Information Complexity and the Quest for Interactive Compression

(Survey)

Omri Weinstein*

Abstract

Information complexity is the interactive analogue of Shannon's classical information theory. In recent years this field has emerged as a powerful tool for proving strong communication lower bounds, and for addressing some of the major open problems in communication complexity and circuit complexity. A notable achievement of information complexity is the breakthrough in understanding of the fundamental direct sum and direct product conjectures, which aim to quantify the power of parallel computation. This survey provides a brief introduction to information complexity, and overviews some of the recent progress on these conjectures and their tight relationship with the fascinating problem of compressing interactive protocols.

1 Introduction

The holy grail of complexity theory is proving lower bounds on different computational models, thereby delimiting computational problems according to the resources they require for solving. One of the most useful abstractions for proving such lower bounds is *communication complexity*; Since its introduction [Yao79], this model has had a profound impact on nearly every field of theoretical computer science, including VLSI chip design, data structures, mechanism design, property testing and streaming algorithms [Wac90, PW10, DN11, BBM12] to mention a few, and constitutes one of the few known tools for proving *unconditional* lower bounds. As such, developing new tools in communication complexity is a promising approaches for making progress within computational complexity, and in particular, for proving strong circuit lower bounds that appear viable (such as Karchmer-Wigderson games and ACC lower bounds [KW88, BT91]).

Of particular interest are "black box" techniques for proving lower bounds, also known as "hardness amplification" methods (which morally enable strong lower bounds on composite problems via lower bounds on a simpler primitive problem). Classical examples of such results are the Parallel Repetition theorem [Raz98, Rao08] and Yao's XOR Lemma [Yao82], both of which are cornerstones of complexity theory. This is the principal motivation for studying the *direct sum and direct product* conjectures, which are at the core of this survey.

Perhaps the most notable tool for studying communication problems is information theory, introduced by Shannon in the late 1940s in the context of (one-way) data transmission problems [Sha48]. Shannon's noiseless coding theorem revealed the tight connection between communication

1

and information, namely, that the amortized description length of a random one-way message (M) is equivalent to the amount of information it contains

$$\lim_{n \longrightarrow \infty} \frac{C(M^n)}{n} = H(M), \tag{1}$$

where M^n denotes n i.i.d observations from M, C is the minimum number of bits of a string from which M^n can be recovered (w.h.p), and $H(\cdot)$ is Shannon's Entropy function. In the 65 years that elapsed since then, information theory has been widely applied and developed, and has become the primary mathematical tool for analyzing communication problems.

Although classical information theory provides a complete understanding of the one-way transmission setup (where only one party communicates), it does not readily convert to the *interactive setup*, such as the (two-party) communication complexity model. In this model, two players (Alice and Bob) receive inputs x and y respectively, which are jointly distributed according to some prior distribution μ, and wish to compute some function $f(x, y)$ while communicating as little as possible. To do so, they engage in a *communication protocol*, and are allowed to use both public and private randomness. A natural extension of Shannon's entropy to the interactive setting is the *Information Complexity* of a function $\mathsf{IC}_\mu(f, \varepsilon)$, which informally measures the average amount of information the players need to disclose each other about their inputs in order to solve f with some prescribed error under the input distribution μ. From this perspective, communication complexity can be viewed as the extension of transmission problems to general tasks performed by two parties over a noiseless channel (the noisy case recently received a lot of attention as well [Bra14]). Interestingly, it turns out that an analogue of Shannon's theorem does in fact hold for interactive computation, asserting that the amortized communication cost of computing many independent copies of any function f is precisely equal to its single-copy information complexity:

Theorem 1.1 ("Information = Amortized Communication", [BR11])**.** *For any $\varepsilon > 0$ and any two-party communication function $f(x, y)$,*

$$\lim_{n \longrightarrow \infty} \frac{\mathsf{D}_{\mu^n}(f^n, \varepsilon)}{n} = \mathsf{IC}_\mu(f, \varepsilon).$$

Here $\mathsf{D}_{\mu^n}(f^n, \varepsilon)$ denotes the minimum communication required for solving n independent instances of f with error at most ε on *each copy*.[1] The above theorem assigns an *operational* meaning to information complexity, namely, one which is grounded in reality (in fact, it was recently shown that this characterization is a "sharp threshold", see Theorem 4.8).

Theorem 1.1 and some of the additional results mentioned in this survey, provide a strong evidence that information theory is the "right" tool for studying interactive communication problems. One general benefits of information theory in addressing communication problems is that it provides a set of simple yet powerful tools for reasoning about transmission problems and more broadly about quantifying relationships between interdependent random variables and conditional events. Tools that include mutual information, the chain rule, and the data processing inequality [CT91]. Another, arguably most important benefit, is the *additivity* of information complexity under composition of independent tasks (Lemma 3.1 below). This is the main reason that information theory, unlike other analytic or combinatorial methods, is apt to give *exact* bounds on rates and capacities

---

[1]Indeed, the "$\leq$" direction of this proof gives a protocol with overall success only $\approx (1 - \varepsilon)^n$ on all n copies, see [BR11].

(such as Shannon's noiseless coding theorem and Theorem 1.1). It is this benefit that has been primarily used in prior works (which are beyond the scope of this survey) involving information-theoretic applications in communication complexity, circuit complexity, streaming, machine learning and privacy ([CSWY01, LS10, CKS03, BYJKS04, JKR09, BGPW13, ZDJW13, WZ14] to mention a few).

One caveat is that mathematically striking characterizations such as the noiseless coding theorem only become possible in the limit, where the number of independent samples transmitted over the channel (i.e., the block-length) grows to infinity. One exception is Huffman's *"one-shot"* compression scheme (aka Huffman coding, [Huf52]), which shows that the expected number of bits $C(M)$ needed to transmit a *single sample* from M, is very close (but not equal!) to the optimal rate

$$H(M) \leq C(M) \leq H(M) + 1. \tag{2}$$

Huffman's theorem of course implies Shannon's theorem (since entropy is additive over independent variables), but is in fact much stronger, as it asserts that the optimal transmission rate can be (essentially) achieved using much a smaller block length. Indeed, what happens for small block lengths is of importance for both practical and theoretical reasons, and it will be even more so in the interactive regime. While Theorem 1.1 provides an interactive analogue of Shannon's theorem, an intriguing question is whether an interactive analogue of Huffman's "one-shot" compression scheme exists. When the number of communication rounds of a protocol is small (constant), compressing it can morally[2] be done by applying Huffman's compression scheme to each round of the protocol, since (2) would entail at most a constant overhead in communication. However, when the number of rounds is huge compared to the overall information revealed by the protocol (e.g., when each round reveals $\ll 1$ bits of information), this approach is doomed to fail, as it would "spend" at least 1 bit of communication per round. Circumventing this major obstacle and the important implications of this (unsettled) question to the direct sum and product conjectures are extensively discussed in Sections 4 and 5.

Due to space constraints, this survey is primarily focused on the above relationship between information and communication complexity. As mentioned above, information complexity has recently found many more exciting applications in complexity theory – to interactive coding, streaming lower bounds, extension complexity and multiparty communication complexity (e.g., [BYJKS04, BM12, BP13, BEO+13]). Such applications are beyond the scope of this survey.

Organization We begin with a brief introduction to information complexity and some of its main properties (Section 2). In Section 4 we give an overview of the direct sum and direct product conjectures and their relationship to interactive compression, in light of recent developments in the field. Section 5 describes state-of-the-art interactive compression schemes. We conclude with several natural open problems in Section 6. In an effort to keep this survey as readable and self-contained as possible, we shall sometimes be loose on technical formulations, often ignoring constant and technical details which are not essential to the reader.

---

[2]This is not accurate, since unlike the one-way transmission setting, in this setting the receiver has "side information" about the transmitted message, e.g., when Bob sends the second message of the protocol, Alice has a prior distribution on this message conditioned on her input X and the first message of the protocol M_1 which she sent before. Nevertheless, using ideas from rejection sampling, such simulation is possible in the "one-shot" regime with $O(1)$ communication overhead per message [HJMR07, BR11].

2 Model and Preliminaries

The following background contains basic definitions and notations used throughout this survey. For a more detailed overview of communication and information complexity, we refer the reader to an excellent monograph by Braverman [Bra12].

For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, a distribution μ over $\mathcal{X} \times \mathcal{Y}$, and a parameter $\varepsilon > 0$, $\mathsf{D}_\mu(f, \varepsilon)$ denotes the communication complexity of the cheapest deterministic protocol computing f on inputs sampled according to μ with error ε. $\mathsf{R}(f, \varepsilon)$ denotes the cost of the best *randomized* public coin protocol which computes f with error at most ε, for *all* possible inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$. When measuring the communication cost of a particular protocol π, we sometimes use the notation $\|\pi\|$ for brevity. Essentially all results in this survey are proven in the former *distributional* communication model (since information complexity is meaningless without a prior distribution on inputs), but most lower bounds below can be extended to the randomized model via Yao's minimax theorem. For the sake of concreteness, all of the results in this article are stated for (total) functions, though most of them apply to partial functions and relations as well.

2.1 Information Theory

Proofs of the claims below and a broader introduction to information theory can be found in [CT91]. The most basic concept in information theory is Shannon's entropy, which informally captures how predictable a random variable is:

Definition 2.1 (Entropy). *The* entropy *of a random variable A is* $H(A) := \sum_a \Pr[A = a] \log(1/\Pr[A = a])$. *The* conditional entropy $H(A|B)$ *is defined as* $\mathbb{E}_{b \sim B}[H(A|B = b)]$.

A key measure in this article is the *Mutual Information* between two random variables, which quantifies the amount of correlation between them:

Definition 2.2 (Mutual Information). *The* mutual information *between two random variables A, B, denoted $I(A; B)$ is defined to be the quantity $H(A) - H(A|B) = H(B) - H(B|A)$. The* conditional mutual information $I(A; B|C)$ *is* $H(A|C) - H(A|BC)$.

