Information Complexity Is Computable*

Mark Braverman[†1] **and Jon Schneider**[2]

1 **Department of Computer Science, Princeton University, Princeton, NJ, USA**
mbraverm@cs.princeton.edu

2 **Department of Computer Science, Princeton University, Princeton, NJ, USA**
js44@cs.princeton.edu

Abstract

The information complexity of a function f is the minimum amount of information Alice and Bob need to exchange to compute the function f. In this paper we provide an algorithm for approximating the information complexity of an arbitrary function f to within any additive error $\epsilon > 0$, thus resolving an open question as to whether information complexity is computable.

In the process, we give the first explicit upper bound on the rate of convergence of the information complexity of f when restricted to b-bit protocols to the (unrestricted) information complexity of f.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases Communication complexity, convergence rate, information complexity

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.87

1 Introduction

In 1948, Shannon introduced the field of information theory as a set of tools for understanding the limits of one-way communication [15]. One of these tools, the information entropy function $H(X)$, measures the amount of information contained in a random source X.

The analogue of information entropy in communication complexity is *information complexity*. The information complexity of a function f is the least amount of information Alice and Bob need to exchange about their inputs to compute a function f. Just as the information entropy of a random source X provides a lower bound on the amount of communication required to transmit X, the information complexity of a function f provides a lower bound on the communication complexity of f [3]. Moreover, just as Shannon's source coding theorem establishes $H(X)$ as the asymptotic communication-per-message required to send multiple independent copies of X, the information complexity of f is the asymptotic communication-per-copy required to compute multiple copies of f in parallel on independently distributed inputs [7, 5].

These properties make information complexity a valuable tool for proving results in communication complexity. Communication complexity lower bounds themselves have a wide variety of applications to other areas of computer science; for example, results in circuit complexity such as Karchmer-Wigderson games and ACC lower bounds rely on communication complexity lower bounds [12, 4]. In addition, techniques from information complexity have been applied to prove various direct sum results in communication complexity

* A full version of this paper is available at http://arxiv.org/abs/1502.02971.

† Research supported in part by an NSF CAREER award (CCF-1149888), NSF CCF-1525342, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry.

[8, 2, 11], including the only known direct sum results for general randomized communication complexity [3]. Information complexity has also been applied to prove a tight asymptotic bound on the communication complexity of the set disjointness function [6].

Despite this, many fundamental properties of information complexity remain unknown [6]. It is unknown how the information complexity of a function changes asymptotically as we allow the protocol to fail with probability ϵ. It is unknown how the information complexity of a function grows if we restrict our attention to protocols of bounded depth. Perhaps most surprisingly, it is even unknown if, given the truth table of a function f, whether it is possible to even compute (to within some additive factor of ϵ) the information complexity of f, $IC_\mu(f)$. (Contrast this with the case of the information entropy $H(X)$, which is easily computed given the distribution of X).

In this paper, we resolve the last of these questions; we prove that the information complexity of f is indeed computable. Our main technical result is an explicit bound on the convergence rate of b-bit information complexity (information complexity when restricted to protocols that have total communication at most b bits) to unrestricted information complexity. More specifically, we show how to convert an arbitrary protocol π into a protocol π' that leaks at most ϵ more information than π, but has communication cost at most $(N\epsilon^{-1})^{O(N)}$ bits, where N is the size of the truth table of f (Theorem 12). Equivalently, we show that the b-bit information complexity of f is at most $b^{-O(N^{-1})}$ larger than the information complexity of f. By then enumerating over all protocols with this communication cost, we obtain an algorithm that computes the information complexity of f to within an additive factor of ϵ in time $2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)}$ (here N is the size of the truth table of f).

1.1 Prior Work

In [13, 14], Ma and Ishwar present a method to compute tight bounds on the information complexity of functions for protocols restricted to r rounds of computation. By examining the limit as r tends to infinity, this method allows them to numerically compute the information complexity of several functions (such as the 2-bit AND function). To make these computations provably correct, one would need effective (computable) estimates on the rate of convergence of r-round information complexity to the true information complexity. Such estimates were unknown prior to the present paper.

