals 0 information to Bob. If, on the other hand, the random bit was generated publicly, Alice’s message reveals 1 bit of information to Bob.

The information cost can be defined for functions. The following two natural definitions were considered in [Bra11].

**Definition 8.14.** Let $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be a function. The prior-free information cost of $f$ is defined as

$$IC^i(f, \epsilon) = \inf_{\pi} \max_{\mu} IC^i(\pi, \mu),$$

where the infimum ranges over all protocols $\pi$ that compute $f$ with probability of error at most $\epsilon$ on each input.
The distributional information cost of $f$ is defined as

$$IC_{D}^i(f,\epsilon) = \max_{\mu} \inf_{\pi} IC^i(\pi,\mu),$$

where the infimum ranges over all protocol $\pi$ that compute $f$ with probability of error at most $\epsilon$ when inputs are sampled from $\mu$.

Analogous definitions exist for the external information cost.

Clearly, we have $IC_{D}^i(f,\epsilon) \leq IC^i(f,\epsilon)$. Braverman [Bra11] showed that the two definitions are essentially equivalent.

**Theorem 8.15** (Braverman [Bra11]). For every function $f : X \times Y \rightarrow Z$ and $\epsilon \geq 0$ we have $IC^i(f,\epsilon) \leq 2IC_{D}^i(f,\epsilon/2)$.

In the light of the above theorem, we shall use the prior-free information cost in what follows understanding that the two definitions can be used interchangeably.

The information cost measures were introduced with the goal of implementing Program 1:

1. $IC^i$ and $IC^\text{ext}$ are analytic measures,
2. $IC^i$ ([Bra11]) obeys the direct sum,
3. $IC^i \leq IC^\text{ext} \leq C_R$ ([BBCR10]) and the hope is that $IC^i = \Theta(C_R)$.

The third step of the program is not completed yet, and the exact relationship between $IC^i$ and $C_R$ is not known.

**Open Problem 8.16.** Is $IC^i(f,\epsilon) = \Theta(C_R,\epsilon(f))$ for all $f$ and $\epsilon \geq 0$?

Observe that it is not known whether $IC^\text{ext} = O(IC^i)$.

**Open Problem 8.17.** Is there a function $f$ and $\epsilon \geq 0$ such that $IC^\text{ext}(f,\epsilon) > IC^i(f,\epsilon)$? If so, what’s the largest separation possible?
It is also not known whether $\text{IC}^{\text{ext}}$ obeys the direct sum or not. Indeed, any restriction-based proof that $\text{IC}^{\text{ext}}$ satisfies the direct sum would settle both the direct-sum problem and Open Problem 8.17. This observation was communicated to us by Mark Braverman. To demonstrate this claim we shall need the following two results.

**Theorem 8.18** (Braverman [Bra11]). For $\epsilon > 0$ we have

$$\text{IC}(f, \epsilon) = \lim_{n \to \infty} \frac{C_{R, \epsilon}(f^n)}{n}.$$ 

**Theorem 8.19** (Barack et al. [BBCR10]). For every distribution $\mu$, protocol $\pi$ and $\alpha > 0$, there exists functions $\pi_x, \pi_y$ and a protocol $\tau'$ such that $|\pi_x(X, \tau'(X, Y)) - \pi(X, Y)| < \alpha$, $P(\pi_x(X, \tau'(X, Y)) \neq \pi_y(Y, \tau'(X, Y))) < \alpha$ and

$$C(\tau') = O\left(\text{IC}^{\text{ext}}(\pi, \mu)(\log(C(\pi)/\alpha))/\alpha^2\right).$$

**Proposition 8.20.** Suppose that the following statement is true: given $\mu, f, n, \epsilon$ and a protocol $\pi$ that solves $f^n$ with probability of error at most $\epsilon$ (on each coordinate) when inputs are sampled from $\mu^n$ there exists a protocol $\tau$ solving $f$ with probability of error at most $\epsilon$ when inputs are sampled according to $\mu$ such that $C(\tau) \leq C(\pi)$ and $\text{IC}^{\text{ext}}(\tau, \mu) \leq C(\pi)/n$. Then for all $f$ and $\epsilon > 0$ we have

1. $\text{IC}^{\text{ext}}(f, \epsilon) = \text{IC}(f, \epsilon)$,

2. $C_{R, \epsilon}(f^n) = \tilde{\Omega}(nC_{R, \epsilon}(f))$.