A related distance measure between *distributions* is the *Kullback-Leibler* (KL) divergence

$$\mathbb{D}(p\|q) := \sum_x p(x) \log \frac{p(x)}{q(x)} = \mathbb{E}_{x \sim p}\left[\log \frac{p(x)}{q(x)}\right].$$

We shall sometimes abuse the notation and write $\mathbb{D}(A|c\|B|c)$ to denote the KL divergence between the associated distributions of the random variables $(A|C = c)$ and $(B|C = c)$. The following connection between divergence and mutual information is well known:

Lemma 2.3 (Mutual information in terms of Divergence).

$$I(A; B|C) = \mathbb{E}_{b,c}[\mathbb{D}(A|bc\|A|c)] = \mathbb{E}_{a,c}[\mathbb{D}(B|ac\|B|c)].$$

Intuitively, the above equation asserts that, if the mutual information between A and B (conditioned on C) is large, then the distribution of $(A|c)$ is "far" from $(A|bc)$ for average values of b, c (this captures the fact that the "additional information" B provides on A given C is large). One of the most useful properties of Mutual Information and KL Divergence is the chain rule:

Lemma 2.4 (Chain Rule)**.** *Let A, B, C, D be four random variables in the same probability space. Then*

$$I(AB; C|D) = I(A; C|D) + I(B; C|AD)$$

$$= \mathbb{E}_{c,d}\left[\mathbb{D}\left(A|cd\|A|d\right)\right] + \mathbb{E}_{a,c,d}\left[\mathbb{D}\left(B|acd\|A|ad\right)\right].$$

Lemma 2.5 (Conditioning on independent variables does not decrease information)**.** *Let A, B, C, D be four random variables in the same probability space. If A and D are conditionally independent given C, then it holds that $I(A; B|C) \leq I(A; B|CD)$.*

Proof. We apply the chain rule for mutual information twice. On one hand, we have $I(A; BD|C) = I(A; B|C) + I(A; D|CB) \geq I(A; B|C)$ since mutual information is nonnegative. On the other hand, $I(A; BD|C) = I(A; D|C) + I(A; B|CD) = I(A; B|CD)$ since $I(A; D|C) = 0$ by the independence assumption on A and D. Combining both equations completes the proof. $\square$

Throughout this article, we denote by $|p - q|$ the *total variation* distance between the distributions p and q. Pinsker's inequality bounds statistical distance in terms of the KL divergence. It will be useful for analysis of the interactive compression schemes in Section 5.

Lemma 2.6 (Pinsker's inequality)**.** $|p - q|^2 \leq \frac{1}{2} \cdot \mathbb{D}\left(p\|q\right)$.

2.2 Interactive Information complexity

Given a communication protocol π, $\pi(x, y)$ denotes the concatenation of the public randomness with all the messages that are sent during the execution of π (for information purposes, this is without loss of generality, since the public string R conveys no information about the inputs). We call this the *transcript* of the protocol. When referring to the random variable denoting the transcript, rather than a specific transcript, we will use the notation $\Pi(x, y)$ — or simply Π when x and y are clear from the context.

Definition 2.7 (Internal Information Cost [CSWY01, BBCR10])**.** *The (internal) information cost of a protocol over inputs drawn from a distribution μ on $\mathcal{X} \times \mathcal{Y}$, is given by:*

$$\mathsf{IC}_\mu(\pi) := I(\Pi; X|Y) + I(\Pi; Y|X). \tag{3}$$

Intuitively, the definition in (3) captures how much additional information the two parties learn about each other's inputs by observing the protocol's transcript. For example, the information cost of the trivial protocol in which Alice and Bob simply exchange their inputs, is simply the sum of their conditional marginal entropies $H(X|Y) + H(Y|X)$ (notice that, in contrast, the *communication* cost of this protocol is $|X| + |Y|$ which can be arbitrarily larger than the former quantity).

Another information measure which makes sense at certain contexts is the *external* information cost of a protocol, $\mathsf{IC}_\mu^{\mathsf{ext}}(\pi) := I(\Pi; XY)$, which captures what an *external* observer learns on average about both player's inputs by observing the transcript of π. This quantity will be of minor interest in this survey (though it plays a central role in many applications). The external information cost of a protocol is always at least as large as its (internal) information cost, since intuitively an external observer is "more ignorant" to begin with. We remark that when μ is a *product* distribution, then $\mathsf{IC}_\mu^{\mathsf{ext}}(\pi) = \mathsf{IC}_\mu(\pi)$ (see, e.g., [Bra12]).

One can now define the *information complexity* of a function f with respect to μ and error ε as the least amount of information the players need to reveal to each other in order to compute f with error at most ε:

Definition 2.8. *The Information Complexity* of f with respect to μ (and error ε) is

$$\mathsf{IC}_\mu(f, \varepsilon) := \inf_{\pi:\, \Pr_\mu[\pi(x,y) \neq f(x,y)] \leq \varepsilon} \mathsf{IC}_\mu(\pi).$$

What is the relationship between the information and communication complexity of f? This question is at the core of this survey. The answer to one direction is easy: Since one bit of communication can never reveal more than one bit of information, the communication cost of any protocol is always an upper bound on its information cost over *any* distribution μ:

Lemma 2.9 ([BR11]). *For any distribution μ, $\mathsf{IC}_\mu(\pi) \leq \|\pi\|$.*

The answer to the other direction, namely, whether any protocol can be compressed to roughly its information cost, will be partially given in the remainder of this article.

2.3   The role of private randomness in information complexity

A subtle but vital issue when dealing with information complexity, is understanding the role of private vs. public randomness. In public-coin communication complexity, one often ignores the usage of private coins in a protocol, as they can always be simulated by public coins. When dealing with *information complexity*, the situation is somewhat the opposite: Public coins are essentially a redundant resource (as it can be easily shown via the chain rule that $\mathsf{IC}_\mu(\pi) = \mathbb{E}_R[\mathsf{IC}_\mu(\pi_R)]$), while the usage of private coins is crucial for minimizing the information cost, and fixing these coins is prohibitive (once again, for communication purposes in the distributional model, one may always fix the entire randomness of the protocol, via the averaging principle). To illustrate this point, consider the simple example where in the protocol π, Alice sends Bob her 1-bit input $X \sim Ber(1/2)$, XORed with some random bit Z. If Z is private, Alice's message clearly reveals 0 bits of information to Bob about X. However, for any fixing of Z, this message would reveal an entire bit(!). The general intuition is that a protocol with low information cost would reveal information about the player's inputs in a "careful manner", and the usage of private coins serves to "conceal" parts of their inputs. Indeed, it was recently shown that the restriction to public coins may cause an exponential blowup in the information revealed compared to private-coin protocols ([GKR14, BMY14]). In fact, we shall see in Section 4 that quantifying this gap between public-coin and private-coin information complexity is tightly related to the question of interactive compression.

For the remainder of this article, communication protocols π are therefore assumed to use private coins (and therefore such protocols are randomized even conditioned on the inputs x, y and R), and it is crucial that the information cost $\mathsf{IC}_\mu(\pi) = I(\Pi; X | YR) + I(\Pi; Y | XR)$ is measured conditioned on the *public* randomness R, but never on the private coins of π.

3   Additivity of Information Complexity

Perhaps the single most remarkable property of information complexity is that it is a fully additive measure over composition of tasks. This property is what primarily makes information complexity a natural "relaxation" for addressing direct sum and product theorems. The main ingredient of the following lemma appeared first in the works of [Raz08, Raz98] and more explicitly in [BBCR10, BR11, Bra12]. In the following, f^n denotes the function that maps the tuple $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$ to $(f(x_1, y_1), \ldots, f(x_n, y_n))$.

Lemma 3.1 (Additivity of Information Complexity). $\mathsf{IC}_{\mu^n}(f^n, \varepsilon) = n \cdot \mathsf{IC}_\mu(f, \varepsilon)$.

Proof. The ($\leq$) direction of the lemma is easy, and follows from a simple argument that applies the single-copy optimal protocol independently to each copy of f^n, with independent randomness. We leave the simple analysis of this protocol as an exercise to the reader.

The ($\geq$) direction is the main challenge. Will will prove it in a contra-positive fashion: Let Π be an ε-error protocol for f^n, such that $\mathsf{IC}_{\mu^n}(\Pi) = I$ (recall that here ε denotes the per-copy error of Π in computing $f(x_i, y_i)$). We shall use Π to produce a *single-copy* protocol for f whose information cost is $\leq I/n$, which would complete the proof. The guiding intuition for this is that Π should reveal I/n bits of information about an average coordinate.

To formalize this intuition, let $(x, y) \sim \mu$, and denote $\mathbf{X} := X_1 \ldots X_n$, $X_{\leq i} := X_1 \ldots X_i$ and $X_{-i} := X_1 \ldots X_{i-1}, X_{i+1}, \ldots, X_n$, and similarly for $\mathbf{Y}, Y_{\leq i}, Y_{-i}$. A natural idea is for Alice and Bob to "embed" their respective inputs (x, y) to a (publicly chosen) random coordinate $i \in [n]$ of Π, and execute Π. However, Π is defined over n input copies, so in order to execute it, the players need to somehow "fill in" the rest $(n-1)$ coordinates, each according to μ. How should this step be done? The first attempt is for Alice and Bob to try and complete X_{-i}, Y_{-i} privately. This approach fails if μ is a non-product distribution, since there's no way the players can sample X and Y privately, such that $(X, Y) \sim \mu$ if μ correlates the inputs. The other extreme – sampling X_{-i}, Y_{-i} using public randomness only – would resolve the aforementioned correctness issue, but might leak too much information: An instructive example to consider is where, in the first message of Π, Alice sends Bob the XOR of the n bits of her uniform input X: $M = X_1 \oplus X_2 \oplus \ldots \oplus X_n$. Conditioned on X_{-i}, Y_{-i}, M reveals 1 bit of information about X_i to Bob, while we want to argue that in this case, only $1/n$ bits are revealed about X_i. So this approach reveals too much information.