Plenty of unsolved problems of this flavor – where the computability of some limiting value is unknown despite it being straightforward to compute individual terms of this limit – occur in information theoretic contexts. One famous problem is the problem of computing the Shannon capacity of a graph, the amortized independence number of the kth power of a graph (this limiting quantity also has an interpretation as the zero-error channel capacity of a certain channel defined by this graph). While computing the independence number of any given graph is possible (albeit NP-hard), the rate at which this limit converges is unknown. Indeed, Alon and Lubetzky have shown that the limiting behavior of this quantity can be quite complex; no fixed number of terms of this limit is guaranteed to give a subpolynomial approximation to the Shannon capacity [1]. Another example, from the realm of quantum information theory, occurs in computing the quantum value of games [9]. Here it is straightforward to compute the quantum value of a game when limited to n bits of entanglement, but no explicit bounds are known for how many bits of entanglement are required to achieve within ϵ of optimal performance.

1.2 Outline of Proof

The main result of our paper is that zero-error information complexity is computable. Formally, we prove the following theorem.

▶ **Theorem 1.** *There exists an algorithm which, given a function $f : \mathcal{A} \times \mathcal{B} \to \{0,1\}$, initial distribution $\mu \in \Delta(\mathcal{A} \times \mathcal{B})$, and a real number $\epsilon > 0$, returns a value C between $IC_\mu(f) - \epsilon$ and $IC_\mu(f) + \epsilon$. This can be performed in time $2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)}$, where $N = |\mathcal{A} \times \mathcal{B}|$.*

Throughout this paper, we will take the perspective of an outside observer watching in as Alice and Bob execute some protocol. This observer starts with some probabilistic *belief* about the inputs of Alice and Bob (initially this is just μ, the distribution of inputs to Alice and Bob). As Alice and Bob execute the protocol, they send each other *signals* – Bernoulli random variables that contain information about their inputs – which cause the observer to update his belief. The total amount of information leaked by the protocol to the participants can then be represented directly in terms of the final belief and initial belief. These notions are defined in more detail in Section 2.

The strategy of the proof is as follows. We start with a general protocol π for solving f, and whose information cost is very close to the information complexity of f. Unfortunately, we do not know anything about π besides the fact that it's a finite, discrete protocol that computes f without error. Note that if we could restrict π to a finite family of protocols (e.g. protocols that sent at most b bits, for an explicit bound $b = b(\epsilon, N)$), then we could just brute force over all such π's and compute the approximate information complexity of f. The proof shows that, indeed, there is always a protocol π' that can be derived from π, and which belongs to such an explicit family. The proof proceeds in several steps. In each step, more structure is added to π (structure that is then exploited by the following steps). The difficulty is, of course, ensuring at each step that π can be replaced with a more structured protocol π' while increasing its information cost by only, say, $\epsilon/10$. Ultimately, we manage to turn π into a protocol with r back-and-forth rounds, where r is an explicit function of N and ϵ. Finally, it is shown that an r-rounds of interaction protocol can be replaced with a b-bit protocol where $b = b(\epsilon, N, r) = b(\epsilon, N)$ is an explicit function, while only increasing its information cost by a controlled amount, completing the proof.

The full proof of Theorem 1 is roughly structured into three parts. In the first part, we begin by showing that we can 'discretize' any protocol π; that is, we can simulate any protocol π with a protocol π' that only uses a bounded number of different types of signals, but that only reveals a marginal amount of additional information. We accomplish this by building a 'mesh' of signals and rounding each signal in π to one of the signals in this mesh. This is described in Section 3.1.

In the second part, we show in Section 3.2 that we can transform any suitably discrete protocol π (i.e. one that uses an explicitly bounded number of distinct signals) into a protocol that uses few rounds. We achieve this via a bundling scheme; the main idea is that, where Alice would ordinarily send Bob one instance of a signal, she instead sends Bob several instances of this signal. Then, the next several times Alice would send that signal to Bob, Bob simply refers to the next unused copy sent by Alice, thus decreasing the number of rounds in the protocol.

In the third part, we show in Section 3.3 that it is never advantageous to send more than $\log N$ bits in any round of a protocol, thus providing an explicit bound on the communication complexity of the protocol. We accomplish this by showing that if it is ever the case that Alice sends one of $M > N$ different messages in a round, Alice can use public randomness

to sample a subset of N of these M messages and use $\log N$ bits to send one of these N messages instead.