**Proof.**

1. $\text{IC}(f, \epsilon) \leq \text{IC}^{\text{ext}}(f, \epsilon) \leq C_{R, \epsilon}(f^n)/n \xrightarrow{n \to \infty} \text{IC}(f, \epsilon)$, where the last step is by Theorem 8.18.

2. Fix $\alpha > 0$. Pick a distribution $\mu$ that achieves maximum of $C_{D, \epsilon + \alpha}(f)$.

Let $\pi$ be a protocol that solves $f^n$ with probability of error at most $\epsilon$ and
\[ C(\pi) \leq C_{R,\epsilon}(f^n) \leq C_{R,\epsilon}(f^n). \] Applying the premise of the proposition, we obtain a protocol \( \tau \) such that \( C(\tau) = C(\pi) \) and \( \text{IC}^\text{ext}(\tau, \mu) \leq C(\pi)/n. \)

Now we apply Theorem 8.19 to obtain a protocol \( \tau' \) such that \( C(\tau') = \tilde{O}(\text{IC}^\text{ext}(\tau, \mu)) = \tilde{O}(C_{R,\epsilon}(f^n)/n) \). Moreover, since \( \tau' \) solves \( f \) with probability of error at most \( \alpha + \epsilon \) when inputs are sampled according to \( \mu \) we have \( C(\tau') \geq C_{D,\epsilon+\alpha}^b(f) = C_{R,\epsilon+\alpha}(f) \) by Proposition 4.3.

\[ \square \]

In the rest of this section, we shall examine the role of the distinction between private and public randomness in establishing the relationship between \( \text{IC}^i \) and \( C_R \). In the process, we shall use various techniques that have appeared in the literature on the information cost.

### 8.3.3 Public vs Private Randomness in the Information Cost: a New Result

**Definition 8.21.** The public internal information cost of a function \( f \) with error parameter \( \epsilon \geq 0 \) is defined as

\[ \text{IC}^i,\text{pub}(f, \epsilon) = \inf_{\pi} \max_{\mu} \text{IC}^i(\pi, \mu), \]

where the infimum ranges over all public-coin protocols (no private coins allowed) \( \pi \) that compute \( f \) with probability of error at most \( \epsilon \) on each input.

Clearly, for every \( \epsilon \geq 0 \) we have \( \text{IC}^i,\text{pub}(f, \epsilon) \geq \text{IC}^i(f, \epsilon) \). We say that the function \( f \) is **bold** if\(^9\) the opposite holds up to logarithmic factors, i.e.,

\[ \text{IC}^i,\text{pub}(f, \epsilon) = \tilde{O}(\text{IC}^i(f, \epsilon)). \]

In this section, we shall show the following result. We believe this result is new.

\(^9\)In particular, \( f \) is not intimidated by going public.
**Theorem 8.22.** The following two statements are equivalent:

1. For all $f$ and $\epsilon \geq 0$ we have $\text{IC}^i(f, \epsilon) = \tilde{\Theta}(C_{R, \epsilon}(f))$.

2. Every $f$ is bold.

**Proof of 1 implies 2.**

Take the best protocol $\pi$ for $f$ provided by the definition of $C_{R, \epsilon}$. Since $\pi$ only has public randomness we have $\text{IC}^i,\text{pub}(f, \epsilon) \leq \text{IC}^i(\pi) \leq C_{R, \epsilon}(f) = \tilde{O}(\text{IC}^i(f, \epsilon))$. \hfill $\square$

The idea for the proof of direction “2 implies 1” is to start with a public-coin protocol $\pi$ that has small information cost, but might have a large communication cost, and derive protocol $\tau$ with communication cost roughly equal to the information cost of $\pi$. This process is called “protocol compression” in the work of Barack et al. [BBCR10]. In fact, we shall use their compression scheme, and show that in the deterministic setting it gives an almost optimal compression.