It turns out that the "right" way of breaking the dependence across the coordinates is to use a combination of public and private randomness. Let us define, for each $i \in [n]$, the public random variable

$$R_i := X_{<i}, Y_{>i}.$$

Note that given R_i, Alice can complete all her missing inputs $X_{>i}$ *privately* according to μ, and Bob can do the same for $Y_{<i}$. Let us denote by $\theta(x, y, i, R_i)$ the protocol transcript produced by running $\Pi(X_1, \ldots, X_{i-1}, x, X_{i+1}, \ldots, X_n$, $Y_1, \ldots, Y_{i-1}, y, Y_{i+1}, \ldots, Y_n)$ and outputting its answer on the i'th coordinate. Let $\Theta(x, y)$ be the protocol obtained by running $\theta(x, y, i, R_i)$ on a uniformly selected $i \in [n]$.

By definition, Π computes f^n with a *per-copy* error of ε, and thus in particular $\Theta(x, y) = f(x, y)$ with probability $\geq 1 - \varepsilon$. To analyze the information cost of Θ, we write:

$$I(\Theta; x|y) = \mathbb{E}_{i, R_i}[I(\theta; x|y, R_i)] = \sum_{i=1}^n \frac{1}{n} \cdot I(\Pi; X_i \mid Y_i, R_i)$$

$$= \frac{1}{n} \sum_{i=1}^n I(\Pi; X_i \mid Y_i, X_{<i} Y_{>i}) = \frac{1}{n} \sum_{i=1}^n I(\Pi; X_i \mid X_{<i} Y_{\geq i})$$

$$\leq \frac{1}{n} \sum_{i=1}^n I(\Pi; X_i \mid X_{<i} \mathbf{Y}) = \frac{1}{n} \cdot I(\Pi; \mathbf{X} \mid \mathbf{Y}),$$

where the inequality follows from Lemma 2.5, since $I(Y_{<i}; X_i | X_{<i}) = 0$ by construction, and the last transition is by the chain rule for mutual information. By symmetry of construction, an analogous

argument shows that $I(\Theta; y|x) \le I(\Pi; \mathbf{Y} \mid \mathbf{X})/n$, and combining these facts gives

$$\mathsf{IC}_\mu(\Theta) \le \frac{1}{n} \left(I(\Pi; \mathbf{X} \mid \mathbf{Y}) + I(\Pi; \mathbf{Y} \mid \mathbf{X}) \right) = \frac{I}{n}. \tag{4}$$

$\square$

4   Direct Sum, Product, and the Interactive Compression Problem

Direct sum and direct product theorems assert a lower bound on the complexity of solving n copies of a problem in parallel, in terms of the cost of a single copy. Let f^n denote the problem of computing n simultaneous instances of the function f (in some arbitrary computational model for now), and $C(f)$ denote the cost of solving a single copy of f. The obvious solution to f^n is to apply the single-copy optimal solution n times sequentially and independently to each coordinate, yielding a linear scaling of the resources, so clearly $C(f^n) \le n \cdot C(f)$. The *strong direct sum* conjecture postulates that this naive solution is essentially optimal. In the context of randomized communication complexity, the strong direct sum conjecture informally asks whether it is true that for any function f and input distribution μ,

$$\mathsf{D}_{\mu^n}(f^n, \varepsilon) =^? \Omega(n) \cdot \mathsf{D}_\mu(f, \varepsilon). \tag{5}$$

More generally, direct sum theorems aim to give an (ideally linear in n, but possibly weaker) lower bound on the communication required for computing f^n with some *constant overall* error $\varepsilon > 0$ in terms of the cost of computing a single copy of f with the same (or comparable) fixed error.

A *direct product* theorem further asserts that unless sufficient resources are provided, the probability of successfully computing all n copies of f will be exponentially small, potentially as low as $(1 - \varepsilon)^{\Omega(n)}$. This is intuitively plausible, since the naive solution which applies the best (ε-error) protocol for one copy of f independently to each of the n coordinates, would indeed succeed in solving f^n with probability $(1 - \varepsilon)^n$. Is this naive solution optimal?

To make this more precise, let us denote by $\mathsf{suc}(\mu, f, C)$ the maximum success probability of a protocol with communication complexity $\le C$ in computing f under input distribution μ. A direct product theorem asserts that any protocol attempting to solve f^n (under μ^n) using some number T of communication bits (ideally $T = \Omega(n \cdot C)$), will succeed only with exponentially small probability: $\mathsf{suc}(\mu^n, f^n, T) \lesssim (1 - \varepsilon)^{\Omega(n)}$. Informally, the strong direct product question asks whether

$$\mathsf{suc}(\mu^n, f^n, o(n \cdot C)) \lesssim^? (\mathsf{suc}(\mu, f, C))^{\Omega(n)}. \tag{6}$$

Note that (6) in particular implies (5) when setting $C = \mathsf{D}_\mu(f, \varepsilon)$. Classic examples of direct product results in complexity theory are Raz's Parallel Repetition Theorem [Raz98, Rao08] and Yao's XOR Lemma [Yao82] (For more examples and a broader overview of the rich history of direct sum and product theorems see [JPY12] and references therein). The value of such results to computational complexity is clear: direct sum and product theorems, together with a lower bound on the (easier-to-reason-about) "primitive" problem, yield a lower bound on the composite problem in a "black-box" fashion (a method also known as *hardness amplification*). For example, the Karchmer-Raz-Wigderson approach for separating $\mathbf{P}$ from $\mathbf{NC}^1$ can be completed via a (still open) direct sum conjecture for Boolean formulas [KRW95] (after more than a decade, some progress on this conjecture was recently made using information-complexity machinery [GMWW14]). Other

fields in which direct sums and products have played a central role in proving tight lower bounds are streaming [BYJKS04, ST13, MWY13, GO13] and distributed computing [HRVZ13].

Can we always hope for such strong lower bounds to hold? It turns out that the validity of these conjectures highly depends on the underlying computational model, and the short answer is no.[3] In the communication complexity model, this question has had a long history and was answered positively for several restricted models of communication [Kla10, Sha03, LSS08, She12, JPY12, MWY13, PRW97]. Interestingly, in the *determistic* communication complexity model, Feder et al. [FKNN95] showed that

$$\mathsf{D}(f^n) \geq n \cdot \Omega\left(\sqrt{\mathsf{D}(f)}\right)$$

for any two-party Boolean function f (where $\mathsf{D}(f)$ stands for the deterministic communication complexity of f), but this proof completely breaks when protocols are allowed to err. Indeed, in the randomized communication model, there is a tight connection between the direct sum question for the function f and its information complexity. By now, this should come as no surprise: Theorem 1.1 asserts that, for large enough n, the communication complexity of f^n scales linearly with the (single-copy) information cost of f, i.e. $\mathsf{D}_{\mu^n}(f^n, \varepsilon) = \Theta\left(n \cdot \mathsf{IC}_\mu(f, \varepsilon)\right)$, and hence the strong direct sum question (5) boils down to understanding the relationship between the single-copy measures $\mathsf{D}_\mu(f, \varepsilon)$ and $\mathsf{IC}_\mu(f, \varepsilon)$. Indeed, it can be formally shown ([BR11]) that the direct sum problem is equivalent [4] to the following problem of "one-shot" compression of interactive protocols:

Problem 4.1 (Interactive compression problem, [BBCR10]). *Given a protocol π over inputs $x, y \sim \mu$, with $\|\pi\| = C, \mathsf{IC}_\mu(\pi) = I$, what is the smallest amount of communication of a protocol τ which (approximately) simulates π (i.e., $\exists\, g \text{ s.t } |g(\tau(x,y)) - \pi(x,y)|_1 \leq \delta$ for a small constant δ)?*

In particular, if one could compress any protocol into $O(I)$ bits, this would have shown that $\mathsf{D}_\mu(f, \varepsilon) = O\left(\mathsf{IC}_\mu(f, \varepsilon)\right)$ which would in turn imply the strong direct sum conjecture. In fact, the additivity of information cost (Lemma 3.1 from Section 3) implies the following general quantitative relationship between (possibly weaker) interactive compression results and direct sum theorems in communication complexity:

Proposition 4.2 (One-Shot Compression implies Direct Sum). *Suppose that for any $\delta > 0$ and any given protocol π for which $\mathsf{IC}_\mu(\pi) = I$, $\|\pi\| = C$, there is a compression scheme that δ-simulates[5] π using $g_\delta(I, C)$ bits of communication. Then*

$$g_\delta\left(\frac{\mathsf{D}_{\mu^n}(f^n, \varepsilon)}{n}, \mathsf{D}_{\mu^n}(f^n, \varepsilon)\right) \geq \mathsf{D}_\mu(f, \varepsilon + \delta).$$

---

[3]In the context of circuit complexity, for example, this conjecture fails (at least in its strongest form): Multiplying an $n \times n$ matrix by a (worst case) n-dimensional vector requires n^2 operations, while (deterministic) multiplication of n different vectors by the same matrix amounts to matrix-multiplication of two $n \times n$ matrices, which can be done in $n^{2.37} \ll n^3$ operations [Wil12].

[4]The exact equivalence of the direct sum conjecture and Problem 4.1 holds for *relations* (Theorem 6.6 in [BR11]). For total functions, one could argue that the requirement in Problem 4.1 is too harsh as it requires simulation of the entire transcript of the protocol, while in the direct sum context for functions we are merely interested in the output of f. However, all known compression protocols satisfy the stronger requirement and no separation is known between those techniques.

[5]The simulation here is in an internal sense, namely, Alice and Bob should be able to reconstruct the transcript of the original protocol (up to a small error), based on public randomness and their own private inputs. See [BRWY12] for the precise definition and the (subtle) role it plays in context of direct product theorems.