Combining the above steps allows us to prove the following bound on the convergence rate of b-bit information complexity.

▶ **Theorem 12** (restated). *Let π be a communication protocol with information cost C that successfully computes function f over inputs drawn from distribution μ over $\mathcal{A} \times \mathcal{B}$. Then there exists a protocol π' with information cost at most $C + \epsilon$ that also successfully computes f over inputs drawn from μ, but has communication cost at most $b(f, \epsilon)$ where*

$$b(f, \epsilon) = (N\epsilon^{-1})^{O(N)} \tag{1}$$

and $N = |\mathcal{A} \times \mathcal{B}|$.

Finally, by reapplying the discretization procedure of Section 3.1, we show that it suffices to consider protocols whose signals all belong to an explicit finite set. By enumerating over all protocols of this depth that use signals from this set and computing the minimum information cost of any such protocol, we can therefore approximate $IC_\mu(f)$ to within ϵ, thus completing the proof.

The proof we provide below shows that zero-error internal and external information complexity are computable. We believe similar techniques can be used to show that ϵ-error information complexity is computable, but do not include such a proof in this paper.

1.3 Open Problems

Naturally, the most immediate open problem arising from our work is understanding whether (and how much) the rate of convergence in Theorem 12 can be improved:

▶ **Open Problem 2.** *What is the (worst case) rate of convergence of the b-bit (or r-round) information complexity of f to $IC_\mu(f)$? In other words, for a given $\epsilon > 0$ and truth table size $N = |\mathcal{A} \times \mathcal{B}|$, how large does $b(N, \epsilon)$ need to be to ensure that the b-bit information complexity $IC_{b,\mu}(f)$ satisfies*

$$IC_{b,\mu}(f) > IC_\mu(f) - \epsilon?$$

In this paper we prove that $b(N, \epsilon) \leq (N\epsilon^{-1})^{O(N)}$. On the other hand, [6] shows that when f is the two-bit AND (and thus $N = 4$ is a constant), the tight estimate for b is $b = \Theta(\epsilon^{-1/2})$. Therefore, the polynomial dependence on ϵ, even when N is a constant, is necessary. On the other hand, we do not have any interesting lower bounds on b in terms of N. In particular, it is not known whether the exponential dependence on N is necessary here.

The second open problem is in a similar vein, asking whether Theorem 1 can be improved.

▶ **Open Problem 3.** *What is the computational complexity of computing the (zero-error internal) information complexity of a function f within error ϵ given its truth table? By how much can the bound of $2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)}$ be improved?*

By the analysis in Section 3.4, any progress on Problem 2 will translate into progress on Problem 3. For comparison, it is not hard to see that the trivial algorithm for computing the *average-case communication complexity* of a function $f : [n] \times [n] \to \{0, 1\}$ (so that $N = n^2$) within an additive error ϵ runs in time $2^{n \cdot N^{N/\epsilon}} = 2^{\exp((N\epsilon^{-1})^{O(1)})}$ (it suffices to look at all protocols of depth at most $\frac{N \log N}{\epsilon}$, of which there are at most $2^{n \cdot N^{N/\epsilon}}$). In other words, there is an exponential gap between the trivial communication complexity upper bound and the bound we obtain in Theorem 1.

2 Preliminaries

2.1 Information Theory

For an introduction to the information theoretic notions used throughout this paper, we refer the reader to [10] (A brief introduction can also be found in the full version of this paper).

2.2 Protocols and Information Complexity

In the two-party communication setting, Alice is given an element a from a finite set $\mathcal{A}$, while Bob is given an element b from a finite set $\mathcal{B}$, where (a, b) is drawn from some distribution μ over $\mathcal{A} \times \mathcal{B}$. Their goal is to compute $f(a, b)$, where $f : \mathcal{A} \times \mathcal{B} \to \{0, 1\}$ is a function known to both parties. They would like to accomplish this while revealing as little information as possible; either to each other (in the case of information cost) or to an outside observer (in the case of external information cost). To do this, they execute a *communication protocol*, which we view as being built out of *signals*.

▶ **Definition 4.** A *signal* σ over a set S is an assignment of a probability $\sigma_s \in [0, 1]$ to each element s in S. For a given element s of S, we define $\sigma(s)$ to be the Bernoulli random variable that equals 1 with probability σ_s.