We start with a few lemmas.

**Lemma 8.23** (Barack et al. [BBCR10]). Let $\pi$ be a protocol, and $\mu$ be a distribution on inputs $\mathcal{X} \times \mathcal{Y}$.

$$\text{IC}^i(\pi, \mu) = \mathbb{E}_R(\text{IC}^i(\pi_R, \mu_R)).$$

**Proof.** First we show how to manipulate the first term in the definition of $\text{IC}^i(\pi, \mu)$.

$$I(X; R_{\pi_R}(X, Y)|Y) = H(R_{\pi_R}(X, Y)|Y) - H(R_{\pi_R}(X, Y)|XY)$$

$$= \ H(R|Y) + H(\pi_R(X, Y)|YR) - H(R|YX) - H(\pi_R(X, Y)|YX R)$$

$$= \ I(R; X|Y) + I(\pi_R(X, Y); X|YR)$$

$$= \ 0 + I(\pi_R(X, Y); X|YR).$$
Where $I(R; X|Y) = 0$ because $R$ and $X$ are independent. Similarly, we obtain $I(Y; R\pi_R(X,Y)|X) = I(\pi_R(X,Y); Y|XR)$. Finally, we have

$$IC^i(\pi, \mu) = I(\pi_R(X,Y); X|YR) + I(\pi_R(X,Y); Y|XR) = E_R(IC^i(\pi_R, \mu)).$$

\[\square\]

Now, let $\pi$ be a public-coin protocol and $\mu$ be a probability distribution on the inputs $\mathcal{X} \times \mathcal{Y}$. For each random string $r$ protocol $\pi_r$ is deterministic, and can be viewed as a complete binary tree of depth $C(\pi)$. With each node $u$ we associate three objects:

1. $o_u \in \{\text{Alice, Bob}\} -$ the owner of the node,

2. $f_u : \begin{cases} \mathcal{X} \to \{0, 1\} & \text{if } o_u = \text{Alice,} \\ \mathcal{Y} \to \{0, 1\} & \text{if } o_u = \text{Bob.} \end{cases}$

This function specifies, which bit is transmitted by the owner of the node.

3. $p_u : \begin{cases} \mathcal{X} \to \{0, 1\} & \text{define by } p_u(x) = P(f_u(y) = 1|\text{reached } u, X = x), \\ & \text{if } o_u = \text{Alice,} \\ \mathcal{Y} \to \{0, 1\} & \text{defined by } p_u(y) = P(f_u(x) = 1|\text{reached } u, Y = y), \\ & \text{if } o_u = \text{Bob.} \end{cases}$

In other words, $p_u$ is the probability estimated by the non-owner of the node that the owner sends 1 as the next bit.

Each internal node $u$ has two outgoing edges, one of which is labeled 1 and the other 0. The leaves are labeled with values from $\mathcal{Z}$. The protocol is executed by starting at the root and using functions $f_u$ to decide which edge to follow,
until a leaf is reached, at which point its label is declared to be the output of the protocol.

The following compression scheme appears in [BBCR10]. Out of protocol \( \pi \) we construct the following protocol \( \tau_{k,\gamma} \), where \( k \) and \( \gamma \) are parameters to be specified later.

**Sampling Stage** Alice and Bob sample random string \( r \). Now, Alice and Bob each has a tree representation of \( \pi_r \). Alice and Bob use public randomness to sample a uniformly random number \( t_u \in [0,1] \) for each node \( u \). They set the current node to be the root.