Proof. Let Π be an optimal n-fold protocol for f^n under μ^n with per-copy error ε, i.e., $\|\Pi\| = \mathsf{D}_{\mu^n}(f^n, \varepsilon) := C_n$. By Lemma 3.1 (equation (4)), there is a single-copy ε-error protocol θ for computing $f(x,y)$ under μ, whose information cost is at most $\mathsf{IC}_{\mu^n}(\Pi)/n \le C_n/n$ (since communication always upper bounds information). By assumption of the claim, θ can now be δ-simulated using $g_\delta(C_n/n, C_n)$ communication, so as to produce a single-copy protocol with error $\le \varepsilon + \delta$ for f, and therefore $\mathsf{D}_\mu(f, \varepsilon + \delta) \le g_\delta(C_n/n \, , \, C_n)$. $\square$

The first general interactive compression result was proposed in the seminal work of Barak, Braverman, Chen and Rao [BBCR10], who showed that any protocol π can be δ-simulated using $g_\delta(I, C) = \tilde{O}_\delta(\sqrt{C \cdot I})$ communication (we prove this result in Section 5.1). Plugging this compression result into Proposition 4.2, this yields the following weaker direct sum theorem:

Theorem 4.3 (Weak Direct Sum, [BBCR10])*. For every Boolean function f, distribution μ, and any positive constant $\delta > 0$,*

$$\mathsf{D}_{\mu^n}(f^n, \varepsilon) \ge \tilde{\Omega}(\sqrt{n} \cdot \mathsf{D}_\mu(f, \varepsilon + \delta)).$$

Later, Braverman [Bra12] showed that it is always possible to simulate π using $2^{O_\delta(I)}$ bits of communication. This result is still far from ideal compression ($O(I)$ bits), but it is nevertheless appealing as it show that any protocol can be simulated using amount of communication which depends solely on its information cost, but *independent* of its original communication which may have been arbitrarily larger (we prove this result in Section 5.2). Notice that the last two compression results are indeed incomparable, since the communication of π could be much larger than its information complexity (e.g., $C \ge 2^{2^{2^I}}$). The current state of the art for the *general* interactive compression problem can be therefore summarized as follows: Any protocol with communication C and information cost I can be compressed to

$$g_\delta(I, C) \le \min \left\{ 2^{O_\delta(I)} \, , \, \tilde{O}_\delta(\sqrt{I \cdot C}) \right\} \tag{7}$$

bits of communication.

The above results may seem as a plausible evidence that it is in fact possible to compress general interactive protocols all the way down to $O(I)$ bits. Unfortunately, this task turns out to be too ambitious: In a recent breakthrough result, Ganor, Kol and Raz [GKR14] proved the following lower bound on the communication of any compression scheme:

$$g_\delta(I, C) \ge \max \left\{ 2^{\Omega(I)} \, , \, \tilde{\Omega}(I \cdot \log C) \right\}. \tag{8}$$

More specifically, they exhibit a Boolean function f which can be solved using a protocol with information cost I, but cannot be simulated by a protocol π' with communication cost $< 2^{\Omega(I)}$ (a simplified construction and proof was very recently obtained by Rao and Sinha [RS15]). Since the *communication* of the low information protocol they exhibit is $\sim 2^{2^I}$, this also rules out a compression to $I \cdot o(\log C)$, or else such compression would have produced a too good to be true ($2^{o(I)}$ communication) protocol. The margin of this text is too narrow to contain the proof of this separation result, but it is noteworthy that proving it was particularly challenging: It was shown that essentially all previously known techniques for proving communication lower bounds apply to information complexity as well [BW12, KLL$^+$12], and hence could not be used to separate

information complexity and communication complexity. Using (the reverse direction of) Proposition 4.2 (see Theorem 6.6 in [BR11]), the compression lower bound in (8) refutes the strongest possible direct sum (5), but leaves open the following gap

$$\tilde{\Omega}_\delta\left(\sqrt{n}\right) \;\leq\; \min_f \frac{\mathsf{D}_{\mu^n}(f^n, \varepsilon)}{\mathsf{D}_\mu(f, \varepsilon + \delta)} \;\leq\; O\left(\frac{n}{\log n}\right). \tag{9}$$

Notice that this still leaves the direct sum conjecture for randomized communication complexity wide open: It is still conceivable that improved compression to $g_\delta(I, C) = I \cdot C^{o(1)}$ is in fact possible, and the quest to beat the compression scheme of [BBCR10] remains unsettled.[6]

Despite the lack of progress in the general regime, several works showed that it is in fact possible to obtain near-optimal compression results in restricted models of communication: When the input distribution μ is a *product distribution* (x and y are independent), [BBCR10] show a near-optimal compression result, namely that π can be compressed into $O(I \cdot polylog(C))$ bits.[7] Once again, using Proposition 4.2 this yields the following direct sum theorem:

Theorem 4.4 ([BBCR10]). *For every product distribution μ and any $\delta > 0$,*

$$\mathsf{D}_{\mu^n}(f^n, \varepsilon) = \tilde{\Omega}(n \cdot \mathsf{D}_\mu(f, \varepsilon + \delta)).$$

Improved compression results were also proven for *public-coin protocols* (under arbitrary distributions) [BBK+13, BMY14], and for bounded-round protocols, leading to near-optimal direct sum theorems in corresponding communication models. We summarize these results in Table 1.

Reference	Regime	Communication Complexity
[HJMR07]	r-round protocols, product distributions[7]	$I + O(r)$
[BR11, BRWY13]	r-round protocols	$I + O\left(\sqrt{r \cdot I}\right) + O(r \log 1/\delta)$
[BMY14] (improved [BBK+13])	Public coin protocols	$O(I^2 \cdot \log\log(C)/\delta^2)$
[BBCR10]	Product distributions[7]	$O(I \cdot \mathrm{poly}\log(C)/\delta)$
[Bra12, BBCR10]	**General protocols**	$\min\{2^{O(I/\delta)},\ O(\sqrt{I \cdot C} \cdot \log(C)/\delta)\}$
[GKR14, RS15]	**Best lower bound**	$\max\{2^{\Omega(I)},\ \Omega(I \cdot \log(C))$

Table 1: Best to date compression schemes, for various regimes. Notice that in the general regime (last two columns), in terms of dependence on the original communication C, the gap is still very large ($\Omega(\log C)$ vs. $\tilde{O}(C^{1/2})$).

4.1 Harder, better, stronger: From direct sum to direct product

As noted above, direct sum theorems such as Theorems 1.1, 4.3 and 4.4 are weak in that they merely assert that attempting to solve n independent copies of f using less than some number T of resources, would fail with some *constant* overall probability (($\mathsf{suc}(\mu^n, f^n, o(\sqrt{n \cdot C})) \leq \varepsilon$ in the general case, and $\mathsf{suc}(\mu^n, f^n, o(n \cdot C)) \leq \varepsilon$ in the product case, where $C = \mathsf{D}_\mu(f, \varepsilon)$). This is somewhat

---

[6]Ramamoorthy and Rao [RR15] recently showed that BBCR's compression scheme can be improved when the underlying communication protocol is *asymmetric*, i,e., when Alice reveals much more information than Bob.

[7] These compression results in fact hold for general (non-product) distributions as well, when compression is with respect to I^{ext}, the external information cost of the original protocol π (which may be significantly larger than I).

unsatisfactory, since the naive solution that applies the single-copy optimal protocol independently to each copy has only exponentially small success probability in solving all copies correctly. Indeed, some of the most important applications of hardness amplification require amplifying the error parameter (e.g., the usage of parallel repetition in the context of the PCP theorem).

As mentioned before, many direct product theorems were proven in limited communication models (e.g. Shaltiel's Discrepancy bound [Sha03, LSS08] which was extended to the generalized discrepancy bound [She12], Parnafes, Raz, and Wigderson's theorem for communication forests [PRW97], Jain's theorem [Jai11] for simultaneous communication and [JY12]'s direct product in terms of the "smooth rectangle bound" to mention a few), but none of them applied to general functions and communication protocols. In a recent breakthrough work, Jain, Pereszlényi and Yao used an information-complexity based approach to prove a strong direct product theorem for any function (relation) in the bounded-round communication model.

Theorem 4.5 ([JPY12]). *Let* $\mathsf{suc}_r(\mu, f, C)$ *denote the largest success probability of an* r-round *protocol with communication at most* C, *and suppose that* $\mathsf{suc}_r(\mu, f, C) \le \frac{2}{3}$. *If* $T = o\left(\left(\frac{C}{r} - r\right) \cdot n\right)$, *then* $\mathsf{suc}_r(\mu^n, f^n, T) \le \exp(-\Omega(n/r^2))$.

This theorem can be essentially viewed as a sharpening of the direct sum theorem of Braverman and Rao for bounded-round communication [BR11]. This bound was later improved by Braverman et. al who showed that $\mathsf{suc}_{r/7}(\mu^n, f^n, o((C - r \log r) \cdot n)) \le \exp(-\Omega(n))$, thus settling the strong direct product conjecture in the bounded round regime. The followup work of [BRWY12] took this approach one step further, obtaining the first direct product theorem for *unbounded-round* randomized communication complexity, thus sharpening the direct sum results of [BBCR10].

Theorem 4.6 ([BRWY12], informally stated). *For any two-party function* f *and distribution* μ *such that* $\mathsf{suc}(\mu, f, C) \le \frac{2}{3}$, *the following holds:*

- *If* $T \log^{3/2} T = o(C \cdot \sqrt{n})$, *then* $\mathsf{suc}(\mu^n, f^n, T) \le \exp\left(-\Omega(n)\right)$.

- *If* μ *is a product distribution, and* $T \log^2 T = o(C \cdot n)$, *then* $\mathsf{suc}(\mu^n, f^n, T) \le \exp(-\Omega(n))$.