▶ **Definition 5.** A *communication protocol* π is a finite rooted binary tree, where each non-leaf node is labeled by either a signal over $\mathcal{A}$ (corresponding to Alice's move) or a signal over $\mathcal{B}$ (corresponding to Bob's move), and each edge is labeled either 0 or 1. Alice and Bob can execute this protocol by starting at the root and repeatedly performing the following procedure; if the signal σ at the current node is a signal over $\mathcal{A}$, Alice sends Bob an instance of $\sigma(a)$, and they both move down the corresponding edge; likewise, if the signal is a signal over $\mathcal{B}$, Bob performs the analogous procedure.

Each leaf node is labeled with a value 0 or 1. We say the communication protocol *successfully computes* f with zero error if the value of the leaf node Alice and Bob finish the protocol on is always equal to $f(a, b)$ for all $(a, b) \in \mathcal{A} \times \mathcal{B}$ (in particular, even (a, b) where $\mu(a, b) = 0$). The communication cost $CC(\pi)$ of protocol π is equal to the depth of the deepest leaf in π.

This agrees with the usual definition of a private coins protocol (indeed, any bit Alice can ever send in any protocol must be a signal over $\mathcal{A}$, and likewise for Bob). A public coins protocol is simply a distribution over private coins protocols. For our purposes, it suffices to solely examine private coins protocols, since the information cost of a public coins protocol is simply the expected information cost of the corresponding private coins protocols.

As is standard, we will let A and B be random variables representing Alice's input and Bob's input respectively, and let Π be the random variable representing the protocol's transcript. We can then define the *information cost* of a protocol and the *information complexity* of a function as follows.

▶ **Definition 6.** The *information cost* of a protocol π is given by

$$IC_\mu(\pi) = I(A; \Pi | B) + I(B; \Pi | A)\,.$$

The *external information cost* of a protocol π is given by

$$IC^{ext}_\mu(\pi) = I(AB; \Pi)\,.$$

▶ **Definition 7.** The *information complexity* of a function f is given by

$$IC_\mu(f) = \inf_\pi IC_\mu(\pi)$$

where the infimum is over all protocols π that successfully compute f. Likewise, the *external information complexity* of a function f is given by

$$IC^{ext}_\mu(f) = \inf_\pi IC^{ext}_\mu(\pi)\,.$$

where again, the infimum is over all protocols π that successfully compute f.

Throughout the remainder of this paper, it will be useful to think of signals as operating on the space $\Delta(\mathcal{A}\times\mathcal{B})$ of probability distributions over $\mathcal{A}\times\mathcal{B}$, which we term *beliefs*. At the beginning of a protocol, an outside observer's belief is simply given by μ, the distribution (a,b) was drawn from. As this observer observes new signals, his belief evolves according to Bayes' rule; for example, if he observes the signal $\sigma(a)$ sent by Alice, his belief changes from the prior belief p to the posterior belief

$$p_0(a,b) = \frac{(1-\sigma_a)p(a,b)}{\sum_{i,j}(1-\sigma_i)p(i,j)} \tag{2}$$

if $\sigma(a) = 0$ (which occurs with probability $P_0 = \sum_{i,j}(1-\sigma_i)p(i,j)$) and to the posterior belief

$$p_1(a,b) = \frac{\sigma_a p(a,b)}{\sum_{i,j}\sigma_i p(i,j)} \tag{3}$$

if $\sigma(a) = 1$ (which occurs with probability $P_1 = \sum_{i,j}\sigma_i p(i,j)$). As shorthand, we will say that σ *shifts* belief p to (p_0,p_1). Note that the probabilities P_0 and P_1 are uniquely recoverable given p_0 and p_1 (in particular, treating beliefs as vectors in $\mathbb{R}^{|\mathcal{A}\times\mathcal{B}|}$, it must be the case that $P_0p_0 + P_1p_1 = p$ and that $P_0 + P_1 = 1$).

Throughout the remainder of the paper, we will let $N = |\mathcal{A}\times\mathcal{B}| = |\mathcal{A}|\cdot|\mathcal{B}|$. Note that N is the size of the truth table of f and is thus (in some sense) the size of the input to the problem of computing the information complexity of f. All logarithms are to base 2 unless otherwise specified.