**Path Construction** Without communicating to Bob, Alice builds a path \( v = v_0, \ldots, v_{C(\pi)} \) as follows. Starting with the current node \( v_i \), if \( o_{v_i} = \text{Alice} \) then \( v_{i+1} \) is obtained from \( v_i \) by applying \( f_{v_i} \) and following the corresponding edge in the tree. Otherwise, Alice follows the 1-edge if \( t_{v_i} \leq p_{v_i} \) and the 0-edge otherwise. Similarly, Bob constructs his path \( w = w_0, \ldots, w_{C(\pi)} \).

**Finding Mistakes** Alice and Bob use the protocol from Proposition 4.5 for finding the first difference between \( v \) and \( w \) with error probability at most \( \gamma \).

**Loop** If no difference is found or Path Construction and Finding Mistakes were performed \( k \) times, Alice and Bob terminate their communication. Otherwise, the non-owner of the node of the first difference fixes the mistake, Alice and Bob set the current node to be the node of the first difference, and they both go to step Path Construction.

The following lemma lies at the heart of the analysis of the above protocol compression.

**Lemma 8.24.** The expected number of times step Finding Mistakes has to be executed in \( \tau_{k,\gamma} \) to arrive at the correct leaf of \( \pi \) is at most \( IC^1(\pi,\mu) \).
Proof. Consider $\tau_{k,\gamma,r}$ - the protocol $\tau_{k,\gamma}$ after $r$ has been fixed. Let $S_{i,r}$ be the indicator variable whether a mistake occurred at the $i$-th level of the tree $\pi_r$, $i \in [C(\pi)]$. The $i$-th bit of protocol $\pi_r$ is denoted by $\pi_r^i$. The concatenation of the first $i-1$ bits of $\pi_r$ is denoted by $\pi_r^{<i}$.

Suppose that during the execution of $\tau_{k,\gamma,r}$ the current node is $u$. Moreover, assume that $u$ is owned by Alice and $f_u(x) = 1$. The probability that Bob makes a mistake at this node is $1 - p_u \leq \ln(1/p_u) < \log(1/p_u) = d_{KL}((\pi_r^i | \pi_r^{<i}XY)|| (\pi_r^i | \pi_r^{<i}Y))$. If $f_u(x) = 0$ we still obtain that Bob makes a mistake at node $u$ with probability $p_u \leq \log(1/(1-p_u)) = d_{KL}((\pi_r^i | \pi_r^{<i}XY)|| (\pi_r^i | \pi_r^{<i}Y))$.

Similar inequalities hold if Bob owns node $u$. Therefore, we get

$$
\mathbb{E}(S_{i,r}) = \mathbb{E}(d_{KL}((\pi_r^i | \pi_r^{<i}XY)|| (\pi_r^i | \pi_r^{<i}Y)) + d_{KL}((\pi_r^i | \pi_r^{<i}XY)|| (\pi_r^i | \pi_r^{<i}X))) = I(X; \pi_r^i | Y\pi_r^{<i}) + I(Y; \pi_r^i | X\pi_r^{<i}).
$$

Thus, we have

$$
\mathbb{E}(\sum_{i=1}^{C(\pi)} S_{i,r}) = \sum_{i=1}^{C(\pi)} \mathbb{E}(S_{i,r}) \leq \sum_{i=1}^{C(\pi)} I(X; \pi_r^i | Y\pi_r^{<i}) + I(Y; \pi_r^i | X\pi_r^{<i}) = I(X; \pi_r(X,Y) | Y) + I(Y; \pi_r(X,Y) | X) = IC^i(\pi_r, \mu).
$$

Let $S_i$ be the indicator random variable that a mistake occurred at step $i$ during execution of $\tau$. Using Lemma 8.23 we obtain

$$
\mathbb{E} \left( \sum_{i=1}^{C(\pi)} S_i \right) = \mathbb{E}_R \left( \sum_{i=1}^{C(\pi)} S_{i,R} \right) \leq \mathbb{E}_R(\text{IC}^i(\pi_R, \mu)) = \text{IC}^i(\pi, \mu).
$$

Now, we put all the pieces together and finish the proof Theorem 8.22.
Proof of 2 implies 1 in Theorem 8.22.