One appealing corollary of the second proposition is that, under the *uniform* distribution, two-party interactive computation cannot be "parallelized", in the sense that the best protocol for solving f^n (up to polylogarithmic factors), is to apply the single-coordinate optimal protocol independently to each copy, which almost matches the above parameters.

The high-level intuition behind the proofs of Theorems 4.5 and 4.6 follows the direct sum approach of [BBCR10] (Proposition 4.2 above): Suppose, towards contradiction, that the success probability of an n-fold protocol using T bits of communication in computing f^n under μ^n is larger than, say, $\exp(-n/100)$. We would like to "embed" a *single-copy* $(x, y) \sim \mu$ into this n-fold protocol, thereby producing a *low information* protocol ($\le T/n$ bits), and then use known compression schemes to compress this protocol, eventually obtaining a protocol with communication ($< C$), and a too-good-to-be-true success probability ($> 2/3$), contradicting the assumption that $\mathsf{suc}(\mu, f, C) \le \frac{2}{3}$. The main problem with employing the [BBCR10] approach and embedding a single-copy (x, y) into π using the sampling argument in Lemma 3.1, is that it would produce a single-copy protocol $\theta(x, y)$ whose success probability is no better than that of π ($\exp(-n/100)$) while we need to produce a single-copy protocol with success $> 2/3$ in order to achieve the above contradiction.

Circumventing this major obstacle is inspired by the idea of repeated conditioning which first appeared the parallel repetition theorem [Raz98]: Let $\mathcal{W}$ be the event that π correctly computes f^n, and $\mathcal{W}_i$ denote the event that the protocol correctly computes the i'th copy $f(x_i, y_i)$. Let $\pi(\mathcal{W})$ denote the probability of $\mathcal{W}$, and $\pi(\mathcal{W}_i|\mathcal{W})$ denote the conditional probability of the event $\mathcal{W}_i$ given $\mathcal{W}$ (clearly, $\pi(\mathcal{W}_i|\mathcal{W}) = 1$). The idea is to show that if $\pi(\mathcal{W}) \geq \exp(-n/100)$ and $\|\pi\| \ll T$ (for the appropriate choice of T which is determined by the best compression scheme), then $(1/n)\sum_{i=1}^{n} \pi(\mathcal{W}_i|\mathcal{W}) < 1$, which is a contradiction. In other words, if one could simulate the message distribution of the conditional distribution $(\pi|\mathcal{W})_i$ (rather than the distribution of $\pi(x_i, y_i)$) using a low information protocol, then (via compression) one would obtain a protocol $\theta(x_i, y_i)$ with *constant* success probability, as desired.

The guiding intuition for why this approach makes sense, is that conditioning a random variable on a "large" event $\mathcal{W}$ does not change its original distribution too much:

$$\mathbb{D}\left(X_1 Y_1, X_2 Y_2, \ldots, X_n Y_n \mid \mathcal{W} \| X_1 Y_1, X_2 Y_2, \ldots, X_n Y_n\right) = \mathbb{D}\left(\mathbf{XY}|\mathcal{W}\|\mathbf{XY}\right)$$

$$= \mathbb{E}\left[\log \frac{\pi(\mathbf{XY}|\mathcal{W})}{\pi(\mathbf{XY})}\right] \leq \mathbb{E}\left[\log \frac{\pi(\mathbf{XY})}{\pi(\mathbf{XY})\pi(\mathcal{W})}\right] = \frac{1}{\log(\pi(\mathcal{W}))} \leq \frac{n}{100}$$

since $\pi(\mathcal{W}) \geq \exp(-n/100)$, which means (by the chain rule and independence of the n copies) that the distribution of an *average* input pair (X_i, Y_i) conditioned on $\mathcal{W}$ is $(1/100)$-close to its original distribution μ, and thus implies that at least the *inputs* to the "protocol" $(\pi|\mathcal{W})_i$ can be approximately sampled correctly (using correlated sampling [Hol07]). The heart of the problem, however, is that $(\pi|\mathcal{W})_i$ is no longer a communication protocol. To see why, consider the simple protocol π in which Alice simply "guesses" Bob's bit x, and $\mathcal{W}$ being the event that her guess is correct. Then simulating $(\pi|\mathcal{W})$ requires Alice to know Bob's input y, which Alice doesn't have! This example shows that it is impossible to simulate the message distribution of $(\pi|\mathcal{W})_i$ exactly. The main contribution of Theorem 4.6 (and Theorem 4.5 in the bounded-round regime) is showing that it is nevertheless possible to *approximate* this conditional distribution using an actual communication protocol, which is statistically close to a low-information protocol:

Lemma 4.7 (Claims 26 and 27 from [BRWY12], informally stated). *There is a protocol θ taking inputs $x, y \sim \mu$ so that the following holds:*

- *θ publicly chooses a uniform $i \in [n]$ independent of x, y, and R_i which is part of the input to π (intuitively, R_i determines the "missing" inputs x_{-i}, y_{-i} of π as in Lemma 3.1).*

- *$\mathbb{E}_i\left[|(\theta|R_i) - (\pi|R_i \mathcal{W})_i|\right] \leq 1/10$ (that is, θ is close to the distribution $(\pi|\mathcal{W})_i$ for average i).*

- *$\mathbb{E}_i\left[I_{\pi|\mathcal{W}}(X_i; \Pi|Y_i R_i) + I_{\pi|\mathcal{W}}(Y_i; \Pi|X_i R_i)\right] \leq 4\|\pi\|/n$ (that is, the information cost of the distribution $(\pi|\mathcal{W})_i$ is low).*

The main challenge in proving this theorem is in the choice of the public random variable R_i, which enables relating the information of the protocol θ to that of $(\pi|\mathcal{W})$ *even under the conditioning on $\mathcal{W}$*. This technically-involved argument is a "conditional" analogue of Lemma 3.1 (for details see [BRWY12]). Note that the last proposition of Lemma 4.7 only guarantees that the information cost of the transcript under the distribution $(\pi|\mathcal{W})$ is low (on an average coordinate), while we need this property to hold for the simulating protocol θ, in order to apply the compression schemes of [BBCR10] which would finish the proof. Unfortunately, a protocol π that is statistically close to a low-information distribution needs not be a low-information protocol itself: Consider, for

example, a protocol π where with probability δ Alice sends her input $X \in \{0,1\}^n$ to Bob, and with probability $1 - \delta$ she sends a random string. Then π is δ-close to a 0-information protocol, but has information complexity of $\approx \delta \cdot n$, which could be arbitrarily high. [BRWY12] circumvented this problem by showing that the necessary compression schemes of [BBCR10] are "smooth" in the sense that they also work for protocols that are merely close to having low-information. In a followup work, Braverman and Weinstein exhibited a general technique for converting protocol which are statistically-close to having low information into actual low-information protocols (see Theorem 3 in [BW14]), which combining Lemma 4.7 also led to a strong direct product theorem in terms of information complexity, sharpening the "Information=Amortized Communication" Theorem of Braverman and Rao:

Theorem 4.8 ([BW14], informally stated). *Suppose that* $\mathsf{IC}_\mu(f, 2/3) = I$, *i.e., solving a single copy of f with probability $2/3$ under μ requires I bits of information. If $T \log(T) = o(n \cdot I)$, then* $\mathsf{suc}(\mu^n, f^n, T) \le \exp\left(-\Omega(n)\right).$

In fact, this theorem shows that the direct sum and product conjectures in randomized communication complexity are equivalent (up to polylogarithmic factors), and they are both equivalent to one-shot interactive compression, in the quantitative sense of Proposition 4.2 (we refer the reader to [BW14] for the formal details).

5   State of the Art Interactive Compression Schemes

In this section we present the two state-of-the-art compression schemes for unbounded-round communication protocols, the first due to Barak et al., and the second due to Braverman [BBCR10, Bra12]. As mentioned in the introduction, a natural idea for compressing a multi-round protocol is to try and compress each round separately, using ideas from the transmission (one-way) setup [Huf52, HJMR07, BR11]. Such compression suffers from one fatal flaw: It would inevitably require sending at least 1 bit of communication at each round, while the information revealed in each round may be $\ll 1$ (an instructive example is the protocol in which Alice sends Bob, at each round of the protocol, an independent coin flip which is ε-biased towards her input $X \sim Ber(1/2)$, for $\varepsilon \ll 1$). Thus any attempt to implement the compression on a round- by-round basis is hopeful only when the number of rounds is bounded but is doomed to fail in general (indeed, this is the essence of the bounded-round compression schemes of [BR11, BRWY13]).

The main feature of the compression results we present below is that they do not depend on the number of rounds of the underlying protocol, but only on the overall communication and information cost.

5.1   Barak et al.'s compression scheme

Theorem 5.1 ([BBCR10]). *Let π be a protocol executed over inputs $x, y \sim \mu$, and suppose $\mathsf{IC}_\mu(\pi) = I$ and $\|\pi\| = C$. Then for every $\varepsilon > 0$, there is a protocol τ which ε-simulates π, where*

$$\|\tau\| = O\left(\sqrt{C \cdot I} \cdot (\log(C/\varepsilon)/\varepsilon)\right). \tag{10}$$

Proof. The conceptual idea underlying this compression result is using public randomness to *avoid communication by trying to guess what the other player is about to say.* Informally speaking, the

players will use shared randomness to sample (correlated) *full paths* of the protocol tree, according to their private knowledge: Alice has the "correct" distribution on nodes that she owns in the tree (since conditioned on reaching these nodes, the next messages only depend on her input x), and will use her "best guess" (i.e., her prior distribution on Bob's next message, under μ, her input x and the history of messages) to sample messages at nodes owned by Bob. Bob will do the same on nodes owned by Alice. This "guessing" is done in a correlated way using public randomness (and no communication whatsoever (!)), in a way that guarantees that if the player's guesses are close to the correct distribution, then the probability that they sample the same bit is large.