3 Computability of Information Complexity

3.1 Discretizing signals

In the first part of the proof, we show that we can convert any protocol for f into a protocol that uses a bounded number of types of signals. In particular, we prove the following theorem.

▶ **Theorem 8.** *Let π be a communication protocol with information cost C. Then, for any $\epsilon > 0$, there exists a communication protocol π' that computes the same function as π with information cost at most $C+\epsilon$ but that only uses $Q = (N\epsilon^{-1})^{O(N)}$ different types of signals.*

Proof Sketch. Recall that signals are simply vectors in $\mathbb{R}^N$. Our general approach will therefore be to build a 'mesh' of signals in $\mathbb{R}^N$ and round each signal in our protocol to one of the nearby signals in the mesh. We can show this rounding procedure preserves the correctness of the protocol but possibly leaks some additional information.

Via the concavity of mutual information, it happens that if we take the width of this mesh to be $poly(\epsilon/N)$, then this rounding procedure leaks at most ϵ additional information. Such a mesh in N dimensions contains at most $(1/poly(\epsilon/N))^N = (N\epsilon^{-1})^{O(N)}$ different signals, as desired. ◀

The above sketch suppresses a number of technical difficulties in proving the above theorem. In particular, in the full paper, we demonstrate how to deal with:

1. Initial distributions μ that lack full support (Section 3.1 in full paper).
2. Signals sent near the boundary of $\Delta(\mathcal{A} \times \mathcal{B})$ (Section 3.2 in full paper).
3. Signals with widely differing magnitudes (Section 3.3 in full paper).

3.2 Bounding the number of alternations

We next show that we can convert a protocol for f that uses a bounded number of distinct signals (yet arbitrarily many of them) into a protocol for f that, while leaking at most ϵ extra information, uses a bounded number of alternations (steps in the protocol where Alice stops talking and Bob starts talking, or vice versa).

We achieve this by 'bundling' signals of the same type together; that is, at a point in the protocol where Alice would send Bob a certain signal, she may instead send him a bundle of t signals. Then, the next $t-1$ times Alice would send Bob this signal, Bob instead refers to the next unused signal in the bundle. If there are unused signals in a bundle, this may increase the information cost of the protocol; however, by choosing the size of the bundle cleverly, we can bound the size of this increase.

▶ **Definition 9.** Let π be a communication protocol and let $v_1, v_2, \ldots, v_k$ be one possible computation path for π. An *alternation* in this computation path is an index i where the signals at v_i and v_{i+1} are sent by different players. The number of alternations in π is the maximum number of alternations over all computation paths of π.

▶ **Theorem 10.** *Let π be a communication protocol with information cost C that only uses Q distinct signals. Then, for any $\epsilon > 0$, there exists a communication protocol π' that computes the same function as π with information cost at most $C + 2\epsilon$ but that uses at most*

$$W = \left(\frac{2Q \log N}{\epsilon} + Q\right) \frac{\log N}{\epsilon}$$

alternations.

Proof Sketch. Label our Q different signals $\sigma^{(1)}$ through $\sigma^{(Q)}$. We will reduce the number of alternations in π by bundling signals of the same type in large groups. That is, if Alice (at a specific point in the protocol) would send Bob signal $\sigma^{(i)}$, she instead sends Bob t copies of signal i (for an appropriately chosen t). Then, the next $t-1$ times in the protocol that Alice would send Bob signal $\sigma^{(i)}$, Bob instead refers to one of the unused t copies Alice originally sent. Once these t copies are depleted and protocol calls for a $(t+1)$st copy, the process repeats and Alice sends a new bundle to Bob (possibly with a different value for t).

We choose t as follows. Without loss of generality, assume Alice is sending a bundle of signals σ to Bob. Let Π_{pre} be the transcript of the protocol thus far. Let $X^t = (X_1, X_2, \ldots, X_t)$ be a random variable corresponding to t independently generated outputs of σ. We consider three cases:

Case 1: It is the case that $I(A; X^1 | \Pi_{pre} B) \geq \frac{\epsilon}{Q}$. In this case we set $t = 1$ (note that this is equivalent to simply following the original protocol).

Case 2: There exists a positive t_0 such that $\frac{\epsilon}{2Q} \leq I(A; X^{t_0} | \Pi_{pre} B) \leq \frac{\epsilon}{Q}$. In this case, we set $t = t_0$.