Let \( f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \) and \( \epsilon \geq 0 \) be given. Let \( \pi \) be the best public-coin protocol provided by the definition of \( \text{IC}^i_{\text{pub}}(f, \epsilon) \). By Proposition 4.3 we know that \( C_{R,3\epsilon}(f) = \max_\mu C^\mu_{D,3\epsilon}(f) \). Fix \( \mu \) to be the distribution that achieves the maximum on the right hand side.

Apply the compression scheme described above to obtain the protocol \( \tau_{k,\gamma} \) with parameters \( k = \text{IC}^i(\pi, \mu)/\epsilon \) and \( \gamma = \epsilon^2/\text{IC}^i(\pi, \mu) \). The probability (taken over public coin flips and inputs) that \( \tau_{k,\gamma} \) outputs an incorrect value is bounded above by the sum of probabilities of the following three events:

1. the number of mistakes in \( \tau_{k,\gamma} \) is more than \( k \), the probability of which is \( P(S \geq k) \leq \frac{\mathbb{E}(S)}{k} = \frac{\text{IC}^i(\pi, \mu)/k}{k} = \epsilon \) by Markov’s inequality and Lemma 8.24,

2. the protocol for finding the first mistake fails, the probability of which is, by union bound, at most \( k\gamma = (\text{IC}^i(\pi, \mu)/\epsilon) \cdot (\epsilon^2/\text{IC}^i(\pi, \mu)) = \epsilon \),

3. given that the output of \( \tau_{k,\gamma} \) matches that of \( \pi \), the protocol \( \pi \) could still output the wrong value with probability at most \( \epsilon \).

It follows that for some choice of public coins the probability of success at least \( 1 - 3\epsilon \) is achieved, when the probability is taken only over the random inputs.

Hence we have \( C_{R,3\epsilon}(f) = C^\mu_{D,3\epsilon}(f) = O(k \log(C(\pi)/\gamma)) = O((\text{IC}^i(\pi, \mu)/\epsilon) \cdot (\log((C(\pi) \text{IC}^i(\pi, \mu))/\epsilon^2))) = \tilde{O}(\text{IC}^i(\pi, \mu)) \). We have

\[
C_{R,\epsilon}(f) = \Theta(C_{R,3\epsilon}(f)) = \tilde{O}(\text{IC}^i(\pi, \mu)) = \tilde{O}(\text{IC}^i_{\text{pub}}(f, \epsilon)) = \tilde{O}(\text{IC}^i(f, \epsilon)),
\]

where the first step follows from sequential repetition and the last step follows since \( f \) is bold. \( \square \)

In particular, we have shown that getting rid of private randomness in the
definition of information cost is equivalent, up to logarithmic factors, to the
direct-sum theorem for the randomized communication complexity.

We end this section with a remark that the question of private vs public
randomness and how it affects the direct sum has appeared before, although in
a very different context. The following theorem is due to Feder et al. [FKNN95].

**Theorem 8.25 (FKNN1995).** Let \( f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\} \) be a partial
function. Then

\[ \overline{C}_R^p(f) = \Theta(C_R(f)). \]

**Proof.** Clearly, \( C_R(f^k) \leq C_R^p(f^k) \). In addition, by Theorem 4.2 we have \( C_R^p(f^k) = O(C_R(f^k) + \log kn) \).

\[ \square \]