The above step gives rise to two paths, P_A and P_B respectively. In the the next step, the players will use (mild) communication to find all inconsistencies among P_A and P_B and correct them one by one (according to the "correct" speaker). By the end of this process, the players obtain a consistent path which has the correct distribution $\Pi(x, y)$. Therefore, the overall communication of the simulating protocol would be comparable to the number of mistakes between P_A and P_B (times the communication cost of fixing each mistake). Intuitively, the fact that π has low information will imply that the number of inconsistencies is small, as inconsistent samples on a given node typically occur when the "receiver's" prior distribution is far from the "speaker's" correct distributions, which will in turn imply that this bit conveyed a lot of information to the receiver (Alas, we will see that if the information revealed by the i'th bit of π is ε, then the probability of making a mistake on the i'th node is $\approx \sqrt{\varepsilon}$, and this is the source of sub-optimality of the above result. We discuss this bottleneck at the end of the proof).

We now sketch the proof more formally (yet still leaving out some minor technicalities). Let $\Pi = M_1, \ldots, M_C$ denote the transcript of π. Each node w at depth i of the protocol tree of π is associated with two numbers, $p_{x,w}$ and $p_{y,w}$, describing the probability (according to each player's respective "belief") that conditioned on reaching w, the next bit sent in π is "1" (the right child of w). That is,

$$p_{x,w} := \Pr[M_i = 1 \mid xrM_{<i} = w] \quad , \text{ and } \quad p_{y,w} := \Pr[M_i = 1 \mid yr, M_{<i} = w]. \tag{11}$$

Note that if w is owned by the Alice, then $p_{x,w}$ is exactly the correct probability with which the i-th bit is transmitted in π, conditioned that π has reached w.

In the simulating protocol τ, the players first sample, without communication and using public randomness, a uniformly random number ρ_w in the interval $[0, 1]$, for every node w of the protocol tree[8]. For simplicity of analysis, in the rest of the proof we assume the public randomness is fixed to the vale $R = r$. Alice and Bob now privately construct the following respective trees $\mathcal{T}_A, \mathcal{T}_B$: For each node w, Alice includes the right child of w in $\mathcal{T}_A$ iff $p_{w,x} < \rho_w$, and the left child ("0") otherwise. Bob does the same by including the right child of w in $\mathcal{T}_B$ iff $p_{w,y} < \rho_w$.

The trees $\mathcal{T}_A$ and $\mathcal{T}_B$ define a unique path $\ell = m_1, \ldots, m_C$ of π, by combining outgoing edges from $\mathcal{T}_A$ in nodes owned by Alice, and edges from $\mathcal{T}_B$ in nodes owned by Bob. Note that ℓ has precisely the desired distribution of $\Pi(X, Y)$. To identify ℓ, the players will now find the inconsistencies among $\mathcal{T}_A$ and $\mathcal{T}_B$ and correct them one by one.

We say that a *mistake* occurs in level i if the outgoing edges of m_{i-1} in $\mathcal{T}_A$ and $\mathcal{T}_B$ are inconsistent. Finding the (first) mistake of τ amounts to finding the first differing index among two C-bit strings (corresponding to the paths P_A and P_B induced by $\mathcal{T}_A$ and $\mathcal{T}_B$). Luckily, there is a randomized protocol which accomplishes this task with high probability $(1 - \gamma)$ using only $O(log(C/\gamma))$

---

[8]Note that there are exponentially many nodes, but the communication model does not charge for local computations or the amount of shared randomness, so these resources are indeed "for free".

bits of communication, using a clever "noisy" binary search due to Feige et al. [FPRU94]. Since errors accumulate over C rounds and we are aiming for an overall simulation error of ε, we will set $\gamma \approx \varepsilon/C$, thus the cost of fixing each inconsistency remains $O(\log(C/\varepsilon))$ bits. The expected communication complexity of τ (over X, Y, R) is therefore

$$\mathbb{E}[\|\tau\|] = \mathbb{E}[\# \text{ mistakes of } \tau] \cdot O(\log(C/\varepsilon)). \tag{12}$$

Though we are not quite done, one should appreciate the simplicity of analysis of the cost of this protocol. The next lemma completes the proof, asserting that the expected number of mistakes τ makes is not too large:

Lemma 5.2. $\mathbb{E}[\# \text{ mistakes of } \tau] \leq \sqrt{C \cdot I}$.

Indeed, substituting the assertion of Lemma 5.2 into (12), we conclude that the expected communication complexity of τ is $O(\sqrt{C \cdot I} \cdot \operatorname{poly}\log(C/\varepsilon))$, and a standard Markov bound yields the bound in (10) and therefore finishes the proof of Theorem 5.1.

Proof of Lemma 5.2. Let $\mathcal{E}_i$ be the indicator random variable denoting whether a mistake has occurred in step i of the protocol tree of π. Hence the expected number of mistakes is $\sum_{i=1}^{C} \mathcal{E}_i$. We shall bound each term $\mathbb{E}[\mathcal{E}_i]$ separately. By construction, a mistake at node w in level i occurs exactly when either $p_{x,w} < \rho_w < p_{y,w}$ or $p_{y,w} < \rho_w < p_{x,w}$. Since ρ_w was uniform in $[0,1]$, the probability of a mistake is

$$|p_{x,w} - p_{y,w}| = |(M_i|x, r, M_{<i} = w) - (M_i|y, r, M_{<i} = w)|,$$

where the last transition is by definition of $p_{x,w}$ and $p_{y,w}$. Note that, by definition of a protocol, if $w := m_{<i}$ is owned by Alice, then $M_i|xyrm_{<i}] = M_i|xyrm_{<i}$ and if it is owned by Bob, then $M_i|y, r, m_{<i} = M_i|x, y, r, m_{<i}$. We therefore have

$$\mathbb{E}[\mathcal{E}_i] = \mathbb{E}_{xym_{<i}\sim\pi}[|(M_i|xrm_{<i}) - (M_i|yrm_{<i})|]$$
$$\leq \mathbb{E}_{xym_{<i}\sim\pi}\left[\max\{|(M_i|xyrm_{<i}) - (M_i|xrm_{<i})|, |(M_i|xyrm_{<i}) - (M_i|yrm_{<i})|\}\right]$$
$$\leq \mathbb{E}_{xym_{<i}\sim\pi}\left[\sqrt{\mathbb{D}\left(M_i|xyrm_{<i}\|M_i|xrm_{<i}\right) + \mathbb{D}\left(M_i|xyrm_{<i}\|M_i|yrm_{<i}\right)}\right] \tag{13}$$
$$\leq \sqrt{\mathbb{E}_{xym_{<i}\sim\pi}\left[\mathbb{D}\left(M_i|xyrm_{<i}\|M_i|xrm_{<i}\right) + \mathbb{D}\left(M_i|xyrm_{<i}\|M_i|yrm_{<i}\right)\right]} \tag{14}$$
$$= \sqrt{I(M_i; X|M_{<i}RY) + I(M_i; Y|M_{<i}RX)} \tag{15}$$

where transition (13) follows from Pinsker's inequality (Lemma 2.6), transition (14) follows from the convexity of $\sqrt{\cdot}$, and the last transition is by Proposition 2.3.

Finally, by linearity of expectation and the Cauchy-Schwartz inequality, we conclude that

$$\mathbb{E}\left[\sum_{i=1}^{C} \mathcal{E}_i\right] \leq \sum_{i=1}^{C} \sqrt{I(M_i; X|M_{<i}RY) + I(M_i; Y|M_{<i}RX)}$$
$$\leq \sqrt{\left(\sum_{i=1}^{C} 1\right) \cdot \left(\sum_{i=1}^{C} I(M_i; X|M_{<i}RY) + I(M_i; Y|M_{<i}RX)\right)}$$
$$= \sqrt{C \cdot I}$$

where the last transition is by the chain rule for mutual information. $\square$

$\square$

A natural question arising from the above compression scheme is whether the analysis in Lemma 5.2 is tight. Unfortunately, the answer is yes, as demonstrated by the following example: Suppose Alice has a uniform C-bit string $X_1 \ldots X_C$ where $X_i \sim Ber(1/2)$, and consider the C-bit protocol in which Alice sends, at each round i, an independent sample M_i such that

$$M_i \sim \begin{cases} Ber(1/2 + \varepsilon) & \text{if } X_i = 1 \\ Ber(1/2 - \varepsilon) & \text{if } X_i = 0 \end{cases}$$

for $\varepsilon = 1/\sqrt{C}$. Since Bob has a perfectly uniform prior on X, a direct calculation shows that in this case $I(M_i; X | M_{<i}) = I(M_i; X) = \mathbb{D}\left(Ber(1/2 + \varepsilon) \| Ber(1/2)\right) = O(\varepsilon^2)$, so the total information cost of the protocol is $O(C \cdot \varepsilon^2) = O(1)$. On the other hand, the probability of making a "mistake" at step i of the simulation above is the total variation distance $|Ber(1/2 + \varepsilon) - Ber(1/2)| \approx \varepsilon$. Therefore, the expected number of mistakes conditioned on, say, $X_1 = \ldots = X_C = 1$, is $C \cdot \varepsilon = \sqrt{C}$, by choice of $\varepsilon = 1/\sqrt{C}$. I.e., this example shows that both Pinsker's and the Cauchy-Schwartz inequalities are tight in the extreme case where each of the C bit of π reveals $\approx I/C$ bits of information. In the next section we present a different compression scheme which can do better in this regime, at least when I is much smaller than C.