Case 3: For all positive t, $I(A; X^t | \Pi_{pre} B) \leq \frac{\epsilon}{2Q}$. In this case, we set t to be the maximum number of times signal σ is ever sent in protocol π.

The remainder of this proof is divided into three parts. In the first part, we argue that the three cases above are comprehensive. In the second part, we argue that the information cost of this new protocol is at most $C + \epsilon$. Finally, in the third part we argue that this bundling process decreases the total number of alternations to at most W. We briefly sketch these arguments here (see the full paper for detailed proofs).

Cases are comprehensive

It is not immediately clear that one of the three above cases must occur; it could be the case that $I(A; X^t|\Pi_{pre}B)$ 'jumps' from below $\epsilon/2Q$ to above ϵ/Q as t increases by one step. To show this cannot happen, we prove a 'diminishing returns' theorem for information revealed by additional copies of X (in particular, we show $I(A; X^{t+1}|\Pi_{pre}B) - I(A; X^t|\Pi_{pre}B)$ is decreasing in t).

Information leakage is small

When not all the signals in a bundle are used, this new protocol leaks some additional information over our original protocol. However, by the selection of t, each bundle is either a Case 1 bundle (which is immediately used up) or leaks at most ϵ/Q information. Since there are at most Q unused bundles (one for each signal type), we leak at most ϵ additional information.

Number of alternations is small

The number of alternations is at most the number of bundles sent. With the exception of Case 3 bundles (of which we send at most one of each type, for a total of Q), each bundle increases the expected information revealed by at least $\epsilon/2Q$. Since the amount of information revealed by the protocol is bounded above by $\log N$, in expectation we send at most $Q + \frac{2Q \log N}{\epsilon}$ bundles. To translate this to a worst case result, we simply terminate the protocol after sending at most W bundles; it then follows from Markov's inequality that we leak at most ϵ additional information by doing this. ◀

3.3 Bounding the number of bits

We finally show that each alternation in any protocol can be executed in a way that requires the exchange of at most $\log N$ bits without any loss in information cost; it follows that a protocol with at most W alternations can be converted into an equivalent protocol with communication complexity at most $W \log N$.

▶ **Theorem 11.** *Let π be a communication protocol with information cost C that has W alternations. Then there exists a communication protocol π' with information cost C that computes the same function as π but that sends a total of at most $W \log N$ bits.*

Proof. We will show how to execute each alternation of a protocol in at most $\log N$ bits. For simplicity, we will assume Alice and Bob have access to public randomness; this can later be converted into a protocol with only private randomness via the observation that some fixed choice of public randomness minimizes the information cost of the protocol.

Assume that Alice is speaking during some alternation of π, and let there be M possible strings she may send to Bob. If p is the belief at the beginning of the alternation, then at the end of the alternation we will have shifted to one of M possible beliefs, $p_1, p_2, \ldots, p_M$. Let α_i equal the probability we end up at belief p_i. We can therefore write $p = \sum_{i=1}^{M} \alpha_i p_i$.

In particular, note that p is contained in the convex hull of the p_i. Since the beliefs p_i lie in a space of dimension $N-1$, by Caratheodory's theorem, it follows that there exists some set $T \subset [M]$ such that $|T| \leq N$ and p is a convex combination of $\{p_i | i \in T\}$. Fix such a T. We can then write $p = \sum_{i \in T} \beta_i p_i$.

Let $\gamma = \min_{i \in T} \frac{\alpha_i}{\beta_i}$. Alice and Bob now use public randomness to flip a weighted coin that comes up heads with probability γ. If this coin comes up heads, then Alice samples an element of T according to the distribution induced by the β_i and sends this element to Bob using at most $\log |T| \leq \log N$ bits (by say, specifying its location in T).

If this coin comes up tails, they construct a new probability distribution α' over $[M]$ given by setting $\alpha'_i = \alpha_i - \gamma\beta_i$ for all $i \in [M]$ (where $\beta_i = 0$ if $i \notin T$) and renormalizing. Note that by our choice of γ, it will be the case that $\alpha'_i \geq 0$ for all i; moreover, for at least one i, $\alpha'_i = 0$. They now repeat this process for this new distribution α'.