## A Upper Bound on the Monotone Universal Composition Relation

The proof of Tardos and Zwick [TZ97] is based on *Hamming error-correcting
codes*, which we briefly review here. Let \( n = 2^r - 1 \). Hamming code encodes
messages of length \( n - r \) over a binary alphabet into codewords of length \( n \) as
follows. The bits at positions \( 2^i \) for \( i = 0, \ldots, r - 1 \) are parity bits, the rest
are (unmodified) message bits. For each \( i = 0, \ldots, r - 1 \), the bit at position
\( 2^i \) is computed as a parity of message bits appearing at indices that contain
\( 2^i \) in their binary expansion. To correct a single error, the receiver computes
parity bits of a received codeword, XORs those bits with the parity bits of the
received codeword, and the result of the XOR operation encodes the index of
the error. Observe that the total number of codewords with at most one bit
of error is \( 2^{n-r}(n+1) = 2^n \), i. e., distance-1 Hamming neighborhoods of the
valid codewords partition the space of all strings of length \( n \). We shall take the
correct message as a representative for a block of this partition.
The trivial upper bound on $C(U_{n,k})$ is $k(n + \log n)$ and a trick due to Karchmer [Kar89] can reduce it to $k(n + \log^* n)$. Next, we show how to modify the algorithm of Tardos and Zwick [TZ97] to handle the monotone case and show $C(U_{n,k}) \leq k(n + 4)$. We believe this result was not pointed out before.

**Proposition A.1.** $C(U_{m,n,1}) \leq n + 4$.

**Proof.** In $U_{m,n,1}$, Alice is given $x \in \{0,1\}^n$ and Bob is given $y \in \{0,1\}^n$ with a promise that there exists $i$ with $x^i = y^i = 1$. Their job is to find such an index $i$. We shall describe a protocol that achieves this task with $n + 4$ bits of communication. Let $n = 2^r - 1 + s$ where $0 \leq s \leq 2^r$, and define $n_1 = 2^{r-1} - 1$ and $n_2 = s + 2^{r-1}$. Let $x_1$ be the prefix of $x$ of length $n_1$ and $x_2$ be the suffix of $x$ of length $n_2$, thus concatenation of $x_1$ with $x_2$ is the whole string $x$. Define $y_1$ and $y_2$ similarly. Alice regards $x_1$ as a codeword (possibly with an error) from the Hamming code of length $n_1$. Let $\text{msg}(x_1)$ denote the representative of a block, to which $x_1$ belongs. Alice sends $\text{msg}(x_1)$ to Bob.

If $|\text{msg}(x_1) \cap y_1| \geq 2$ then Bob sends 0 to Alice together with $y_1$. Alice finds an index $i$ such that $x_1^i = \text{msg}(x_1)^i = y_1^i = 1$, which is guaranteed to exist, and sends it to Bob. In this case, the whole communication takes $(2^{r-1} - r) + (1 + 2^{r-1} - 1) + (r - 1) \leq n$ bits.

Otherwise, if $|\text{msg}(x_1) \cap y_1| \leq 1$ Bob sends 1 to Alice together with $y_2$. If Alice finds an index $i$ such that $y_2^i = x_2^i = 1$, then she sends 0 to Bob along with index $i$. In this case the whole communication takes $(2^{r-1} - r) + (1 + n_2) + (1 + \log n_2) \leq n + 4$ bits.

The remaining case is when Alice does not find the desired index in the second half of the input, so she notifies Bob by sending 1 to him. If $|\text{msg}(x_1) \cap y_1| = 1$, Bob sends 0 to Alice together with the only index $i$ from the intersection, since Alice must have $x_1^i = \text{msg}(x_1)^i = 1$, for otherwise the promise of the problem would be violated. If $|\text{msg}(x_1) \cap y_1| = 0$, Bob sends 1 to Alice, and
Alice responds with the index $i$ such that $x_1^i \neq \text{msg}(x_1)^i$, as this is the only candidate for the answer. In both last cases the whole communication takes $(2^{r-1} - r) + (1 + n_2) + 1 + r \leq n + 2$. \hfill \qed

**Corollary A.2.** $C(U_{n,k}^m) \leq k(n + 4)$.

**References**