5.2 Braverman's compression scheme

Theorem 5.3 ([Bra12]). *Let π be a protocol executed over inputs $x, y \sim \mu$, and suppose $\mathsf{IC}_\mu(\pi) = I$. Then for every $\varepsilon > 0$, there is a protocol τ which ε-simulates π, where $\|\tau\| = 2^{O(I/\varepsilon)}$.*

Proof. To understand this result, it will be useful to view the interactive compression problem as the following correlated sampling task: Denote by π_{xy} the distribution of the transcript $\Pi(x, y)$, and by π_x (resp. π_y) the conditional marginal distribution $\Pi | x$ ($\Pi | y$) of the transcript from Alice's (Bob's) point of view (for notational ease, the conditioning on the public randomness r of the protocol is included here implicitly. Note that in general π is still randomized even conditioned on x, y, since it may have private randomness). By the product structure of communication protocols, the probability of reaching a leaf (path) $\ell \in \{0, 1\}^C$ of π is

$$\pi_{xy}(\ell) = p_x(\ell) \cdot p_y(\ell) \tag{16}$$

where $p_x(\ell) = \prod_{w \subseteq \ell, w \text{ odd}} p_{x,w}$ is the product of the transition probabilities defined in (11) on the nodes owned by Alice along the path from the root to ℓ, and $\pi_y(\ell)$ is analogously defined on the even nodes. Thus, the desirable distribution from which the players wish to jointly sample, decomposes to a natural product distribution [9]. Similarly,

$$\pi_x(\ell) = p_x(\ell) \cdot q_x(\ell) \quad \text{and} \quad \pi_x(\ell) = q_y(\ell) \cdot p_y(\ell) \tag{17}$$

where $q_x(\ell) = \prod_{w \subseteq \ell, w \text{ even}} p_{x,w}$ is Alice's prior "belief" on the *even nodes* owned by Bob along the path to ℓ (see (11)), and $q_y(\ell) = \prod_{w \subseteq \ell, w \text{ odd}} p_{x,w}$ is Bob's prior belief on the odd nodes owned by Alice. Thus, the player's goal is to sample $\ell \sim \pi_{x,y}$, where Alice has the correct distribution on

---

[9] As we shall see, the rejection sampling approach of the compression protocol below crucially exploits this product structure of the target distribution, and it is curious to note this simplifying feature of interactive compression as opposed to general correlated sampling tasks.

odd nodes (and only an estimate on the odd ones), and Bob has the correct distribution on even nodes (and an estimate on the even ones).

We claim that the information cost of π being low (I) implies that Alice's prior "belief" q_x on the even nodes owned by Bob, is "close" to the true distribution p_y on these nodes (and vice versa for q_y and p_x on the odd nodes). To see this, recall the equivalent interpretation of mutual information in terms of KL-divergence:

$$I = I(\Pi; X|Y) + I(\Pi; Y|X) = \mathbb{E}_{(x,y)\sim\mu}\left[\mathbb{D}\left(\pi_{xy}\|\pi_y\right) + \mathbb{D}\left(\pi_{xy}\|\pi_x\right)\right]$$

$$= \mathbb{E}_{x,y,\ell\sim\pi_{x,y}}\left[\log\frac{\pi_{xy}(\ell)}{\pi_y(\ell)} + \log\frac{\pi_{xy}(\ell)}{\pi_x(\ell)}\right] = \mathbb{E}_{x,y,\ell\sim\pi_{x,y}}\left[\log\frac{p_x(\ell)}{q_y(\ell)} + \log\frac{p_y(\ell)}{q_x(\ell)}\right], \qquad (18)$$

where the last transition follows from substituting the terms according to (16) and (17). The above equation asserts that the typical log-ratio p_x/q_y is at most I, and the same holds for p_y/q_x. The following simple corollary essentially follows from Markov's inequality[10], so we state it without a proof.

Corollary 5.4. *Define the set of transcripts* $B_\varepsilon := \{\ell : p_x(\ell) > 2^{(I+1)/\varepsilon} \cdot q_y(\ell) \quad or \quad p_y(\ell) > 2^{(I+1)/\varepsilon} \cdot q_x(\ell) \}$. *Then* $\pi_{x,y}(B_\varepsilon) < \varepsilon$.

The intuitive operational interpretation of the above claim is that, for almost all transcripts ℓ, the following holds: If a *uniformly random* point $\in [0,1]$ falls below $p_y(\ell)$, then the probability it falls below q_x as well is $\gtrsim 2^{-I}$. This intuition gives rise to the following rejection sampling approach: The players interpret the public random tape as a sequence of points $(\ell_i, \alpha_i, \beta_i)$, uniformly distributed in $\mathcal{U} \times [0,1] \times [0,1]$, where $\mathcal{U} = \{0,1\}^C$ is the set of all possible transcripts of π. Their goal will be to discover the first index i^* such that $\alpha_{i^*} \leq p_x(\ell_{i^*})$ and $\beta_{i^*} \leq p_y(\ell_{i^*})$. Note that, by design, the probability that a random point ℓ_i satisfies these conditions is precisely $p_x(\ell_i) \cdot p_y(\ell_i) = \pi_{xy}(\ell_i)$, and therefore ℓ_{i^*} has the correct distribution.

The players consider only the first $t := 2|\mathcal{U}|\ln(1/\varepsilon)$ points of the public tape, as the probability that a single node satisfies the desirable condition is exactly $1/|\mathcal{U}|$, and thus by independence of the points, the probability that $i^* > t$ is at most $(1 - 1/|\mathcal{U}|)^t = \varepsilon^2 < \varepsilon/16$).

To do so, each player defines his own set of "potential candidates" for the index i^*. Alice defines the set

$$\mathcal{A} := \{i < T \; : \; \alpha_i \leq p_x(\ell_i) \text{ and } \beta_i \leq 2^{8I/\varepsilon} \cdot q_x(\ell_i)\}.$$

Thus $\mathcal{A}$ is the set of transcript which have the correct distribution on the odd nodes (which Alice can verify by herself), and "approximately" satisfies the desirable condition on the even nodes, on which Alice only has a prior estimate (q_x). Similarly, Bob defines

$$\mathcal{B} := \{i < t \; : \; \beta_i \leq p_y(\ell_i) \text{ and } \alpha_i \leq 2^{8I/\varepsilon} \cdot q_y(\ell_i)\}.$$

By Corollary 5.4, $\Pr[\ell^* \notin \mathcal{A} \cap \mathcal{B}] \leq \varepsilon/8$, so for the rest of the proof we assume that $\ell^* \in \mathcal{A} \cap \mathcal{B}$. In fact, ℓ^* is the first element of $\mathcal{A} \cap \mathcal{B}$. Note that for each point $(\ell_i, \alpha_i, \beta_i)$, $\Pr[\ell_i \in \mathcal{A} \cap \mathcal{B}] \leq 2^{8I/\varepsilon}/|\mathcal{U}|$.

---

[10]One needs to be slightly careful, since the log ratios can in fact be negative, while Markov's inequality applies only to non-negative random variables. However, it is well known that the contribution of the negative summands is bounded, see [Bra12] for a complete proof.

Since we consider only the first $t = 2|\mathcal{U}|\ln(1/\varepsilon)$ points, this implies $\mathbb{E}[|\mathcal{A}|] \leq 2^{8I/\varepsilon} \cdot 2\ln(1/\varepsilon)$, and Chernoff bound further asserts that

$$\Pr[|\mathcal{A}| > 2^{10I/\varepsilon}] \ll \varepsilon/16.$$

Thus, if we let $\mathcal{E}_1$ denote the event that $\ell^* \notin \mathcal{A} \cap \mathcal{B}$, and $\mathcal{E}_2 := \{i^* > t \text{ or } |\mathcal{A}| > 2^{10I/\varepsilon} \text{ or} |\mathcal{B}| > 2^{10I/\varepsilon} \}$, then by a union bound $\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \leq 2\varepsilon/8 + 3\varepsilon/16 < \varepsilon/2$. Thus, letting $\tau_{x,y}$ denote the distribution of $\ell_{i^*}|\neg(\mathcal{E}_1 \cup \mathcal{E}_2)$, the above implies

$$|\tau_{x,y} - \pi_{x,y}| \leq \varepsilon/2,$$

as desired. We will now show a (2-round) protocol τ in which Alice and Bob output a leaf $\ell \sim \tau_{x,y}$, thereby completing the proof. To this end, note we have reduced the simulation task to the problem of finding and outputting the first element in $\mathcal{A} \cap \mathcal{B}$, where $|\mathcal{A}| \leq 2^{10I/\varepsilon}$ and $|\mathcal{B}| \leq 2^{10I/\varepsilon}$. The idea is simple: Alice wishes to send her entire set $\mathcal{A}$ to Bob, who can then check for intersection with his set $\mathcal{B}$. Alas, explicitly sending each element $\ell \in \mathcal{A}$ may be too expensive (requires $\log|\mathcal{U}|$ bits), so instead Alice will send Bob sufficiently many $(O(I/\varepsilon))$ random hashes of the elements in $\mathcal{A}$, using a publicly chosen sequence of hash functions. Since for $a \in \mathcal{A}$ and $b \in \mathcal{B}$ such that $a \neq b$, the probability (over the choice of the hash functions) that $h_j(a) = h_j(b)$ for all $j \in O(I/\varepsilon)$ is bounded by $2^{-O(I/\varepsilon)} < \frac{\varepsilon}{4|\mathcal{A}| \cdot |\mathcal{B}|}$, a union bound ensures that the probability there is an $a \in \mathcal{A}$, $b \in \mathcal{B}$ such that $a \neq b$ but the hashes happen to match, is bounded by $\varepsilon/4$, which completes the proof. For completeness, the protocol τ is described in Figure 1.

The simulation protocol τ

1. Alice computes the set $\mathcal{A}$. If $|\mathcal{A}| > 2^{10I/\varepsilon}$ the protocol fails.

2. Bob computes the set $\mathcal{B}$. If $|\mathcal{B}| > 2^{10I/\varepsilon}$ the protocol fails.

3. For each $a \in \mathcal{A}$, Alice computes $d = \lceil 20I/\varepsilon + \log 1/\varepsilon + 2 \rceil$ random hash values $h_1(a), \ldots, h_d(a)$, where the hash functions are evaluated using public randomness.