Note that throughout this modified round, Alice sends in total at most $\log N$ bits to Bob. Moreover, each time they use public randomness, the round either terminates (Alice sends a message to Bob) or the size of the support of α shrinks by one, guaranteeing that the round eventually terminates. Finally, at the end of this process, the probability Bob receives message $i \in [M]$ from Alice is equal to α_i, hence making this modified round information-theoretically equivalent to the original round.

Applying this to every alternation in a protocol π with W rounds results in a protocol π' with communication complexity of at most $W \log N$, as desired. ◀

3.4 Computing Information Complexity

Combining the results of Theorems 8, 10 and 11, we obtain the following result.

▶ **Theorem 12.** *Let π be a communication protocol with information cost C that successfully computes function f over inputs drawn from distribution μ over $\mathcal{A} \times \mathcal{B}$. Then there exists a protocol π' with information cost at most $C + \epsilon$ that also successfully computes f over inputs drawn from μ, but has communication cost at most $b(f, \epsilon)$ where*

$$b(f, \epsilon) = (N\epsilon^{-1})^{O(N)}$$

and $N = |\mathcal{A} \times \mathcal{B}|$.

By a similar rounding technique to that in Section 3.1, we can further ensure each signal in π belongs to a set $\mathbf{S}$ of size $(N\epsilon^{-1})^{O(N^2)}$ (see Section 3.7 of the full paper for details). We can now proceed to prove our main theorem.

Proof of Theorem 1. Fix an $\epsilon > 0$; we will show how to approximate the information complexity of f to within an additive factor of ϵ.

By Theorem 12, there exists some protocol with information cost at most $IC_\mu(f) + \epsilon$ with communication complexity at most $B(f, \epsilon)$ and that only uses signals in the set $\mathbf{S}$. The number of such protocols is finite; in particular each such protocol has at most $2^{B(f,\epsilon)}$ nodes, each of which is labelled by one of $|\mathbf{S}|$ signals. Since $|\mathbf{S}| = (N\epsilon^{-1})^{O(N^2)}$, it follows that the total number of protocols is at most

$$\begin{aligned} |\mathbf{S}|^{2^{B(f,\epsilon)}} &= (N\epsilon^{-1})^{O(N^2)2^{(N\epsilon^{-1})^{O(N)}}} \\ &= 2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)} \end{aligned}$$

The information cost of a protocol with depth B (and thus at most 2^B nodes) can be computed in time $2^{O(B)}$. It follows that computing the minimum information cost of the

above protocols can be done in time $2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)}$, and hence one can approximate $IC_\mu(f)$ to within an additive factor ϵ in time $2^{\exp\left((N\epsilon^{-1})^{O(N)}\right)}$, as desired. ◀

Acknowledgments. We would like to thank Ankit Garg and Noga Ron-Zewi for providing helpful comments on an earlier draft of this paper.

References

1. Noga Alon and Eyal Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *Information Theory, IEEE Transactions on*, 52(5):2172–2176, 2006.
2. Ziv Bar-Yossef, Thathachar S Jayram, Ravindra Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 209–218. IEEE, 2002.
3. Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.
4. Richard Beigel and Jun Tarui. On acc [circuit complexity]. In *Foundations of Computer Science, 1991. Proceedings., 32nd Annual Symposium on*, pages 783–792. IEEE, 1991.
5. Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015.
6. 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*, pages 151–160. ACM, 2013.
7. Mark Braverman and Akhila Rao. Information equals amortized communication. *Information Theory, IEEE Transactions on*, 60(10):6058–6069, 2014.
8. Amit Chakrabart, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 270–278. IEEE, 2001.
9. Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Computational Complexity, 2004. Proceedings. 19th IEEE Annual Conference on*, pages 236–249. IEEE, 2004.
10. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.
11. Rahul Jain. New strong direct product results in communication complexity. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 18, page 2, 2011.
12. Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require superlogarithmic depth. *SIAM Journal on Discrete Mathematics*, 3(2):255–265, 1990.
13. Nan Ma and Prakash Ishwar. Two-terminal distributed source coding with alternating messages for function computation. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 51–55. IEEE, 2008.
14. Nan Ma and Prakash Ishwar. Some results on distributed source coding for interactive function computation. *Information Theory, IEEE Transactions on*, 57(9):6180–6195, 2011.
15. Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.