4. Alice sends the values $\{h_j(a_i)\}_{a_i \in \mathcal{A},\ 1 \leq j \leq d}$ to Bob.

5. Bob finds the first index i such that there is a $b \in \mathcal{B}$ for which $h_j(b) = h_j(a_i)$ for $j = 1..d$ (if such an i exists). Bob outputs ℓ_b and sends the index i to Alice.

6. Alice outputs ℓ_i.

Figure 1: A simulating protocol for sampling a transcript of $\pi(x, y)$ using $2^{O(I/\varepsilon)}$ communication.

$\square$

6 Concluding Remarks and Open Problems

We have seen that direct sum and product theorems in communication complexity are essentially equivalent to determining the best possible interactive compression scheme. Despite the exciting progress described in this survey, this question is still far from settled, and the natural open problem

is closing the gap in (9). The current frontier is trying to improve the dependence on C over the scheme of [BBCR10], even at a possible expense of increased dependence on the information cost:

Open Problem 6.1 (Improving compression for internal information). *Given a protocol π over inputs $x, y \sim \mu$, with $\|\pi\| = C, \mathsf{IC}_\mu(\pi) = I$, is there a communication protocol τ which (0.01)-simulates π such that $\|\tau\| \leq \mathrm{poly}(I) \cdot C^{1/2-\varepsilon}$, for some absolute positive constant $0 < \varepsilon < 1/2$?*

In fact, by a recent result of Braverman and Weinstein [BW14], even a much weaker compression scheme in terms of I, namely $g(I, C) \leq 2^{o(I)} \cdot C^{1/2-\varepsilon}$ would already improve over the the state of the art compression scheme ($\tilde{O}(\sqrt{C \cdot I})$) and would imply new direct sum and product theorems.

Another interesting direction which was unexplored in this survey, is closing the (much smaller) gap in (4.4), i.e, determining whether a logarithmic dependence on C is essential for interactive compression with respect to the *external information cost* measure.

Open Problem 6.2 (Closing the gap for external compression). *Given a protocol π over inputs $x, y \sim \mu$, with $\|\pi\| = C, \mathsf{IC}_\mu^{\mathsf{ext}}(\pi) = I$, is there a communication protocol τ which δ-simulates π such that $\|\tau\| \leq \mathrm{poly}(I) \cdot o(\log(C))$?*

It is believed that the $(\log C)$ factor is in fact necessary (see e.g., that candidate separation sampling problem suggested in [Bra13]), but this conjecture remains to be proved.

Recall that in Section 4.1 we saw direct product theorems for randomized communication complexity, asserting a lower bound on the success rate of computing n independent copies of f in terms of the success of a single copy. When n is very large, such theorems can be superseded by trivial arguments, since f^n must require at least n bits of communication just to describe the output. One could hope to achieve hardness amplification without blowing up the output size – a classical example is Yao's XOR lemma in circuit complexity. In light of the state-of-the-art direct product result, we state the following conjecture:

Open Problem 6.3 (A XOR Lemma for communication complexity). *Is it true that for any 2-party function f and any distribution μ on $\mathcal{X} \times \mathcal{Y}$,*

$$\mathsf{D}_{\mu^n}(f^{\oplus n}, 1/2 + e^{-\Omega(n)}) = \tilde{\Omega}(\sqrt{n}) \cdot \mathsf{D}_\mu(f, 2/3)?$$

(here $f^{\oplus n}((x_1, y_1), \ldots, (x_n, y_n)) := f(x_1, y_1) \oplus \ldots \oplus f(x_n, y_n)$).

We remark that the "direct-sum" analogue of this conjecture is true: [BBCR10] proved that their direct sum result for f^n can be easily extended to the computation of $f^{\oplus n}$, showing (roughly) that $\mathsf{D}_{\mu^n}(f^{\oplus n}, 3/4) = \tilde{\Omega}(\sqrt{n}) \cdot \mathsf{D}_\mu(f, 2/3)$. However, this conversion technique does not apply to the direct product setting.

Acknowledgements

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 67–76, 2010.

[BBK+13] Joshua Brody, Harry Buhrman, Michal Koucký, Bruno Loff, Florian Speelman, and Nikolay K. Vereshchagin. Towards a reverse newman's theorem in interactive information complexity. In *Proceedings of the 28th Conference on Computational Complexity, CCC 2013, K.lo Alto, California, USA, 5-7 June, 2013*, pages 24–33, 2013.

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012.

[BEO+13] Mark Braverman, Faith Ellen, Rotem Oshman, Toniann Pitassi, and Vinod Vaikuntanathan. Tight bounds for set disjointness in the message passing model. *CoRR*, abs/1305.4696, 2013.

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 151–160, New York, NY, USA, 2013. ACM.

[BM12] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:131, 2012.

[BMY14] Balthazar Bauer, Shay Moran, and Amir Yehudayoff. Internal compression of protocols to entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:101, 2014.

[BP13] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 688–697, 2013.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In Rafail Ostrovsky, editor, *FOCS*, pages 748–757. IEEE, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[Bra13] Mark Braverman. A hard-to-compress interactive task? In *2013 51st Annual Allerton Conference on Communication, Control, and Computing, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*, pages 8–12, 2013.

[Bra14] Mark Braverman. Interactive information and coding theory. 2014.

[BRWY12] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:143, 2012.

[BRWY13]   Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:35, 2013.

[BT91]   Richard Beigel and Jun Tarui. On acc. In *FOCS*, pages 783–792, 1991.

[BW12]   Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. In *APPROX-RANDOM*, pages 459–470, 2012.

[BW14]   Mark Braverman and Omri Weinstein. An interactive information odometer with applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:47, 2014.

[BYJKS04]   Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[CKS03]   Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *IEEE Conference on Computational Complexity*, pages 107–117, 2003.

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

[CT91]   Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[DN11]   Shahar Dobzinski and Noam Nisan. Limitations of vcg-based mechanisms. *Combinatorica*, 31(4):379–396, 2011.

[FKNN95]   Tomàs Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 24(4):736–750, 1995. Prelim version by Feder, Kushilevitz, Naor FOCS 1991.

[FPRU94]   Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994.

[GKR14]   Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:49, 2014.

[GMWW14]   Dmitry Gavinsky, Or Meir, Omri Weinstein, and Avi Wigderson. Toward better formula lower bounds: An information complexity approach to the krw composition conjecture. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 213–222, New York, NY, USA, 2014. ACM.

[GO13]   Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. In *IEEE Conference on Computational Complexity*, pages 287–298, 2013.

[HJMR07] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *IEEE Conference on Computational Complexity*, pages 10–23. IEEE Computer Society, 2007.

[Hol07] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007.

[HRVZ13] Zengfeng Huang, Bozidar Radunovic, Milan Vojnovic, and Qin Zhang. Communication complexity of approximate maximum matching in distributed graph data. Technical Report MSR-TR-2013-35, April 2013.

[Huf52] D.A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[Jai11] Rahul Jain. New strong direct product results in communication complexity. 2011.

[JKR09] T. S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once ac^0 formulae. In *IEEE Conference on Computational Complexity*, pages 329–340, 2009.

[JPY12] Rahul Jain, Attila Pereszlenyi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 167–176. IEEE, 2012.

[JY12] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. *CoRR*, abs/1209.0263, 2012.

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *STOC*, pages 77–86, 2010.

[KLL+12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *CoRR*, abs/1204.1505, 2012.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Computational Complexity*, 5(3/4):191–204, 1995. Prelim version CCC 1991.

[KW88] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. In *STOC*, pages 539–550, 1988.

[LS10] Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *Computational Complexity*, 19(2):153–181, 2010.

[LSS08] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *CCC*, pages 71–80, 2008.

[MWY13] Marco Molinaro, David Woodruff, and Grigory Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *SODA*, page to appear, 2013.

[PRW97]   Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing (STOC '97)*, pages 363–372, New York, May 1997. Association for Computing Machinery.

[PW10]   Mihai Patrascu and Ryan Williams. On the possibility of faster sat algorithms. In Moses Charikar, editor, *SODA*, pages 1065–1075. SIAM, 2010.

[Rao08]   Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, 2008.

[Raz98]   Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998. Prelim version in STOC '95.

[Raz08]   Alexander Razborov. A simple proof of bazzi's theorem. Technical Report TR08-081, ECCC: Electronic Colloquium on Computational Complexity, 2008.

[RR15]   Anup Rao and Sivaramakrishnan Natarajan Ramamoorthy. How to compress asymmetric communication. *Electronic Colloquium on Computational Complexity (ECCC), 2015.*, 2015.

[RS15]   Anup Rao and Makrand Sinha. Simplified separation of information and communication. *Electronic Colloquium on Computational Complexity (ECCC), 2015.*, 2015.

[Sha48]   Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948. Monograph B-1598.

[Sha03]   Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003. Prelim version CCC 2001.

[She12]   Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012.

[ST13]   Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. *CoRR*, abs/1304.1217, 2013.

[Wac90]   Juraj Waczulk. Area time squared and area complexity of vlsi computations is strongly unclosed under union and intersection. In Jrgen Dassow and Jozef Kelemen, editors, *Aspects and Prospects of Theoretical Computer Science*, volume 464 of *Lecture Notes in Computer Science*, pages 278–287. Springer Berlin Heidelberg, 1990.

[Wil12]   Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 887–898, New York, NY, USA, 2012. ACM.

[WZ14]   David P. Woodruff and Qin Zhang. An optimal lower bound for distinct elements in the message passing model. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 718–733, 2014.

[Yao79]   Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *STOC*, pages 209–213, 1979.

[Yao82]   Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *FOCS*, pages 80–91. IEEE, 1982.

[ZDJW13]   Yuchen Zhang, John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2328–2336, 2013.